Entropy Calculations of Single Molecules by Combining the Rigid−Rotor and Harmonic-Oscillator Approximations with Conformational Entropy Estimations from Molecular Dynamics Simulations

Ernesto Suárez, Natalia Díaz, and Dimas Suárez*

Julián Clavería 8, Departamento de Química Física y Analítica, Universidad de Oviedo, Oviedo, 33006 Spain

S *Supporting Information*

ABSTRACT: As shown by previous theoretical and computational work, absolute entropies of small molecules that populate different conformers can be predicted accurately on the basis of the partitioning of the intramolecular entropy into vibrational and conformational contributions. Herein, we further elaborate on this idea and propose a protocol for entropy calculations of single molecules that combines the rigid rotor harmonic oscillator (RRHO) entropies with the direct sampling of the molecular conformational space by means of classical molecular dynamics simulations. In this approach, the conformational states are characterized by discretizing the time evolution of internal rotations about single bonds, and subsequently, the mutual information expansion (MIE) is used to approach the full conformational entropy from the converged probability density functions of the individual torsion angles, pairs of torsions, triads, and so on. This RRHO&MIE protocol could have broad applicability, as suggested by our test calculations on systems ranging from hydrocarbon molecules in the gas phase to a polypeptide molecule in aqueous solution. For the hydrocarbon molecules, the ability of the RRHO&MIE protocol to predict absolute entropies is assessed by carefully comparing theoretical and experimental values in the gas phase. For the rest of the test systems, we analyze the advantages and limitations of the RRHO&MIE approach in order to capture high order correlation effects and yield converged conformational entropies within a reasonable simulation time. Altogether, our results suggest that the RRHO&MIE strategy could be useful for estimating absolute and/or relative entropies of single molecules either in the gas phase or in solution.

■ INTRODUCTION

A common assumption that lies at the heart of many entropy calculations is that the absolute entropy of a single molecule can be separated into several meaningful contributions.[1,2] Perhaps the most straightforward and useful division of entropy is into the whole-body translational and rotational ($S_{trans} + S_{rot}$) and the intramolecular configurational (S_{config}) contributions, the latter one accounting for the entropy of internal degrees of freedom. Interestingly, the development of reliable and cost-effective strategies for computing the configurational entropy of complex molecular systems is a topic that has received much attention during recent years given that, for instance, the ability to compute accurate ΔS_{config} values could be very useful in understanding both the experimental and theoretical data on folding[3−8] and/or association[9−14] of biomolecular systems. Unfortunately, there are still many open questions about S_{config}, concerning its relationship to molecular structure and the importance of correlation among internal motions given that, for relatively large molecular systems, such correlations have been studied only through linear approximations or low-order truncated mutual information expansions.[15−23]

According to Grubmüller and co-workers,[24,25] methods that can compute entropy values can be classified into three broad categories: (a) methods based on the computation of free energy differences using thermodynamic integration,[26] (b) the hypothetical scanning approach developed by Meirovitch and co-workers,[27,28]

and (c) an array of direct methods that extract the entropy of a single molecule from configurations generated by carrying out a conventional molecular dynamics (MD) or Monte Carlo simulation. In this work, we are basically interested in the latter category since our approach aims to estimate biomolecular entropies directly from MD simulations.

In what follows, we will briefly review a family of direct methods for entropy calculations that are more relevant to our approach taking into account that these methods can be divided into *parametric* (quasi-harmonic) and *nonparametric* methods depending whether or not they assume a functional form for the probability density function of the internal degrees of freedom. In addition, we also distinguish a third category of direct methods, the hybrid methods that combine elements of the two kinds (i.e., parametric and nonparametric).

Quasi-Harmonic Methods. The original quasiharmonic analysis was the first example of a direct method applied to biomolecular systems. It was first introduced by Karplus and Kushick, showing that the difference in configurational (i.e., non kinetic) entropy between two molecular conformations can be estimated from their respective covariance matrices.[15] The basic idea is to consider the underlying configurational density function $P(\mathbf{q})$ in the classical configurational entropy, $S_{config} = -k_B \int_C P(\mathbf{q}) \ln$

$P(\mathbf{q})\,d\mathbf{q}$, as a multivariate normal distribution, leading thus to the following entropy expression:

$$S_{\text{config}} = \frac{1}{2}k_B[n + \ln((2\pi)^n \det(\boldsymbol{\sigma}))] \tag{1}$$

where $\det(\boldsymbol{\sigma})$ is the determinant of the covariance matrix of the n internal coordinates. This equation can be applied to estimate only *relative* entropy values. However, the implementation of the quasi harmonic (QH) method suffers from some practical drawbacks due to the required transformation to internal coordinates and the consequent approximations made in the Jacobian. The need for this transformation comes from the fact that removal of the center of mass translation (unavoidable for convergence reasons) makes $\boldsymbol{\sigma}$ singular if Cartesian coordinates are used.[17]

To overcome the limitations of the original QH method, Schlitter has proposed an *ad hoc* approximation to the entropy in which the Cartesian covariance matrix is modified by adding a diagonal matrix so that the resulting $\boldsymbol{\sigma}$ matrix is nonsingular. Starting with the formula for the entropy of a one-dimensional quantum-mechanical harmonic oscillator (HO), Schlitter proposed the following heuristic entropy expression for a system of multiple particles:

$$S'_{\text{HO}} = \frac{1}{2}k_B \sum_i \ln\left(1 + \frac{k_B T e^2}{\hbar^2}\langle q_i^2\rangle_c\right)$$
$$= \frac{1}{2}k_B \ln\left[\prod_i \left(1 + \frac{k_B T e^2}{\hbar^2}\langle q_i^2\rangle_c\right)\right] \tag{2}$$

where $\langle q_i^2\rangle_c$ is the classical variance of the eigenvectors of the mass-weighted covariance matrix $\boldsymbol{\sigma}' = \mathbf{M}\boldsymbol{\sigma}$. Of particular importance is the fact that by adding both translational and rotational entropy contributions to the Schlitter's expression for the configurational entropy, it is possible to estimate absolute rather than relative entropies that are directly comparable to the rigid–rotor harmonic-oscillator (RRHO) entropies obtained from normal-mode analysis and standard statistical thermodynamic formulas.

Following the introduction of Schlitter's method, the QH analysis has been upgraded in order to compute absolute entropies.[18] To this end, the reformulated quasiharmonic approximation (QHA) constructs a pseudoHessian matrix ($\mathbf{H}$) of the molecular system directly from the Cartesian covariance matrix $(\mathbf{H})_{ij} = k_B T(\boldsymbol{\sigma}^{-1})_{ij}$. The corresponding eigenvectors of $\mathbf{H}$ can be seen to represent motional modes around the average system configuration. By associating each of the quasi-harmonic modes with a one-dimensional harmonic oscillator, the total configurational entropy can be approximated as a sum of harmonic contributions.

Despite the improvements introduced since the original formulation of the QH method, either Schlitter's approach or the renovated QHA method suffer from three potential flaws: (a) only linear correlations are taken into account, and therefore, supralinear correlations among the system variables are ignored; (b) formally, the multiminima potential energy surface is exceedingly smoothed by defining only one minimum and ignoring any anharmonicity (including multimodality) of the essentially multi-modal probability density function; (c) there is no clear and unambiguous way to separate the overall rotation from the internal motions (e.g., there are uncertainties up to 80 J/mol K due to the arbitrariness in the choice of reference atoms for the preliminary structure superposition).[29] Accordingly, these methods provide an upper limit to the true absolute entropy S_{tot}.[30]

Furthermore, given a covariance matrix, the function that maximizes entropy is precisely a Gaussian distribution function.[31]

To mitigate their well-known limitations, other authors have proposed several refinements of the QH methods that estimate the importance of the anharmonicity and/or supralinear correlation effects.[16,19,32−34] For example, three years after Karplus introduced the QH analysis, Berendsen et al. proposed a simple strategy to account for the anharmonicity in the configurational probability density function.[16,32] In this approach, the configurational part of the classical entropy $S_{\text{config}}(\mathbf{q})$, where $\mathbf{q} = (q_1, q_2, ..., q_n)$ is an array of internal coordinates, is computed by combining the sum of marginal configurational entropies of the individual q_i variables with the correlation contributions captured by the Karplus model:

$$S_{\text{config}}(q) = -k_B \sum_i \int P(q_i) \ln P(q_i) dq_i + [S_{\text{config}}^{QH} - S_{\text{config, diag}}^{QH}] \tag{3}$$

where S_{config}^{QH} is the configurational entropy computed by eq 1 and $S_{\text{config,diag}}^{QH}$ is the uncorrelated or diagonal Gaussian contribution, calculated by zeroing the nondiagonal elements of the covariance matrix.

Very recently, Baron et al. have proposed a more refined method that estimates both the effects of anharmonicity and supralinear correlations on the classical entropy (i.e., not only configurational entropy).[19,33] The starting point is provided by the renovated QHA approach that uses only Cartesian coordinates.[18] Subsequently, the entropy corrections are estimated from the quasiharmonic coordinates or modes and their corresponding probability densities obtained from the simulations. The classical entropy correction for the nonharmonic behavior, $\Delta S_{\text{cl}}^{\text{ah}} = S_{\text{cl}}^{\text{ah}} - S_{\text{cl}}^{\text{ho}}$, is estimated as the difference between the sum of the marginal entropies obtained directly for the individual modes, $S_{\text{cl}}^{\text{ah}}$, and the entropic term obtained considering the quasiharmonic coordinates as harmonic oscillators, $S_{\text{cl}}^{\text{ho}}$. On the other hand, the pairwise supralinear correction proposed by Baron et al., $\Delta S_{\text{cl}}^{\text{pc}} = \sum_{n>m} (S_{\text{cl,mn}}^{\text{ah}} - S_{\text{cl,m}}^{\text{ah}} - S_{\text{cl,n}}^{\text{ah}})$, is obtained from the classical entropies $S_{\text{cl,mn}}^{\text{ah}}$ computed using the joint probability distribution of the modes m and n, and the corresponding marginal entropies $S_{\text{cl,m}}^{\text{ah}}$ and $S_{\text{cl,n}}^{\text{ah}}$.[19] The estimated entropy finally reads

$$S \approx S_{\text{QHA}} + \Delta S_{\text{cl}}^{\text{ah}} + \Delta S_{\text{cl}}^{\text{pc}} \tag{4}$$

In principle, as remarked upon by the authors, this method can be generalized for the inclusion of higher order correlations.[33] Computational results following this approach have been reported for a microsecond MD trajectory of a peptide model,[33] showing the importance of sufficient phase-space sampling to estimate entropic contributions. It has also been shown that, in accordance with previous studies,[13,19] the pairwise supralinear correlation is normally large while the effect of the anharmonicity on the entropy calculations is relatively small.

Nonparametric Methods. On the basis of the mathematical tools of information theory and advanced statistics, it is possible to estimate the full-dimensional configurational probability density function $P(\mathbf{q})$ without resorting to any analytical approximation, unlike the QHA and Schlitter methods. This is the case of the method of Hnizdo et al.[20] that is based on the use of a series of kth nearest-neighbor (NN) entropy estimators,[35] $\hat{S}_k$, with k being large enough to make a smooth estimation but small enough to make the estimation as local as possible (e.g., $k \in \{1, 2, ..., 5\}$). In

this approach, each molecular configuration is represented as a vector with d components (e.g., the number of internal degrees of freedom). The NN estimation rests on the simple assumption that the configurational probability density function can be approximated *locally* in a nonparametric manner around each sample point q_i using the volume of a d-dimensional sphere centered at q_i and with a radius chosen such that it contains k neighbor data points. Given that the NN $\hat{S}_k$ estimators yield asymptotically unbiased and consistent entropies as the number of data points N increases, they can provide accurate results for any probability distribution provided that sufficiently large samples of molecular simulation data are available. In practice, however, this method has been applied to small molecules due to the large computational cost of the NN searching algorithms when the dimensionality of the problem, d, is larger than 10−15.

Another nonparametric approach has been proposed by Gilson and co-workers based on the mutual information expansion (MIE),[14,21] which is a systematic expansion of the entropy of a multidimensional system in mutual-information terms of increasing order n that capture the n-body correlations among the molecular internal coordinates. The size of the problem can be reduced up to manageable limits by neglecting all fourth- and higher-order MIE terms, thus allowing the calculation of S_{config} values for several small molecules as well as the change in S_{config} upon binding for protein−ligand systems,[14,21] but at the cost of sampling millions of molecular configurations for reaching converged results. A combination of the NN and MIE techniques[36] has also been proposed.

Hybrid Approaches. Very recently, Hensen et al. have developed a new direct method that combines and improves different techniques with the specific aim of making entropy calculations for relatively large biomolecules feasible.[24,25] These authors distinguish three parts or blocks in their method. First, they replace the kth NN entropy estimators by adaptive anisotropic ellipsoidal kernels that capture the configurational density in sufficient detail for up to 45-dimensional spaces. Second, they generate minimally coupled subspaces of internal degrees of freedom by applying a linear orthogonal transformation to Cartesian coordinates in such a way that the mutual information among the resulting coordinates is minimized. The new coordinates are subsequently clustered according to their degree of correlation (correlation among different clusters is neglected). Each oversized cluster ($d > 15$) is subdivided into smaller groups with maximum dimensionality $d = 15$, and its configurational entropy is computed as a sum of the estimated entropy of its components (subclusters) and then corrected by means of mutual information functions. For the stiffest degrees of freedom resulting from the orthogonal transformation, Hensen et al. also propose to employ a generalized quasiharmonic Schlitter formula that accounts for their quantum mechanical nature. Thus, the basic idea behind the method of Hensen et al. is that the combination of parametric and sophisticated nonparametric approaches could help overcome many limitations of the QH methods, but without seriously compromising its applicability to relatively large systems.

Clearly, a key element in the adaptive strategy pursued by Hensen et al. is their statistically based clustering of internal degrees of freedom into subsets that are weakly correlated and the separation into *softer* and *stiffer* degrees of freedom. In this respect, other authors have also designed direct methods that assume (*a priori*) a separation between internal degrees of freedom. For example, the method of Thorpe and Ohkuboa[37] takes advantage of the fact that typical molecular mechanics (MM) Hamiltonians in

implicit solvent are easily separable into *vibrational* (identified as stretching, bending, and improper torsional motions) and *non-vibrational* (identified as torsions) degrees of freedom. These authors show that entropic contributions with respect to an arbitrary temperature for medium-sized systems can be obtained from standard thermodynamical statistical formulas and molecular partition functions which, in turn, can be derived from the density of state functions computed with the weighted histogram analysis method and replica-exchange MD simulations. In this way, it turns out that entropy differences for small- and medium-sized systems can be obtained from conformational and vibrational energy terms. More recently, Brüschweiler and Li[23] have claimed that the configurational entropy is separable into contributions from *hard* and *soft* degrees of freedom (the latter ones identified again with torsion angles) and that correlation effects among the *soft* variables cancel in good approximation on the basis of the results of test calculations on some dipeptide systems. These authors propose then to assess entropy changes between two states of a system at the same temperature, assuming that the entropic contributions of the hard variables do not change. As a first order approximation for conformational entropies is used, this approach is computationally very efficient and can be applicable to protein systems.

Combining the Rigid−Rotor and Harmonic-Oscillator Approximations with Conformational Entropy Estimations. In previous works, we have investigated the role of conformational entropy in the absolute and relative stability of collagen model peptides[22] as well as in the binding of small peptides to the active site of matrix metalloproteases.[38] However, the methodological scope of these previous articles was very narrow, as they are focused on the properties of the biomolecular systems being considered. Hence, in this work, we deal with the methodological details of our protocol for entropy calculations of single molecules whose molecular conformational space is sampled by means of classical MD simulations. Following the original proposal by Karplus and co-workers,[12,39,40] we assume that the total entropy (S_{tot}) of a single molecule (excluding translation and rotation) can be partitioned into a vibrational ($\overline{S}_{vib}$) and a pure conformational contribution ($S_{conform}$):

$$S_{tot} = \overline{S}_{vib} + S_{conform} \tag{5}$$

This simple partitioning scheme can be shown to be formally exact as long as one neglects the entropic contributions from the high energy regions between any pair of wells on the potential energy surface of the molecular system.[12] In addition to the entropy partitioning, our approach is further characterized by the two following features:

- First, we employ the harmonic approximation to compute the mean value of S_{vib} over a time series of representative MD snapshots. The required energy minimization and normal-mode calculations can be done even for relatively large molecular systems provided that molecular mechanics or low level quantum mechanical methods are used. Of course, these normal mode and entropy calculations can be carried out within the framework defined by the conventional RRHO approximations and the standard statistical thermodynamic formulas,[41] thus allowing one to estimate absolute entropies.

- Second, the conformational states along the MD trajectory are determined by means of the discretization of the time evolution of internal rotations about single bonds. This transformation, which does not require any *a priori* separation between softer or stiffer degrees of freedom, implies

that the conformational entropy of the whole MD trajectory ($S_{conform}$) should be naturally computed using the Shannon informational entropy:[36,42,43]

$$S_{conform} = k_B \sum_j^{N_{conf}} (-p_j \ln p_j) \qquad (6)$$

where p_j is interpreted as the statistical weight of the jth conformer. However, at this point, we resort to the above-mentioned MIE method in order to approach the full conformational entropy from the converged probability density functions of the individual torsion angles, pairs of torsions, triads, and so on.

The assumption of the entropy partitioning expressed in eq 5 discriminating between vibrational and conformational entropies; the computation of the mean values of the RRHO entropy contributions accounting for the translational, rotational, and vibrational contributions to the absolute entropy; the discretization of the torsional angles; and the concomitant use of MIE for estimating the conformational entropy from data provided by classical MD simulations constitute, altogether, the basic features of the RRHO&MIE entropy method examined in this work. Nevertheless, it must be noticed that some of these methodological ingredients have been employed in previous calculations of the gas-phase entropy of flexible molecules.[44−49] Thus, it has been shown that the absolute entropy of a *mixture of conformers* can be computed with reasonable accuracy by averaging the RRHO entropy of all of the conformers present at a given temperature and then adding an entropy of *mixing* (ΔS_{mix}) that accounts for the entropic gain in the *mixture* of conformers:

$$S = \bar{S} + \Delta S_{mix} = \sum_\alpha p_\alpha S_\alpha + R \sum_\alpha (-p_\alpha \ln p_\alpha) \qquad (7)$$

where p_α, the molar fraction of the α conformer, is typically estimated with the Maxwell−Boltzmann distribution formula in terms of quantum chemical enthalpies or approximate free energies. All of the distinctive conformers including enantiomeric conformers of the same energy are identified using either a direct counting method or automatic conformational search algorithms. Obviously, p_α and ΔS_{mix} in the "mixture of conformers" method are equivalent, respectively, to the statistical weight of the α conformer and the conformational entropy $S_{conform}$ in the framework defined by the RRHO&MIE protocol, and consequently, the two approaches would be essentially identical. However, we will see that conceptual and practical differences subsist in the way that the $S_{conform}$ and ΔS_{mix} terms are handled and that the RRHO&MIE approach is more suitable for dealing with relatively large systems.
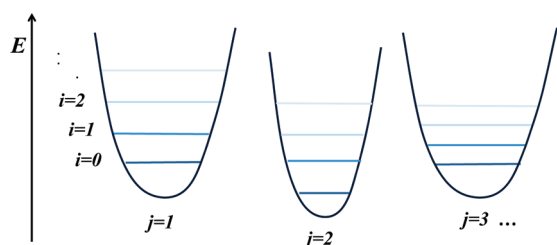
To introduce further details of the RRHO&MIE method and illustrate its potential benefits and limitations, the rest of this paper is organized as follows. First, we will comment on the partitioning of the total entropy (excluding translation and rotation) into the vibrational and conformational components that forms the basis of the RRHO&MIE approach. The separation between vibrational and conformational effects relies heavily on the discretization of the torsional degrees of freedom, and therefore, we will describe this transformation in a detailed manner. Then, after having restated a few definitions about mutual information functions, we will point out that the original MIE expression can be reformulated in such a way that all redundancy in the calculation of the n-order terms is removed. The discretization process combined with the reformulated MIE equation allow us to compute

$S_{conform}$ values including higher order terms beyond the second- or third-order terms that have been considered in most of the previous works. The performance of the RRHO&MIE protocol will be critically discussed on the basis of a series of test calculations on different systems ranging from hydrocarbon molecules in the gas phase to a polypeptide molecule in aqueous solution. For the hydrocarbon molecules (three C_6H_{14} and five C_7H_{16} isomers), their conformational entropies derived from classical MD simulations are combined with their RRHO entropies obtained by carrying out quantum chemical frequency calculations, and the resulting absolute entropies are then compared with experimental data. In this way, we will examine to what extent the assumptions made in the formulation of the RRHO&MIE protocol and the use of classical MD simulations affect the quality of the computed absolute entropies. Second, we will focus on more complex test systems: a series of dipeptide molecules in the gas phase. Although experimental absolute entropies for these test systems are not available, they constitute important cases of study in our validation calculations because they present larger correlation effects among internal degrees of freedom due to the presence of intramolecular H-bond and polar interactions. On the basis of the results obtained for the dipeptide molecules in the gas phase, we will see that the combined discretization process and the MIE approximations are able to capture high order correlation effects and simultaneously yield converged conformational entropies within a reasonable simulation time. Finally, we will analyze the results obtained for a polypeptide molecule in aqueous solution, which can be a representative of the kind of molecular systems for which the RRHO&MIE entropy calculations could be particularly interesting. Besides analyzing the source of conformational correlations, we will also compare the convergence properties and absolute values of S_{config} as provided by the RRHO&MIE (using an implicit solvent model and MM normal mode calculations) and the QHA methods. Overall, we hope that the methodological proposals and the results of our test calculations could be useful to extend the range of applicability of approximate entropy methods for studying more challenging systems involving polypeptide-folding or molecular association processes.

■ THEORY AND COMPUTATIONAL METHODS

a. Decomposition of the Intramolecular Entropy. As shown in previous works by Gilson and Zhou,[40] the entropy of a system with multiple potential energy wells can be formally decomposed into two parts: the entropy that arises from vibrational motions within a single well and the entropy due to conformational transitions between different energy wells. In the case of a single molecule either in the gas phase or in the condensed phase, this entropy decomposition is clearly an approximation whose goodness would depend on the actual molecular structure being studied, temperature, environmental effects, etc. It is true that other entropy methods have been formulated that can derive the total configurational entropy directly from Cartesian coordinates, thus avoiding any assumption about the additivity or lack thereof of the $S_{conform}$ and S_{vib} contributions[19] (although there may remain the problem of separating the overall rotation from the internal motions[29]). However, although we recognize that the total entropy is strictly a global property, its formal decomposition into the conformational and vibrational terms constitutes the basis for the present work. As a matter of fact, many experimental and theoretical methods have provided meaningful results by accepting that free

2641

dx.doi.org/10.1021/ct200216n |*J. Chem. Theory Comput.* 2011, 7, 2638–2653

Scheme 1. Schematic Example of a Multiminima Potential Energy Surface

Scheme 2. High Energy Vibrational Levels Can Be Assigned to Different Energy Basins

energies or entropy contributions can be attributed to particular degrees of freedom and/or physical interactions.[1,2]

In the Supporting Information, we provide an alternative derivation of the formal decomposition of the single molecule entropy assuming that the translational and rotational degrees of freedom have been removed and that the resulting potential energy surface in terms of the remaining internal degrees of freedom can be approximated by a set of *distinguishable* energy basins, as shown in Scheme 1. The final entropy decomposition formula is

$$S_{tot} = \sum_j p_j S_{vib}^j + S_{conform} \tag{8}$$

where p_j and S_{vib}^j stand for the probability and vibrational entropy associated with the jth energy basin, respectively, while $S_{conform}$ is the conformational entropy that arises from the populations of the different energy basins.

The practical implementation of the entropy decomposition (eq 8) implies that each molecular configuration of a single molecule (e.g., one MD snapshot) employed in the entropy calculations should be associated with one molecular j conformer (or j energy basin) representing its conformational state. In our protocol, this assignation is only relevant for the evaluation of the conformational entropy $S_{conform}$ and can be achieved by discretizing the time evolution of the torsion angles (see below). The average value of S_{vib} is obtained separately by means of energy minimization calculations followed by normal-mode analyses within the context of the harmonic oscillator model. In terms of accuracy, the formal entropy decomposition should provide nearly exact results for small molecules in the gas phase (ideal conditions) at room temperature given that, in this case, only the lowest vibrational levels would be populated and they could be tagged univocally to a single conformer.

Could the entropy decomposition provide meaningful results in other more complicated situations such as that depicted in Scheme 2? As temperature increases, high energy vibrational levels that can be simultaneously assigned to different conformers would become populated, partially blurring the distinction between vibrational and conformational motions. Although the practical implementation of eq 8 would not be impeded in this case, the entropy given by eq 8 would be overestimated due to the double-counting of the contributions of those high energy vibrational levels accessible from the two energy basins. Therefore, we propose that, in general, the decomposition of the single molecule entropy could be useful, as it would provide an upper bound to the actual value. Nonetheless, its performance needs to be assessed by carrying out test calculations.

b. On the Use of the Harmonic Oscillator Model for Vibrational Entropies. We think that one remarkable advantage of discriminating between vibrational and conformational entropies is the utilization of the HO model for computing the average
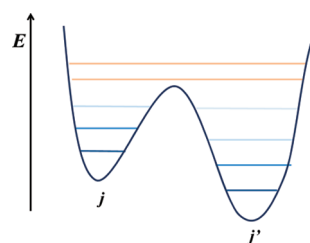
vibrational contribution over a series of representative structures. Clearly, a straightforward computational protocol can be applied for obtaining the mean values of S_{vib}. Starting with a set of representative MD snapshots, optimization calculations relax the internal geometry of each molecular structure to that corresponding to a particular j energy basin, whose vibrational entropy, S_{vib}^j, is subsequently estimated by means of conventional normal mode calculations. In fact, the combination of the HO approximation, the rigid rotor model, and the standard formulas of statistical thermodynamics based upon canonical partition functions can be used to estimate absolute entropies that in some cases admit a direct comparison with experimental data (e.g., in the gas phase).[50] We can also benefit from many efficient implementations for performing either geometry optimizations and second derivative calculations using molecular mechanics[51] and/or quantum mechanical methods depending on the size of the molecular system. Part of the limitations of the HO model (e.g., the lack of anharmonic effects, errors arising from the level of theory, etc.) could be mitigated by using empirical corrections in the form of scaling factors.[52,53] We also note that the RRHO entropies (complemented with the conformational contributions) could be useful within the context of approximate free energy methods like the so-called molecular mechanics Poisson−Boltzmann method.[54] Moreover, previous computational experience with these approaches has shown that the average normal mode entropy converges quite well in terms of the length of simulations and the number of molecular configurations required.[54,55]

c. Discretization of the Torsional Degrees of Freedom. To take into account the $S_{conform}$ contribution, which arises from the population distribution of the molecular conformers, we propose to discretize the probability density functions of those torsion angles that are commonly used to define the molecular conformational state. To this end, we apply the following protocol.

First, we collect the time series with the values of the torsion angles along the MD simulation. For each torsion angle θ, we have a sample of size N, $\{\theta_1, ..., \theta_N \mid \theta_i \in [0, 2\pi)\}$, where N is the number of MD snapshots ($\sim 10^5 - 10^6$). To obtain an analytic representation of the underlying probability density function for the torsion θ, we employ the von Mises kernel estimator (the von Mises density is the circular analogue of the Gaussian distribution). Specifically, the probability density function of a given torsion, $\rho(\theta)$, is then approximated by the arithmetic mean of N von Mises distributions centered on the θ_i values:

$$\hat{\rho}(\theta; \upsilon) = \frac{1}{2\pi N I_0(\upsilon)} \sum_{i=1}^{N} \exp\{\upsilon \cos(\theta - \theta_i)\}$$

where $I_r(\upsilon)$ is the modified Bessel function of order r, and υ, which is the so-called concentration parameter, is the inverse of the smoothing parameter of the kernel estimator. The value of υ

is obtained by applying the recently derived "von Mises-scale plug-in rule"[56] for the smoothing parameter, which results in the following expression that depends on the number N of data points:

$$v = [3N\hat{\kappa}^2 I_2(2\hat{\kappa})\{4\pi^{1/2} I_0(\hat{\kappa})^2\}^{-1}]^{2/5}$$

with $\hat{\kappa}$ being an estimate of the concentration parameter of the global data, for which we take $\hat{\kappa} = 1$, as this value leads to a slightly oversmooth distribution that is more convenient for our purposes. By setting an empirical value for $\hat{\kappa}$ rather than for v, we keep the dependence on N.

The advantage of using the von Mises kernel estimator instead of a normalized histogram method is that we can characterize the analytical properties of $\rho(\theta) \approx \hat{\rho}(\theta;v)$ in order to automatically optimize the location of the maximum and minimum values of $\rho(\theta)$, which is a prerequisite for discretizing the time evolution of the torsion angle θ. This task is performed by the analytical evaluation of the first and second derivatives of $\hat{\rho}(\theta;v)$ over a grid of $\phi_k = k(\pi/(180))$ points with $k = 0, 1, ..., 359$. On one hand, the approximate positions of the $\hat{\rho}(\theta;v)$ critical points are first determined by averaging two consecutive grid points k and $k + 1$ for which $\hat{\rho}'(\phi_k;v)\,\hat{\rho}'(\phi_{k+1};v) \leq 0$. For the torsional distributions, the maximum critical points are easily identified thanks to their largely negative second derivate, and therefore, we first locate the maxima of $\hat{\rho}(\theta;v)$ and use the intermediate position between two consecutive maxima as the initial guess for searching the minima. The minima of $\hat{\rho}(\theta;v)$ are found by means of a steepest descent search that adopts a convergence threshold for the residual gradient of 10^{-4} and employs a linear interpolation approximation for evaluating the gradient on the basis of the $\hat{\rho}'(\phi_k;v)$ values.

Once the $\theta_{\min,i}$ values corresponding to the $\hat{\rho}(\theta;v)$ minima are found (let us suppose that there are m minima and $\theta_{\min,i} < ... < \theta_{\min,m}$), the configurational space of θ defined by the $[0,2\pi]$ interval is divided into m nonoverlapping intervals $([\theta_{\min,1},\theta_{\min,2}), ..., [\theta_{\min,m-1},\theta_{\min,m}), [\theta_{\min,m},2\pi) \cup [0,\theta_{\min,1}))$ that, in turn, define the different conformational states accessible to θ. In this way, the initial time series containing N data points, $\{\theta_1, ..., \theta_N\}$, is easily transformed into a set of N integer numbers $\{a_1, ..., a_N\}$ labeling the conformational states populated by the torsion angle. For example, if θ corresponds to an internal rotation about a $C(sp^3)-C(sp^3)$ bond, its associated a_i variable could have values of 1, 2, and 3 representing the $g+$, $g-$, and *anti* conformations, respectively. Therefore, the continuous variable θ characteristic of the torsion angle becomes a discrete random variable A, whose probability mass function, $P(A)$, can be estimated by the maximum likelihood method fed with its corresponding outcomes $\{a_1, ..., a_n\}$. Finally, we note that the loss of entropy during the $\theta \rightarrow A$ transformation can be expected to be *vibrational*, and that we assume that such a contribution can be reasonably accounted for by normal mode calculations. In other words, the entropy due to the fluctuations of the torsion angles around a local minimum should be recovered by the S_{vib} calculations. This means that, in our approach, it is not necessary to distinguish between *soft* or *hard* degrees of freedom (or between *vibrational* or *nonvibrational* ones) because the conformational entropy turns out to be purely *informational* as a consequence of the discretization process. The probability density function for the torsion angle θ about the C2−C3 bond of 2-methyl-hexane is shown in Figure 1.

d. Mutual Information Expansion: Application to Conformational Entropy Calculations and Reformulation into a Computationally Efficient Scheme More Suitable for Very Large Systems. As shown in the previous section, the conforma-

tional state of a torsion angle can be associated with a one-dimensional random variable A. Analogously, the conformational state of a set of M torsion angles can be described by an M-dimensional random vector $(A_1, ..., A_M)$, or alternatively, the conformational state can also be associated with an ordered set $\{A_1, ..., A_M\}$, where A_i specifies the conformational state of the i torsion and M is the size in terms of the number of torsion angles of our system $\mathscr{A} = \{A_1, ..., A_M\}$. Of course, for medium-sized and large molecules, the number of potentially accessible conformers is huge ($\sim 3^M$), and in this case, obtaining the underlying probability mass function $P(A)$ is practically impossible due to sampling limitations. As other authors have done previously,[21,36] we will use the mutual information expansion as a workaround to this problem. The basic idea here is that if we are able to obtain converged values of the probability mass functions of the individual torsion angles $p(A_i)$, pairs of torsions $p(A_i, A_j)$, triads $p(A_i, A_j, A_k)$, and so on, then we can approach the full-dimensional informational entropy of the $(A_1, A_2, ..., A_M)$ variables (i.e., the conformational entropy) by including systematically n-order correlations among the A_i variables as measured by mutual information functions.[42] More specifically, the mutual information expansion leading to the total entropy can be written as

$$S(A_1, ..., A_M) = \sum_{i=1}^{M} S(A_i) - \sum_{i<j} I_2(A_i, A_j)$$
$$+ \sum_{i<j<k} I_3(A_i, A_j, A_k)$$
$$- \sum_{i<j<k<l} I_4(A_i, A_j, A_k, A_l) + ...$$

where $S(A_i)$ is the informational entropy of the ith torsion angle along the MD simulation while $I_2(A_i, A_j)$, $I_3(A_i, A_j, A_k)$, and so forth are the corresponding mutual information functions that capture the general dependence among the $(A_1, A_2, ..., A_M)$ variables (unlike the covariance function used in the QH methods that only measures linear correlations). More specifically, the mutual information shared by two variables, A_i and A_j, is computed by combining the informational entropies of the single- and two-variable probability mass functions of the torsion angles:

$$I_2(A_i, A_j) = S(A_i) + S(A_j) - S(A_i, A_j)$$

where $S(A_i, A_j)$ is the joint entropy of A_i and A_j. Both $S(A_i)$ and $S(A_i, A_j)$ can be computed using a Shannon-type expression, $-k_B \sum p_\alpha \ln p_\alpha$, where p_α is the corresponding probability mass function and the sum runs over the possible states accessible to the individual torsion angles A_i or to the pair of torsions (A_i, A_j). Similarly, the third-order function $I_3(A_1, A_2, A_3)$ includes correlation effects among three torsion angles:
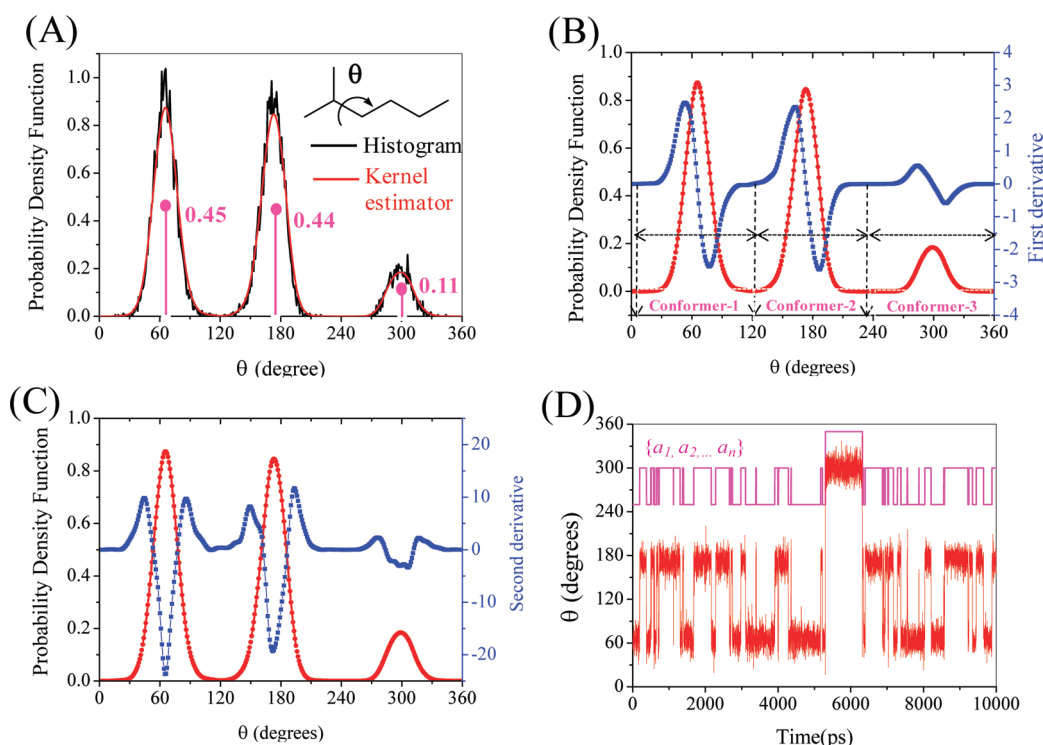
$$I_3(A_i, A_j, A_k) = S(A_i) + S(A_j)$$
$$+ S(A_k) - S(A_i, A_j) - S(A_i, A_k) - S(A_j, A_k)$$
$$+ S(A_i, A_j, A_k)$$

In general, the mutual information shared among k variables $\mathscr{J} = \{A_1, A_2, ..., A_k\}$ can be generalized by

$$I_k(\mathscr{J}) = \sum_{l=1}^{k} (-1)^{l+1} \sum_{\substack{\mathscr{I} \subset \mathscr{J} \\ |\mathscr{I}| = l}} S(\mathscr{I})$$

where for every value of l, the inner sum runs over all possible $\binom{k}{l}$ subsets of $\mathscr{J}$ with l elements, $|\mathscr{I}|$ being the cardinality of $\mathscr{I}$

2643

dx.doi.org/10.1021/ct200216n |*J. Chem. Theory Comput.* 2011, 7, 2638–2653

Figure 1. (A) Probability density function for the torsion angle θ about the C2—C3 bond of 2-methyl-hexane as obtained from a histogram representation and a Von Mises kernel estimator. The probability mass function of the three conformational states is also indicated by the vertical bars. (B and C) Superposition of the probability density function and its first and second derivatives as estimated by the Von Mises kernel. (D) Time evolution of the torsion angle θ and its associated discrete variable A (see text for details).

(i.e., the number of elements). A special case arises when $k = 1$, that is, $I_1(J_1) = S(J_1)$. The total entropy of the M variables $\mathscr{A} = \{A_1, ..., A_M\}$ is usually approximated by a truncated mutual information expansion up to order n,[21,42] which becomes an exact expression when $n = M$:

$$S^{(n)}(\mathscr{A}) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{\substack{\mathscr{J} \subset \mathscr{A} \\ |\mathscr{J}|=k}} I_k(\mathscr{J}) \qquad (9)$$

However, we note again that in previous calculations of configurational entropies based on the MIE approach, only low order (i.e., $n \leq 3$) approximations have been used.

A computational shortcoming of the usual form of the MIE is that, for any subset $\mathscr{J}$ of $\mathscr{A}$ with cardinality lower than n, its entropy must be evaluated more than once. If the order of the expansion is not too high and a reasonable amount of rapid access memory is available, all of the required $S(\mathscr{J})$ terms can be stored and used repeatedly in the calculation of the high order terms. For large and strongly correlated systems (e.g., $M > 100$, $n = 3-5$), however, this simple approach could become unfeasible due to memory depletion, while a direct implementation in which the $S(\mathscr{J})$ terms would be recomputed on-the-fly as needed would be prohibitively expensive. In order to remove all redundancy in the calculation of the MIE terms, the original expression can be reformulated in the following manner:

$$S^{(n)}(\mathscr{A}) = \sum_{k=1}^{n} \left[\sum_{i=0}^{n-k} (-1)^i \binom{M-k}{i} \right] \sum_{\substack{\mathscr{J} \subset \{A_1, ..., A_M\} \\ |\mathscr{J}|=k}}^{M} S(\mathscr{J})$$

$$(10)$$

where the entropy of each subset $\mathscr{J}$ is computed and used only once. The formal proof of the equivalence between eqs 10 and 9 is given in the Supporting Information. To the best of our knowledge, the above expression has not been reported so far, although its implementation could greatly simplify the MIE computational problem for very large systems with hundreds or even more torsion angles. Finally, we note that all of the MIE calculations reported in this work were carried out using a FORTRAN90 code that has been developed in our laboratory and that will be reported elsewhere.[57]

e. Molecular Dynamics Simulation Settings. *Small Systems: Gas-Phase MD Simulations.* MD calculations for each individual molecule were carried out by means of the Amber10 package.[58] The generalized AMBER force field[59] was used for the alkane molecules, while the dipeptides (Ace—X—X—Nme, with X = Ala, Ser, Asn, Leu, and Lys) were represented by the AMBER03 force field.[60] To derive the atomic charges of the alkane molecules, we performed HF/6-31G(d) geometry optimizations of the fully extended conformers followed by single-point B3LYP/cc-pVTZ calculations using the Gaussian 03 program.[61] Other parameters were generated automatically using the antechamber module included in the Amber10 package. Subsequently, 2.0 μs MD trajectories for all of the alkane and dipeptide compounds were run in the gas phase at 298 K and 1.0 atm using a 1.0 fs time step. Coordinates were saved every picosecond of simulation time (2×10^6 structures).

Polypeptide System: MD Simulation in Solution. We simulated the following hexapeptide sequence, Ace—Pro—Phe—Glu~Leu—Arg—Ala—NH$_2$ (termed the PFG peptide), which corresponds to one of the peptide sequences that has been selected from a peptide library mixture for probing the cleavage

site motifs of matrix metalloproteinases (MMPs).[62] Starting coordinates were obtained from conformational search calculations using the LMOD program[63] linked to the Amber package. In the LMOD calculations, we employed the AMBER03 force field coupled with the Hawkins–Cramer–Truhlar pairwise generalized-Born (HCT-GB) model.[64] The lowest energy LMOD structure of PFG was then surrounded by a periodic truncated octahedral box of TIP3P water molecules that extended ~12 Å from the protein atoms (~1400 water molecules). The solvent molecules were initially relaxed by means of energy minimizations and 50 ps of MD. Subsequently, the full system was minimized and heated gradually to 300 K during 50 ps of MD. During the MD simulation, the system remains coupled to a thermal and a hydrostatic bath at $T = 300$ K and $P = 1.0$ atm, the time step of integration was 2.0, the SHAKE procedure on the X–H bonds was applied, and the particle-mesh-Ewald approach was used for nonbonded interactions. A 2.0 μs trajectory was computed, and coordinates were saved for analysis every picosecond.

f. Normal Mode Calculations. *Alkane Molecules: Quantum Mechanical Calculations.* For the alkane molecules, we evaluated the average value of their rotational and vibrational entropies by carrying out quantum mechanical (QM) calculations on selected MD snapshots. To this end, 2000 equally spaced snapshots were extracted from the MD trajectories of each compound. All of the MD snapshots were minimized and scored in terms of their relative MM energies. The relative energies of the relaxed structures were compared in order to filter out all of the energetically equivalent structures, obtaining thus a relatively small set of energetically distinguishable molecular conformers (i.e., this means that only one structure for each pair of enantiomeric conformers is retained). The statistical weight of each energetically unique conformer was estimated by its relative abundance in the initial data set containing the 2000 structures. These conformers were subsequently minimized at the B3LYP/cc-pVTZ level of theory[65,66] and further characterized by analytical frequency calculations. Then, thermal contributions to the gas-phase entropy of the translational, rotational, and vibrational degrees of freedom were obtained within the context of the RRHO approximation and using the B3LYP/cc-pVTZ moments of inertia and vibrational frequencies. Entropy contributions of overall rigid-body rotation take into account the corresponding external symmetry number for each conformer.

Polypeptide Systems: Normal Mode MM Calculations. Taking into account the relatively large size of the polypeptide system, we decided to use MM methodologies for carrying out the required normal mode calculations. Moreover, since the solute and solvent fluctuations are coupled to each other during the MD simulation of the TIP3P water box, we also used the HCT-GB implicit solvent model for removing the explicit consideration of solvent degrees of freedom. Thus, we extracted 10 000 equally spaced snapshots from the 2.0 μs trajectories of PFG. These structures were postprocessed through the removal of all solvent and counterion molecules. Then, solute entropic contributions were estimated for each structure using the NAB package.[67] Prior to the normal mode calculations, the geometries of the system described by the AMBER03 force field were minimized until the root-mean-squared deviation of the elements in the gradient vector was less than 10^{-5} kcal/(mol Å). It may be interesting to note that the large majority of the minimized structures (>98%) corresponded to different minima on the potential energy surface, as expected from the relatively large size of this system. These minimizations and the subsequent normal mode calculations[51] were carried out

using the HCT-GB solvent model. Finally, the RRHO entropic contributions were averaged over the 10 000 snapshots.

■ RESULTS AND DISCUSSION

Absolute Entropies of Alkanes. Table 1 and Figures 2 and 3 collect the results of our entropy calculations for the C_6H_{14} (**1**–**3**) and C_7H_{16} (**4**–**8**) isomers. Of particular interest in these test calculations is the fact that the experimentally available gas-phase entropies of the alkane molecules can be directly compared with our theoretical estimates, since we employ both the RRHO approximation and standard statistical formulas for obtaining the translational–rotational–vibrational entropies of polyatomic molecules,[41] which are subsequently complemented by adding the conformational entropy contributions arising from the discretization of the time evolution of the torsion angles during the classical MD simulations. Furthermore, consideration of these molecules, which are still small enough for exhaustively exploring their conformational space, allows us to show more clearly the relationship of our method with the closely related protocol of the "mixture of conformers" model, which has provided accurate entropy values of small molecules.[44,45]
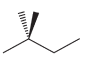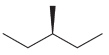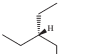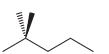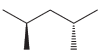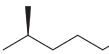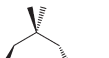
Automatic checks of the relaxed snapshots show that the number of energetically distinguishable conformers (N_E in Table 1) that are populated at 298 K in the gas phase vary between 4 and 15 for the various alkane molecules, except for 2,2-dimethyl-butane (**2**), which shows a single energy level because torsional motions of the methyl groups and about the C2–C3 bond interconnect structures that are degenerate in terms of their potential energy. Precisely, the RRHO entropy value of 2,2-dimethyl-butane at the B3LYP/cc-pVTZ level, 357.77 J/(K mol), is only less than 1 J/(K mol) below the experimental value (378.65), thus showing that the B3LYP/cc-pVTZ level of theory accounts well for the majority of the gas-phase entropy of alkane molecules having a single conformer. This also suggests that frequency scaling, which has been commonly used to correct for systematic errors in the ab initio computation of vibrational frequencies[52] is probably not required at the B3LYP/cc-pVTZ level. For the rest of the test compounds, however, several energy-distinguishable conformers are significantly populated at 298 K. Their individual RRHO entropy values are similar, but they exhibit non-negligible differences as large as 5 J/(K mol).

The mean values of the RRHO absolute entropies ($\overline{S}$ in Table 1) underestimate the experimental data up to tenths of a J/(K mol) for the flexible hydrocarbon molecules. The correlation plot shown in Figure 2A shows more clearly the differences between the experimental and the average theoretical entropies: the resulting squared correlation coefficient is quite low, $R^2 = 0.70$ (a unit slope is imposed), and a large offset at the intercept arises (-14 J/(K mol)). Moreover, the relative entropy values among the C_6H_{14} or C_7H_{16} isomers are poorly described by the average RRHO values: for example, the ΔS_{exp} value between **1** and **2** is -30.17 J/(K mol), whereas the computed $\Delta\overline{S}$ value is only -7.89 J/(K mol). Therefore, it is clear that neglecting conformational entropy contributions significantly affects the quality of the entropy calculations even for simple hydrocarbon molecules.
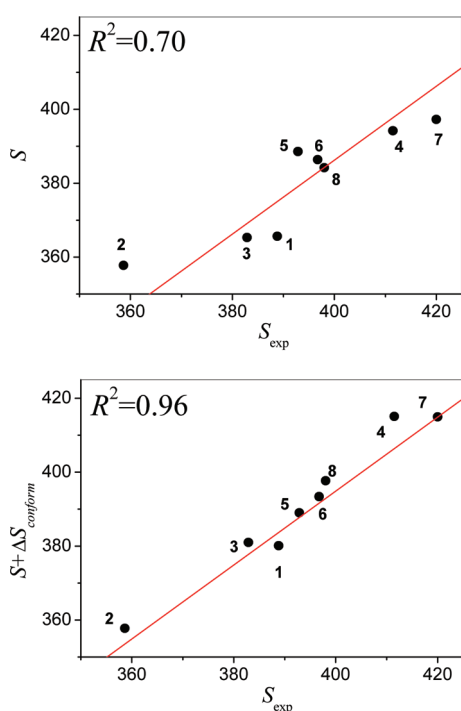
Concerning the conformational entropies of the alkane molecules estimated from the classical MD simulations, we first assess their statistical convergence by plotting the $S_{conform}$ values vs simulation time at various expansion orders for 2-methyl-hexane (**7**) and 3,3-dimethyl-pentane (**8**; see Figure 3). Thus, all of the $S_{conform}$ profiles in Figure 3 converge to nearly zero-slope curves

2645

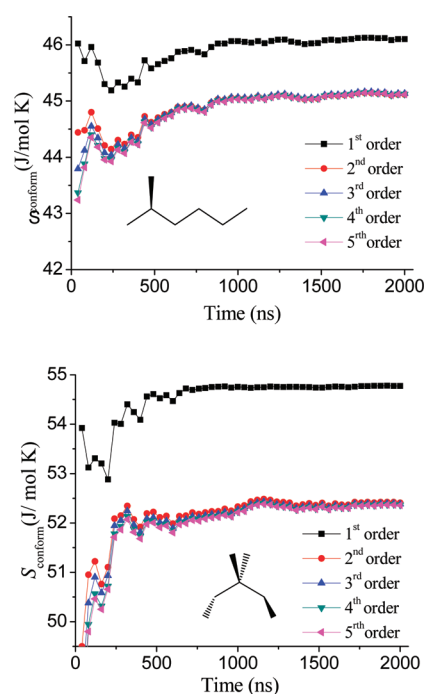dx.doi.org/10.1021/ct200216n |*J. Chem. Theory Comput.* 2011, 7, 2638–2653

Table 1. Average Value of the Translation-Rotational-Vibrational Entropy ($\bar{S}$) and Converged Values of the Conformational Entropy at 298 K (values in J/mol K) for the Alkane Compounds Studied in This Work[a]

	Alkane molecules		N_E	S_{exp} [b]	$\bar{S}$ [c]	$S_{conform}$ [d]	Ω_{term}	$S_{conform}^{term}$	$\bar{S}+\Delta S_{conform}$ [e]
1	hexane		9	388.82 ±0.84	365.94	32.44	3^2	18.27	380.11
2	2,2-dimethyl-butane		1	358.65 ±0.84	357.77	45.67	3^5	45.67	357.77
3	3-methyl-pentane		5	382.88 ±0.67	365.27	43.11	3^3	27.40	380.98
4	3-ethyl-pentane		8	411.50	393.78	48.7	3^3	27.40	415.08
5	2,2-dimethyl-pentane		4	392.88	388.56	46.13	3^5	45.67	389.02
6	2,4-dimethyl-pentane		4	396.73	386.38	43.56	3^4	36.54	393.40
7	2-methyl-hexane		15	419.99	397.25	45.12	3^3	27.40	414.96
8	3,3-dimethyl-pentane		4	398.02	381.90	52.35	3^4	36.54	397.71

[a] The number of energetically indistinguishable conformers at $T = 0$ (Ω_{term}) and the number of energetically distinguishable conformers (N_E) at 298 K are also indicated. [b] From references: hexane,[68] 2,2-dimethyl-butane,[69] 3-methyl-pentane,[70] 3-ethyl-pentane,[71] 2,2-dimethyl-pentane,[71] 2-4-dimethyl-heptane,[71] 2-methyl-hexane,[71] and 3,3-dimethyl-pentane.[70] [c] $\bar{S} = S_{trans} + \bar{S}_{rot}^{RR} + \bar{S}_{vib}^{HO}$ and using the B3LYP/cc-pVTZ geometries and frequencies. The S_{rot} and S_{vib} contributions are averaged according to the relative frequency of the energetically distinguishable conformers during the classical AMBER MD simulations. [d] Including the fifth-order MIE conformational entropy derived from the classical MD trajectories. [e] $\Delta S_{conform} = S_{conform} - S_{conform}^{term}$ (see text for details).

Figure 2. Correlation plots between theoretical absolute entropies (in J/(K mol)) for the eight alkane molecules considered in this work before (A) and after (B) adding the conformational entropy contribution to the mean RRHO entropies.

Figure 3. Convergence plots of the gas-phase conformational entropy (J/mol K) for 2-methyl-hexane and 3,3-dimethyl-pentane.

with respect to simulation time after ∼1.5 and ∼1.0 μs for **7** and **8**, respectively, regardless of the expansion order. An even

faster convergence was observed for the rest of the alkane molecules (see Figure S1 in the Supporting Information), and therefore, we conclude that the reported $S_{conform}$ values for these small alkane molecules essentially lack any statistical uncertainty. Although the

2646

dx.doi.org/10.1021/ct200216n |J. Chem. Theory Comput. 2011, 7, 2638–2653

RRHO entropy is clearly the largest part, the $S_{conform}$ values, which range between ~30 and ~50 J/(K mol), represent a significant contribution. Analyzing now the relative importance of correlation effects in $S_{conform}$, we find that the first-order MIE, which assumes that all torsion angles are independent variables, overestimates the conformational entropy. This behavior is not entirely unexpected because, for these small systems, the convergence plots in Figure 3 and Figure S1 (Supporting Information) strongly suggest that all correlation effects are effectively taken into account by our calculations, and as a consequence, the total entropy diminishes with respect to the first order value. However, we see in Figure 3 and Figure S1 that the entropy curves at first order recover most of the total conformational entropy; that is, entropy reduction due to correlation effects is rather small, only 1—5 J/(K mol). Moreover, it turns out that this reduction is basically due to pair correlation, as the converged second order values are practically indistinguishable from the rest of the higher order $S_{conform}$ entropies in all cases. Therefore, from these test calculations on the selected alkane molecules, we conclude that (a) $S_{conform}$ is a significant entropic contribution and (b) the degree of correlation among the torsional degrees of freedom is rather low so that the second order MIE approximation gives sufficiently accurate values for the examined alkane molecules.

Before comparing our theoretical absolute entropies for the various C_6H_{14} and C_7H_{16} isomers with experimental data, it is necessary to realize that our $S_{conform}$ calculations, which are based on classical MD simulations, imply that individual atoms in a covalently bound molecule are distinguishable particles. Therefore, in order to compare our data with experimental third-law entropies, we have to remove from the $S_{conform}$ values the conformational entropy that arises from the number of possible rearrangements (Ω_{term}) that a single molecule can formally undergo through internal rotations about bonds to terminal symmetrical groups (e.g., $-CH_3$) without altering any molecular property. For instance, in the case of 2,2-dimethyl-butane, internal rotation of each of the four terminal methyl groups as well as around the C2—C3 bond generates three conformers per rotatable bond of identical energy and molecular properties, so that 2,2-dimethyl-butane has a total of $\Omega_{term} = 3^5$ possible intramolecular arrangements that would result in an entropy contribution $S_{conform}^{term} = R \ln \Omega_{term} = 45.67$ J/(K mol) (the $S_{conform}^{term}$ values for the rest of the alkane molecules can be likewise computed). The addition of the entropy differences, $\Delta S_{conform} = S_{conform} - S_{conform}^{term}$, to the RRHO mean values $\overline{S}$ allow us to properly compare between theoretical and experimental data. Note also that entropic effects due to the presence of enantiomeric conformers are taken into account automatically by the $S_{conform}$ calculations. We see in Figure 2 that the linearity of the correlation plot between the experimental and RRHO-based entropies is quite improved after having added the $\Delta S_{conform}$ term to $\overline{S}$. Thus, the squared correlation coefficient is now 0.96, and the intercept is around −2.6 J/(K mol).

The acceptable goodness of the linear fit between experimental and theoretical data in Figure 2 confirms the importance of $S_{conform}$ and supports the usefulness of decomposing the entropy into its vibrational and conformational parts. In the Supporting Information, we present more calculations of the absolute entropies of the eight alkane molecules by using the above-mentioned "mixture of conformers" model.[44,45] These calculations, which are based entirely on QM data, suggest that a significant fraction of the observed error in the RRHO&MIE conformational entropy calculations can emerge from small unbalances in the probability density functions of torsion angles. On the other hand, further
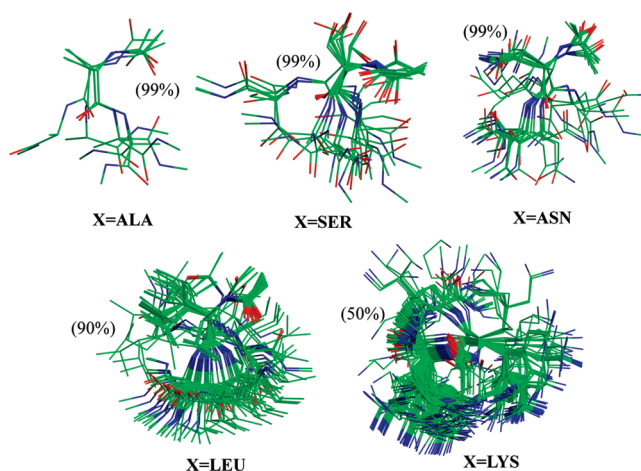
Figure 4. Superposition of the energetically distinguishable conformers of the Ace−X−X−Nme systems. The percentage of the MD snapshots represented by the conformers is given in parentheses.

support for applying the entropy decomposition can be found in previous works,[44−49] following the "mixture of conformers" strategy. Therefore, it may also be relevant to comment here about the quality of the results of the "mixture of conformers" model. For example, the mean unsigned difference (MUD) between experimental and theoretical gas-phase entropies reported by Guthrie[47] for 128 organic compounds with up to 10 carbon atoms is 3.7 J/(K mol) (data derived from unscaled B3LYP/6-31G** frequencies and a semiempirical estimation of the ΔS_{mix} term). A similar MUD (4.4 J/(K mol)) was observed in our calculations on the eight alkane compounds.

Conformational Entropy of Dipeptides. The purpose of introducing the MIE approach within the context of the conformational entropy calculations is to capture correlation among torsional motions. For the alkane molecules studied in this work, it turns out that such correlation effects are almost negligible, and in any case, they are entirely accounted for by the pairwise approximation (i.e., second order correlation). Therefore, it is necessary to analyze more complex molecules in order to assess the ability of our approach for estimating higher order entropic contributions and its dependence on the dimensionality of the problem as defined by the number M of rotatable bonds. To this end, we examined five dipeptide molecules (Ace−X−X−Nme) of increasing size with X = Ala, Ser, Asn, Leu, and Lys (i.e., the peptides are capped by acetyl (Ace) and N-methyl (Nme) groups). The potential ability of these molecules to dynamically form and break intramolecular interactions through direct H-bond contacts or through-space electrostatic forces can introduce a significant correlation in their torsional motions and simultaneously maintain an important flexibility.

Figure 4 shows the superposition of the most populated conformers of the Ace−X−X−Nme systems that were obtained from minimizing 2000 equally spaced MD snapshots in each system and selecting the energetically distinguishable structures. As expected, all of the dipeptides are flexible molecules in the gas phase that experience frequent conformational transitions along the MD simulations. For the X = Ala system with $M = 8$ rotatable bonds, 99% of conformational variability is represented by only four structures, but the number of populated conformers grows up rapidly with M: the X = Lys system with $M = 16$ rotatable bonds populates more than 1200 different structures. Of course,

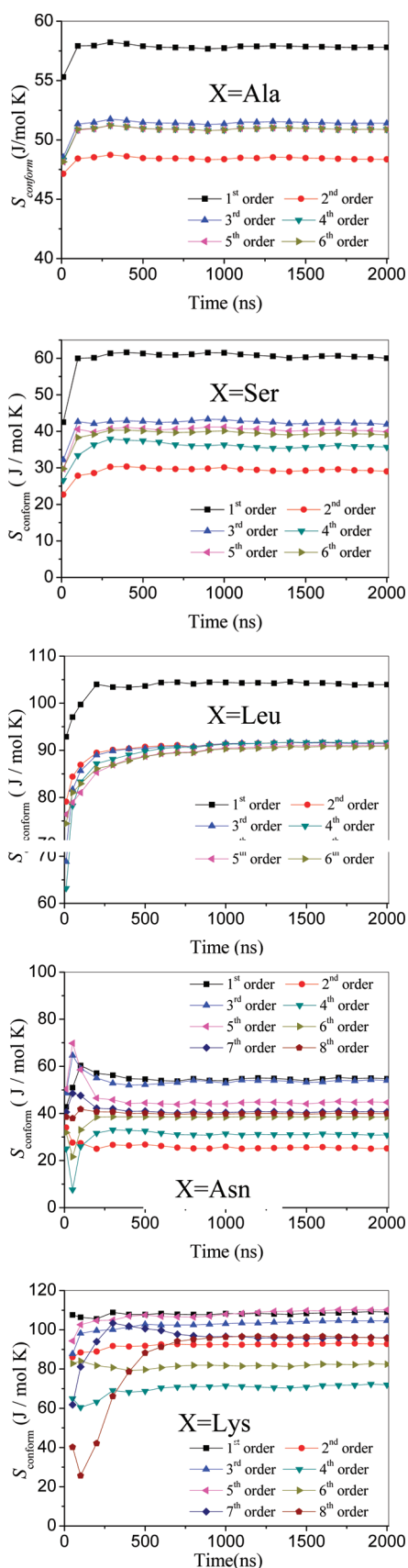**Figure 5.** Convergence plots of the gas-phase conformational entropy $(J/(K\,mol))$ for the Ace$-$X$-$X$-$Nme molecules with X = Ala, Ser, Asn, Leu, and Lys.

the structure and relative abundance of the conformers are influenced by H-bond interactions that interconnect main chain groups (e.g., between the Ace$-$C$=$O carbonyl and the H$-$N$-$X amide at the second residue position) and/or side chain groups. Thus, we found that, in general, at least two or three H-bond contacts have abundances of about 55$-$65% so that they are quite stable interactions.

The conformational entropies at various orders of the Ace$-$X$-$X$-$Nme molecules are plotted along the simulation time in Figure 5. For the majority of the systems, it turns out that their $S_{conform}$ profiles converge to nearly zero-slope curves after $\sim$0.5 μs, thus suggesting that the MD simulations have exhaustively explored their phase space. However, the limiting $S_{conform}$ values show a clear dependence on the expansion order n of the MIE method (see also Table 2). In consonance with expectations, the magnitude and relative weight of the n-order corrections increase with M on going from X = Ala $(M = 8)$ to X = Lys $(M = 16)$. Thus, with respect to the unimportance of high order correlation effects in Figure 3, we see now that both the intramolecular interactions and the partial rigidity introduced by the amidic bonds in the dipeptide molecules can have an impact on the conformational entropy thanks to the appearance of correlation among torsional motions. For the Ace$-$Ala$-$Ala$-$Nme system, which has only one more rotatable bond than the C_7H_{16} isomers and populates just a few conformers (see Figure 4), the second and third order corrections to $S_{conform}$ are noticeable: -9.4 and 3.0 J(K mol). These correlation effects are more clearly seen in the rest of the Ace$-$X$-$X$-$Nme systems as their $S_{conform}$ curves span a range of tenths of a J/(K mol). In order to achieve convergence in the $S_{conform}$ values with respect to the expansion order n, MIE calculations up to sixth to eighth order were required for several systems (see Table 2). Moreover, we see in Figure 5 that n-order corrections to the conformational entropy fluctuate in both sign and magnitude depending on the molecular system, revealing thus the true complexity of the mutual information expansion. Nevertheless, the extended sampling $(2 \times 10^6$ configurations) and the ability of the MIE calculations to estimate high order correlation effects allow us to obtain reasonably well-converged $S_{conform}$ values for all of the Ace$-$X$-$X$-$Nme systems. In the case of the X = Asn and X = Lys systems, the difference between the $S_{conform}$ estimations up to seventh and eighth order amounts to only $\sim$0.2 kJ/mol in terms of free energies at 300 K. Similarly, the corresponding free energy differences for the $S_{conform}$ values of the rest of the Ace$-$X$-$X$-$Nme dipeptides at the fifth and sixth orders are also negligible.

An interesting comparison can be made between the X = Leu $(M = 14)$ and Asn $(M = 13)$ systems. On one hand, the hydrophobic system has a considerable conformational entropy $(\sim 90\,J/(K\,mol))$, which is in consonance with its relatively large number of accessible conformers. Curiously, high order $(n > 3)$ corrections to $S_{conform}$ are very small, which is very probably related to the fact that the Ace$-$Leu$-$Leu$-$Nme molecule forms only two intramolecular H-bond interactions with moderate abundances (<35%). On the other hand, much fewer conformers are populated during the MD trajectory of the X = Asn system, but correlation effects are now rather important, most likely because three to four H-bond interactions with abundances between 30% and 65% involving either side chain or backbone groups are constantly being formed and broken during the simulation, thus coupling the conformational changes of torsions separated by several covalent bonds. For the X = Lys system with $M = 16$ rotatable bonds, in addition to the relevance of correlation effects,

Table 2. Limiting Values of the Conformational Entropy at Various Orders (in J/(mol K)) for Dipeptides in the Gas Phase after 2.0 μs of Simulation Time[a]

| | N_E | M | $S_{conform}$ | | | | | | | |
			1	2	3	4	5	6	7	8
Ace−(Ala)$_2$−Nme	14	8	57.80	48.36	51.40	50.88	50.88	50.93		
Ace−(Ser)$_2$−Nme	25	10	60.03	29.02	41.92	35.66	39.90	39.06		
Ace−(Asn)$_2$−Nme	69	13	54.76	25.14	54.00	30.86	44.70	38.41	40.66	39.90
Ace−(Leu)$_2$−Nme	232	14	103.96	91.49	91.44	91.64	91.10	90.84		
Ace−(Lys)$_2$−Nme	1357	16	109.08	92.70	104.64	71.97	110.18	82.41	95.91	95.74

[a] The number of rotatable bonds considered in the conformational entropy calculations (M) and the number of energetically distinguishable conformers (N_E) observed in a sample of 2000 MD snapshots are also indicated. A total of 2×10^6 MD snapshots were used in all of the calculations.

its $S_{conform}$ curves at high orders exhibit a slower convergence with respect to the simulation time (see Figure 5) as a consequence of the larger dimensionality of the problem, which demands a greater sampling effort to obtain reliable probability mass functions for all of the torsion angles and their combinations. Nevertheless, as mentioned above, the $S_{conform}$ values for X = Lys were converged to ∼96 J/(K mol) at orders n = 7 and 8, a value which happens accidentally to be quite close to the second order one (∼93; see Table 2).

Absolute Entropy of the PFG Peptide. On the basis of the test calculations on the alkane molecules, it seems that there is no clear preference between the "mixture of conformers" and the conformational entropy frameworks for estimating the total entropy of small molecules. However, a sharp difference between the two formally equivalent approaches appears in larger biomolecules for which it is almost impossible to explore all their conformational space as required by the "mixture of conformers" approach. But as mentioned in the Introduction, the partitioning of the single molecule entropy, the discretization of the torsional degrees of freedom, and the adoption of the MIE technique can provide altogether a workaround to the sampling limitation for estimating the total entropy of complex molecular systems from the output provided by extensive MD simulations. To better illustrate the potential applicability of such an approach, we present here the results obtained for the PFG hexapeptide molecule, which is known to be a ligand of the MMP enzymes.

The 2.0 μs MD trajectory of the PFG peptide in explicit solvent, which was started from the initial structure favored by the LMOD algorithm, populates two different conformational regions of the solute molecule characterized by radii of gyration of ∼5.0 and ∼6.0 Å, respectively (Figure S3 in the Supporting Information). Secondary structure analyses assign a helical conformation to the central (i.e., 2−6) residues in ∼55% of the analyzed snapshots (we note in passing that the fact that PFG tends to adopt a helical structure seems in consonance with the ability of the MMPs to bind and hydrolyze collagen peptide chains that have also a helical structure). However, although PFG possesses some secondary structure, it still exhibits a relatively large dynamical flexibility through either its backbone or side chain motions, and therefore, it is conceivable that conformational entropy could be large enough to play a significant role in the free energy change upon binding of PFG to MMPs.[38]

Most likely, the full theoretical understanding of the role played by entropy in the activity of PFG would require the computation of its absolute entropy in aqueous solution, as suggested by our previous calculations on the complexes formed between the MMP-2 enzyme and small peptide substrates.[38] In our approach,

such calculations could be achieved by combining the RRHO entropy of PFG using an implicit solvent model with the solute conformational entropy derived from the MD simulation so that the resulting entropy would be combined with other free energy terms as defined by approximate methodologies like the MM-PBSA protocol.[54] As the PFG hexapeptide contains 111 atoms and virtually all of the MD snapshots correspond to energetically distinguishable conformers, we decided to perform energy minimization and normal mode calculations on a subset of 10 000 MD snapshots using the NAB package and the AMBER03 force field (note that DFT calculations would be computationally too expensive). As shown in Figure 6A, although the RRHO entropies significantly fluctuate, the resulting time series over 2.0 μs is rather stable and the corresponding mean value of S (1415 J/(K mol)) was estimated to within a standard error of only 0.3 J/(K mol). This small uncertainty suggests that the average RRHO entropy of PFG can be considered sufficiently converged for most purposes.

Turning our attention to the convergence plots of conformational entropy in Figure 6B derived from 10^6 MD snapshots, we note first that the $S_{conform}$ values show a large dependence on the MIE order. In fact, the sum of marginal entropies leads to a limiting value of 137 J/(K mol) that is ∼60% above the entropy estimations made from second to fifth order approximations. Given that PFG is a flexible molecule that contains a significant number of rotatable bonds (M = 25) and has several polar groups capable of forming H-bond interactions, the entropy reduction caused by correlation effects is well understood. However, for the same reasons, the $S_{conform}$ calculations at the fourth and fifth orders now have a poor convergence with respect to simulation time: there remains an uncertainty of a few J/(K mol) in their limiting values after 2.0 μs. Besides the sampling limitations at high order, it is also clear that the $S_{conform}$ values of PFG have a non-negligible uncertainty with regard to the n order employed in the calculations. For example, the limiting value of $S_{conform}$ at second order is ∼3 kJ/mol in terms of free energies below that at third order.

Segregation of the conformational entropy into backbone and side chain contributions allows us to further analyze the origin of the large correlation effects in the dynamics of PFG. We see in Figures 6C,D that the backbone (M = 11) and side chain (M = 14) $S_{conform}$ curves show an acceptable convergence at the various orders. Curiously, the entropy curves reflecting the conformational changes of the ψ and ϕ torsion angles present marked oscillations and converge more slowly than those of the amino acid side chains, indicating thus that the solvent-exposed side chains move faster than the backbone chain, and therefore, their motions are more efficiently sampled by the MD simulations. Correlation
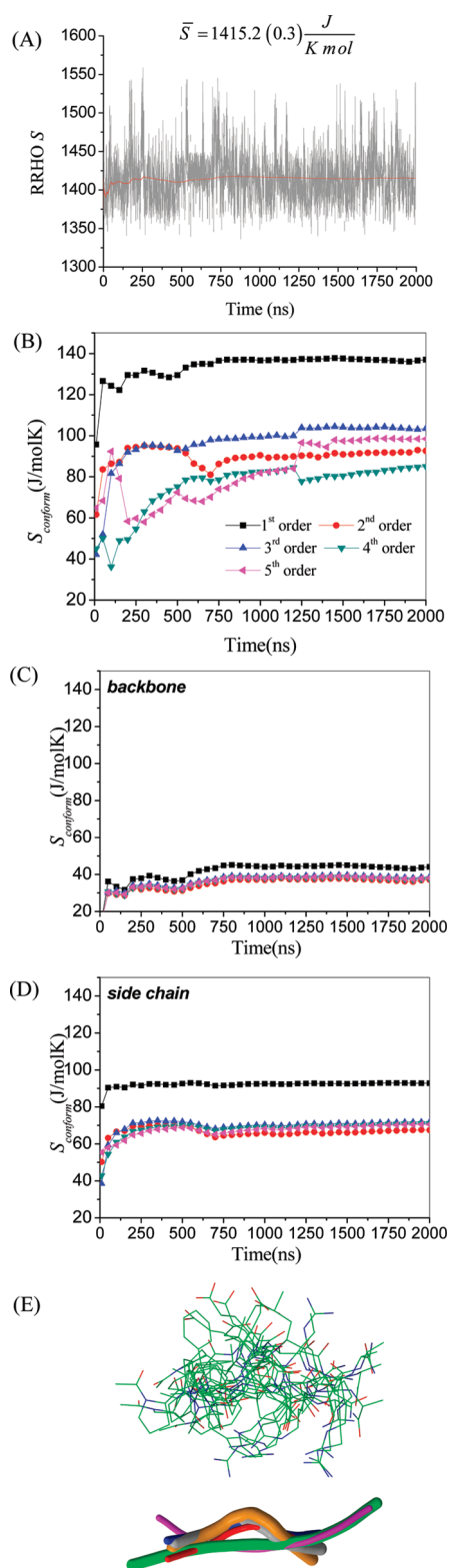
Figure 6. (A) RRHO entropy (in J/(K mol)) calculated for 10 000 snapshots extracted at 200 ps intervals from the 2.0 μs MD simulations of PFG. The average value and its standard error in parentheses are also indicated. (B−D) Convergence plots of the PFG conformational entropy (in J/(K mol)). (E) Superposition of the most populated representative structures derived from clustering analyses both in wireframe and ribbon model representation. Thickness of the ribbon models corresponds to the number of snapshots represented by each model.
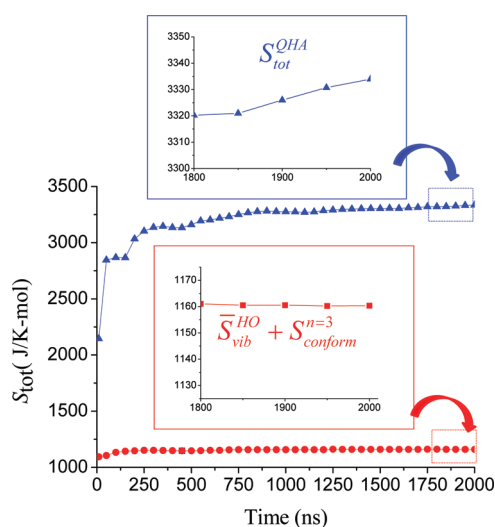
Figure 7. Convergence plots of the total entropy of PFG (excluding translation and rotation) estimated by the QHA method and the addition of the mean vibrational entropy and the conformational entropy at third order.

among the ψ and ϕ angles is only moderate ($\sim$6 J/(K mol)) and is largely captured by the second order approximation. Similarly, second- or third-order corrections to $S_{conform}$ are able to include most of the correlation effects due to the conformational motions of the side chains, but the magnitude of the concomitant entropy reduction is larger ($\sim$ 22 J/(K mol)). When comparing the plots in Figure 6B−D, it is clear that the source of the stronger correlation effects in the conformational entropy of the PFG peptide stem from the coupling between the backbone and side chain motions. We also see that, in terms of their convergence properties, the separate $S_{conform}$ values for the backbone and side-chain torsions are much more reliable than the global data as the segregated entropy plots reach stable plateaus and the free energy differences between the third-, fourth-, and fifth-order estimations are very small ($\sim$0.1 kJ/mol in terms of free energies).

Finally, we compare the additive entropy $\overline{S}_{vib}^{HO} + S_{conform}^{n=3}$ with the results provided by the quasi-harmonic approximation QHA, which constructs a pseudo-Hessian matrix directly from the covariance matrix of the Cartesian coordinates. To remove the overall translation of the center of mass and the overall rotation of the protein, the 2×10^6 MD snapshots employed in the QHA calculations were superposed on top of each other using a least-squares fit. As mentioned in the Introduction, the QHA method exhibits several disadvantages (e.g., neglecting supralinear correlations, approximating multimodal distributions to a unimodal one, etc.) that ultimately result in a large nonsystematic *over-estimation* of S_{config}. This effect is clearly observed in Figure 7 as the QHA calculations lead to a very large entropy value ($\sim$3333 J/(K mol)). Although the entropy decomposition should give an upper limit to the true entropy, the $\overline{S}_{vib}^{HO} + S_{confirm}^{n=3}$ value ($\sim$1160 J/(K mol)) is much lower than the QHA one even though the $-TS_{conform}$ contribution has a statistical uncertainty of several kJ/mol. Moreover, besides the larger overestimation of S_{config}, the QHA calculations also exhibit worse convergence properties (see the two insets in Figure 7).

From the results of the test calculations on PFG, it can be concluded that the total conformational entropy has a considerable weight (10%) in its single-molecule entropy and that the effects of dynamic correlations among torsional angles are far

2650

dx.doi.org/10.1021/ct200216n |*J. Chem. Theory Comput.* 2011, 7, 2638–2653

from being small. These calculations point out that both under-sampling and poor convergence with respect to the MIE order are two closely related problems of the $S_{conform}$ calculations that need to be assessed in practical applications. However, although the uncertainty in the total $S_{conform}$ of PFG can have an impact of several kJ/mol on free energy, we believe that meaningful results can be achieved from partially converged entropy curves like those shown in Figures 6B−D for PFG. For example, computation of entropy differences (e.g., upon peptide binding to a host molecule) could benefit from partial cancellation of errors and provide approximately constant ΔS values at different MIE orders.[22,38] Alternatively, consideration of a subset of torsion angles (e.g., backbone ψ and ϕ angles) could be enough for capturing the relative change in entropy occurring in peptide folding or molecular association processes. In any case, an intensive MD sampling followed by the estimation of conformational entropy differences including correlation effects would be required to fully understand the thermodynamical forces controlling many biomolecular processes that alter the conformational dynamics of the involved molecules. In this respect, the present test calculations support the ability of our approach for including high order correlation effects that have been neglected so far in most of the previous studies using nonparametric methods for estimating the configurational entropy.

SUMMARY AND CONCLUSIONS

In this work, we have pursued the implementation of the partitioning of the intramolecular entropy into vibrational and conformational contributions as originally proposed by Karplus et al., in order to estimate the absolute entropy of single biomolecules from MD simulations. In our approach, a key element consists of the characterization of the conformational state of a given molecule by means of the discretization of the time evolution of its torsional motions about single bonds. This process leads naturally to the computation of the conformational entropy as Shannon entropy, which is subsequently added to the average translational—rotational—vibrational entropy computed by means of normal model calculations on a series of MD snapshots. To help overcome sampling limitations in the computation of the conformational probability mass function of large molecules, we use the mutual information expansion, which has also been employed in other entropy methods, to systematically include correlation effects among torsion angles. Although this protocol has been developed for treatment with relatively large systems, it must be emphasized that its core assumption, that is, the combination of the RRHO entropy with the conformational entropy, is formally equivalent to the "mixture of conformers" strategy that has been routinely used to predict absolute entropies of small molecules with good accuracy.

On the basis of the different test calculations that have been presented in this work, we can draw the following conclusions regarding the applicability and/or reliability of our approach: (a) The gross of the absolute entropy is computed with the RRHO approximation, which constitutes a straightforward computational protocol that avoids the need of discriminating between stiff or soft degrees of freedom. (b) From a quantitative point of view, the combination of the RRHO entropy with the conformational (or mixing) entropy yields results that are quite close to experimental data, as shown by our calculations on the alkane molecules or by other results previously reported in the literature. (c) For computing reliable conformational entropies of small- or

medium-sized molecules that exhibit a rich dynamical behavior, it is essential to capture correlation effects arising from the coupling of torsional motions through intramolecular interactions, and in this respect, the use of the MIE method together with the discretization of the torsional motions constitutes an interesting alternative capable of computing high order corrections quite efficiently. (d) Segregation of conformational entropy into different components (e.g., backbone and side chain terms) can be easily implemented, thus revealing the origin and relative importance of correlation effects.

Finally, it is also important to comment on some limitations of the RRHO-conformational entropy calculations on the basis of the results obtained for the PFG polypeptide. Thus, it is clear that the RRHO&MIE calculations for this kind of system may demand a considerable amount of computer time, particularly if energy minimizations and normal mode calculations have to be carried out on systems containing thousands of atoms, even though an MM method was employed. Similarly, the computation of conformational entropy corrections at high orders ($n > 5-8$) can be rather expensive too. In addition, application of the present approach to large molecules can suffer from convergence issues with respect to the simulation time needed to extract the probable mass functions of individual torsions and groups of torsions, and with reference to the MIE order that is required to capture correlation effects. We believe that the calculation of relative conformational entropies rather than absolute ones using low order approximations and/or for a subset of torsions may benefit from error cancellation. However, the problems of the RRHO&MIE technique should be mitigated by increasing the MD sampling, which in turn, is becoming more and more accessible thanks to the continuous improvement in the efficiency in computer hardware and simulation algorithms. Moreover, the fact that the current protocol has been shown to exhibit a much better convergence behavior than the QHA method in the case of the polypeptide entropy calculation is a promising result that should stimulate further methodological and computational experimentation aimed at overcoming the computational bottlenecks and/or convergence problems.

ASSOCIATED CONTENT

S **Supporting Information.** Derivation of the entropy partitioning (eq 8). Mathematical proof of the equivalence between eqs 9 and 10. Figure S1 showing convergence plots of conformational entropies for various alkane molecules. Figure S2 and Tables S1 and S2 summarizing the results of the "mixtures of conformers" entropy calculations on the alkane molecules. Figure S3 showing the time evolution of the PFG radius of gyration. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*Phone: +34-985103689. Fax: +34-985103125. E-mail: dimas@uniovi.es.

ACKNOWLEDGMENT

■ REFERENCES

(1) Carlsson, J.; Aqvist, J. Calculations of Solute and Solvent Entropies from Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5385.

(2) Brady, G. P.; Sharp, K. E. Entropy in Protein Folding and in Protein-Protein Interactions. *Curr. Opin. Struct. Biol.* **1997**, *7*, 215.

(3) Fitter, J. A. Measure of Conformational Entropy Change During Thermal Protein Unfolding Using Neutron Spectroscopy. *Biophys. J.* **2003**, *84*, 3924.

(4) Bachmann, A.; Kiefhaber, T.; Boudko, S.; Engel, J.; Bächinger, H. P. Collagen Triple-Helix Formation in All-Trans Chains Proceeds by a Nucleation Growth Mechanism with a Purely Entropic Barrier. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13897.

(5) Creamer, T. P.; Rose, G. D. Side-Chain Entropy Opposes a-Helix Formation but Rationalizes Experimentally Determined Helix-Forming Propensities. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5937.

(6) Chapagain, P. P.; Parra, J. L.; Gerstman, B. S.; Liu, Y. Sampling of States for Estimating the Folding Funnel Entropy and Energy Landscape of a Model Alpha-Helical Hairpin Peptide. *J. Chem. Phys.* **2007**, *127*, 075103.

(7) Choa, S. S.; Levya, Y.; Wolynesa, P. G. Quantitative Criteria for Native Energetic Heterogeneity Influences in the Prediction of Protein Folding Kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 434.

(8) Schäfer, H.; Daura, X.; Mark, A. E.; van Gunsteren, W. F. Entropy Calculations on a Reversibly Folding Peptide: Changes in Solute Free Energy Cannot Explain Folding Behavior. *Proteins: Struct., Funct., Genet.* **2001**, *56*, 43.

(9) Grünberg, R.; Nilges, M.; Leckner, J. Flexibility and Conformational Entropy in Protein-Protein Binding. *Structure* **2006**, *14*, 683.

(10) Diehl, C.; Genheden, S.; Modig, K.; Ryde, U.; Akke, M. Conformational Entropy Changes Upon Lactose Binding to the Carbohydrate Recognition Domain of Galectin-3. *J. Biomol. NMR* **2009**, *45*, 157.

(11) Stone, M. J. Nmr Relaxation Studies of the Role of Conformational Entropy in Protein Stability and Ligand Binding. *Acc. Chem. Res.* **2001**, *34*, 379.

(12) Chang, C. A.; Chen, C.; Gilson, M. K. Ligand Configurational Entropy and Protein Binding. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1534.

(13) Baron, R.; McCammon, J. A. (Thermo)Dynamic Role of Receptor Flexibility, Entropy, and Motional Correlation in Protein-Ligand Binding. *ChemPhysChem* **2008**, *9*, 983.

(14) Killian, B. J.; Yudenfreund-Kravitz, J.; Somani, S.; Dasgupta, P.; Pang, Y.-P.; Gilson, M. K. Configurational Entropy in Protein—Peptide Binding: Computational Study of Tsg101 Ubiquitin E2 Variant Domain with an Hiv-Derived Ptap Nonapeptide. *J. Mol. Biol.* **2009**, *389*, 315.

(15) Karplus, M.; Kushick, J. N. Method for Estimating the Configurational Entropy of Macromolecules. *Macromolecules* **1981**, *14*, 325.

(16) Edholm, O.; Berendsen, H. J. C. Entropy Estimation from Simulations of Non-Diffusive Systems. *Mol. Phys.* **1984**, *51*, 1011.

(17) Schlitter, J. Estimation of Absolute and Relative Entropies of Macromolecules Using the Covariance Matrix. *Chem. Phys. Lett.* **1993**, *215*, 617.

(18) Andricioaei, I.; Karplus, M. On the Calculation of Entropy from Covariance Matrices of the Atomic Fluctuations. *J. Chem. Phys.* **2001**, *115*, 6289.

(19) Baron, R.; Gunsteren, W. F. v.; Hünenberger, P. H. Estimating the Configurational Entropy from Molecular Dynamics Simulations: Anharmonicity and Correlation Corrections to the Quasi-Harmonic Approximation. *Trends Phys. Chem.* **2006**, *11*, 87.

(20) Hnizdo, V.; Darian, E.; Fedorowicz, A.; Demchuk, E.; Li, S.; Singh, H. Nearest-Neighbor Nonparametric Method for Estimating the Configurational Entropy of Complex Molecules. *J. Comput. Chem.* **2006**, *28*, 655.

(21) Killian, B. J.; Kravitz, J. Y.; Gilson, M. K. Extraction of Configurational Entropy from Molecular Simulations Via an Expansion Approximation. *J. Chem. Phys.* **2007**, *127*.

(22) Suárez, E.; Díaz, N.; Suárez, D. Entropic Control of the Relative Stability of Triple-Helical Collagen Peptide Models. *J. Phys. Chem. B* **2008**, *112*, 15248.

(23) Li, D.-W.; Bruschweiler, R. In Silico Relationship between Configurational Entropy and Soft Degrees of Freedom in Proteins and Peptides. *Phys. Rev. Lett.* **2009**, *102*, 118108.

(24) Hensen, U.; Grubmüller, H.; Lange, O. F. Adaptive Anisotropic Kernels for Nonparametric Estimation of Absolute Configurational Entropies in High-Dimensional Configuration Spaces. *Phys. Rev. E* **2009**, *80*, 011913.

(25) Hensen, U.; Lange, O. F.; Grubmüller, H. Estimating Absolute Configurational Entropies of Macromolecules: The Minimally Coupled Subspace Approach. *PLoS ONE* **2010**, *5*, e9179.

(26) *Free Energy Calculations. Theory and Applications in Chemistry and Biolog*; Chipot, C., Pohorille, A., Eds.; Springer-Verlag: Berlin, 2007.

(27) Meirovitch, H. Methods for Calculating the Absolute Entropy and Free Energy of Biological Sys. *J. Mol. Recognit.* **2010**, *2*, 153.

(28) Meirovitch, H.; Cheluvaraja, S.; White, R. P. Methods for Calculating the Entropy and Free Energy and Their Application to Problems Involving Protein Flexibility and Ligand Binding. *Curr. Protein Pept. Sci* **2009**, *10*, 229.

(29) Schäfer, H.; Mark, A. E.; Gunsteren, W. F. v. Absolute Entropies from Molecular Dynamics Simulation Trajectories. *J. Chem. Phys.* **2000**, *113*, 7809.

(30) Chang, C.; Chen, W.; Gilson, M. K. Evaluating the Accuracy of the Quasiharmonic Approximation. *J. Chem. Theory Comput.* **2005**, *1*, 1017.

(31) Cover, T. M.; Thomas, J. C. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2006.

(32) Di Nola, A.; Berendsen, H. J. C.; Edholm, O. Free Energy Determination of Polypeptide Conformations Generated by Molecular Dynamics. *Macromolecules* **1984**, *17*, 2044.

(33) Baron, R.; Hünenberger, P. H.; McCammon, J. A. Absolute Single-Molecule Entropies from Quasi-Harmonic Analysis of Microsecond Molecular Dynamics: Correction Terms and Convergence Properties. *J. Chem. Theory Comput.* **2009**, *5*, 3150.

(34) Numata, J.; Wan, M.; Knapp, E. Conformational Entropy of Biomolecules: Beyond the Quasi-Harmonic Approximation. *Genome Inform.* **2007**, *18*, 192.

(35) Goria, M. N.; Leonenko, N. N.; Mergel, V. V.; Novi-Inverardi, P. L. A New Class of Random Vector Entropy Estimators and Its Applications in Testing Statistical Hypotheses. *J. Nonparametr. Stat.* **2005**, *17*, 277.

(36) Hnizdo, V.; Tan, J.; Killian, B. J.; Gilson, M. K. Efficient Calculation of Configurational Entropy from Molecular Simulations by Combining the Mutual-Information Expansion and Nearest-Neighbor Methods. *J. Comput. Chem.* **2008**, *29*, 1605.

(37) Ohkuboa, Y. K.; Thorpe, I. F. Evaluating the Conformational Entropy of Macromolecules Using an Energy Decomposition Approach. *J. Chem. Phys.* **2006**, *124*, 024910.

(38) Díaz, N.; Suarez, D.; Suarez, E. Kinetic and Binding Effects in Peptide Substrate Selectivity of Matrix Metalloproteinase-2: Molecular Dynamics and Qm/Mm Calculations. *Proteins* **2010**, *78*, 1.

(39) Karplus, M.; Ichiye, T.; Pettit, B. M. Configurational Entropy of Native Proteins. *Biophys. J.* **1987**, *52*, 1083.

(40) Zhou, H.-X.; Gilson, M. K. Theory of Free Energy and Entropy in Noncovalent Binding. *Chem. Rev.* **2009**, *109*, 4092.

(41) McQuarrie, D. A. *Statistical Mechanics*; University Science Books: Sausalito, CA, 2000.

(42) Matsuda, H. Physical Nature of Higher-Order Mutual Information: Intrinsic Correlations and Frustration. *Phys. Rev. E* **2000**, *62*, 3098.

(43) Shannon, C. E.; Weaver, W. *The Mathematical Theory of Communication*; The University of Illinois Press: Urbana, IL, 1964.

(44) DeTar, D. F. Theoretical Ab Initio Calculation of Entropy, Heat Capacity, and Heat Content. *J. Phys. Chem. A* **1998**, *102*, 5128.

(45) DeTar, D. F. Calculation of Entropy and Heat Capacity of Organic Compounds in the Gas Phase. Evaluation of a Consistent Method without Adjustable Parameters. Applications to Hydrocarbons. *J. Phys. Chem. A* **2007**, *111*, 4464.

(46) Block, D. A.; Armstrong, D. A.; Rauk, A. Gas Phase Free Energies of Formation and Free Energies of Solution of Rc-Centered Free Radicals from Alcohols: A Quantum Mechanical-Monte Carlo Study. *J. Phys. Chem. A* **1999**, *103*, 3562.

(47) Guthrie, J. P. Use of Dft Methods for the Calculation of the Entropy of Gas Phase Organic Molecules: An Examination of the Quality of Results from a Simple Approach. *J. Phys. Chem. A* **2001**, *105*, 8495.

(48) Bouchoux, G.; Bimbong, R. G.-B.; Nacer, F. Gas-Phase Protonation Thermochemistry of Glutamic Acid. *J. Phys. Chem. A* **2009**, *113*, 6666.

(49) Bouchoux, G.; Bourcier, S.; Blanc, V.; Desaphy, S. Gas Phase Protonation Thermochemistry of Phenylalanine and Tyrosine. *J. Phys. Chem. B* **2009**, *113*, 5549.

(50) East, A. L. L.; Radom, L. Ab Initio Statistical Thermodynamical Models for the Computation of Third-Law Entropies. *J. Chem. Phys.* **1997**, *106*, 6665.

(51) Brown, R. A.; Case, D. A. Second Derivatives in Generalized Born Theory. *J. Comput. Chem.* **2006**, *27*, 1662.

(52) Scott, A. P.; Radom, L. Harmonic Vibrational Frequencies: An Evaluation of Hartree—Fock, Møller—Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. *J. Phys. Chem.* **1996**, *100*, 16502.

(53) Johnson, R. D.; Irikura, K. K.; Kacker, R. N.; Kessel, R. Scaling Factors and Uncertainties for Ab Initio Anharmonic Vibrational Frequencies. *J. Chem. Theory Comput.* **2010**, *6*, 2822.

(54) Gohlke, H.; Case, D. A. Converging Free Energy Estimates: Mm-Pb(Gb)Sa Studies on the Protein—Protein Complex Ras—Raf. *J. Comput. Chem.* **2003**, *25*, 238.

(55) Lee, M. S.; Olson, M. A. Calculation of Absolute Protein-Ligand Binding Affinity Using Path and Endpoint Approaches. *Biophys. J.* **2009**, *90*, 864.

(56) Taylor, C. C. Automatic Bandwidth Selection for Circular Density Estimation. *Comput. Stat. Data An.* **2008**, *52*, 3493.

(57) Suárez, E.; Suárez, D. Manuscript in preparation.

(58) Case, D. A.; Darden, T. A.; Cheatham, I. T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, K. F.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER*; 10th ed.; University of California: San Francisco, 2008.

(59) Case, D. A.; Cheatham, T. E.; Darden, T.; Gholke, H.; R., L.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668.

(60) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *14*, 1999.

(61) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, C.02 ed.; Gaussian Inc.: Wallingford, CT, 2004.

(62) Turk, B. E.; Huang, L. L.; Cantley, L. C. Determination of Protease Cleavage Site Motifs Using Mixture-Based Oriented Peptide Libraries. *Nat. Biotechnol.* **2001**, *19*, 661.

(63) Kolossváry, I.; Guida, W. C. Low Mode Search. An Efficient, Automated Computational Method for Conformational Analysis: Application to Cyclic and Acyclic Alkanes and Cyclic Peptides. *J. Am. Chem. Soc.* **1996**, *118*, 5011.

(64) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, Rna, and Phosphoramidate-DNA Helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401.

(65) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648.

(66) Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007.

(67) *Modeling Unusual Nucleic Acid Structures*; Macke, T., Case, D. A., Eds.; American Chemical Society: Washington, DC, 1998.

(68) Scott, D. W. Correlation of the Chemical Thermodynamic Properties of Alkane Hydrocarbons. *J. Chem. Phys.* **1974**, *60*, 3144.

(69) Kilpatrick, J. E.; Pitzer, K. S. The Thermodynamics of 2,2-Dimethylbutane, Including the Heat Capacity, Heats of Transition, Fusion and Vaporization and the Entropy. *J. Am. Chem. Soc.* **1946**, *68*, 1066.

(70) Finke, H. L.; Messerly, J. F. 3-Methylpentane and 3-Methylheptane: Low-Temperature Thermodynamic Properties. *J. Chem. Thermodyn.* **1973**, *5*, 247.

(71) Huffman, H. M.; Gross, M. E.; Scott, D. W.; McCullough, J. P. Low Temperature Thermodynamic Properties of Six Isomeric Heptanes. *J. Phys. Chem.* **1961**, *65*, 495.

2653

dx.doi.org/10.1021/ct200216n |*J. Chem. Theory Comput.* 2011, 7, 2638–2653