JCTC Journal of Chemical Theory and Computation

2D Entropy of Discrete Molecular Ensembles

J. Wang† and R. Brüschweiler*,‡

Carlson School of Chemistry and Biochemistry, Clark University, Worcester, Massachusetts, and Department of Chemistry and Biochemistry and National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32306

Received April 29, 2005

Abstract: A method is presented for the estimation of the conformational entropy of discrete macromolecular ensembles associated with multiple rotameric dihedral angle states. A covariance matrix is constructed of all mobile dihedral angles, which are represented as complex numbers on the unit circle, and subjected to a principal component analysis. The total entropy is decomposed into additive contributions from each eigenmode, for which a 2D entropy is computed after convolution of the projection coefficients of the conformer ensemble for that mode with a 2D Gaussian function. The method is tested for ensembles of linear polymer chains for which the exact conformational entropies are known. These include chains with up to 15 dihedral angles exhibiting two or three rotamers per dihedral angle. The performance of the method is tested for molecular ensembles that exhibit various forms of correlation effects, such as ensembles with mutually exclusive combinations of rotamers, ensembles with conformer populations biased toward compact conformers, ensembles with Gaussian distributed pairwise rotamer energies, and ensembles with electrostatic intramolecular interactions. For all these ensembles, the method generally provides good estimates for the exact conformational entropy. The method is applied to a protein molecular dynamics simulation to assess the effect of side-chain−backbone and side-chain−side-chain correlations on the conformational entropy.

1. Introduction

The thermodynamic stability of macromolecular states, such as ordered versus disordered states, is determined by their free energies, reflecting the balance between enthalpic and entropic contributions. Reliable estimates of entropy changes are, therefore, essential for the prediction and understanding of free energy changes.[1−9] For macromolecular systems, such as polymers and proteins, an important contribution is the conformational entropy. Unfortunately, the conformational entropy cannot be calculated analytically except for the simplest energy potentials. As an alternative, computer simulations are often used to sample relevant parts of the conformational space of macromolecules. Although such simulations produce discrete sets of conformers, the straightforward application of Boltzmann's equation $S = k \ln W$ to a computed trajectory with W snapshots generally bears little relevance with respect to the entropy. The ensemble of conformers first needs to be converted into probability distributions of relevant degrees of freedom before an entropy can be evaluated. In the quasiharmonic analysis method by Karplus and Kushick, the distribution of the various degrees of freedom of the discrete molecular ensemble, generated, for example, by molecular dynamics (MD) simulations, is approximated by a multivariate Gaussian distribution.[3,10] The quantity that enters the expression for the conformational entropy is the determinant of the covariance matrix of the coordinate fluctuations, which includes correlation effects between different degrees of freedom up to second order. Extensions of this approach have been developed that include

* Corresponding author phone: (850) 644−5173; fax: (850) 644−1366; e-mail: bruschweiler@magnet.fsu.edu.
† Clark University.
‡ Florida State University.

10.1021/ct050118b CCC: $33.50
Published on Web 11/03/2005

quantum-mechanical zero-point vibrational effects[12−17] and that address pure intramolecular reorientational entropic contributions.[18]

The quasiharmonic approximation does not always hold because the probability distribution of soft degrees of freedom, such as dihedral angles, is often significantly non-Gaussian as a result of anharmonic motions and the population of multiple rotameric states. Various methods that address these effects have been described.[19−23] In the method by Edholm and Berendsen, the conformational entropy is separately determined for each internal coordinate from the probability distribution of the ensemble along the coordinate by representing it as a histogram with a variable bin width.[20,21] A correction for correlation effects between internal coordinates is made by adding the difference of the quasiharmonic entropies in the presence and absence of correlations. More rigorous and computationally rather expensive alternative methods are Meirovitch's hypothetical scanning and local states methods that are capable of including correlation effects beyond second order (see ref 9 and references therein) and the method by Demchuk and coworkers that was applied to systems with one and two internal rotational degrees of freedom.[22,23]

In the present work, we describe a new method for estimating the conformational entropy of discrete molecular ensembles. It includes correlation effects up to second order in the complex representation $e^{i\varphi}$ of the molecule's internal torsion angles φ. A two-dimensional Gaussian distribution is assigned to each conformer along each eigenmode, and the conformational entropy is then determined as the sum of the entropy terms $-\int p(z) \log p(z)\, dz$ calculated along each mode. The method is first tested for a rotational isomeric state (RIS) model of polymer chains for which entropies can be determined analytically for reference. Different kinds of correlation effects are introduced to test the ability of the model to adequately reflect the entropy reduction associated with such effects. The model is finally applied to a MD trajectory of the protein ubiquitin.

2. Methods

We consider a linear polymer chain with N_a atoms connected by bonds of uniform length b and fixed bond angles. Each conformation (conformer) is fully specified by the $N_d = N_a - 3$ intervening dihedral angles φ_k, where $k = 1, ..., N_d$. The dihedral angles are represented as points on the unit circle in the complex plane, $z_k = e^{i\varphi_k}$, which circumvents the modulo 2π ambiguity of φ_k. Each conformer j is then specified by a vector $|d^{(j)}\rangle$:

$$|d^{(j)}\rangle = \{e^{i\varphi_1(j)}, e^{i\varphi_2(j)}, ..., e^{i\varphi_{N_d}(j)}\} \quad (1)$$

For an ensemble of N_c conformers, a complex covariance matrix $\mathbf{C}$ can be defined with elements

$$C_{kl} = \langle e^{i\varphi_k} e^{-i\varphi_l}\rangle - \langle e^{i\varphi_k}\rangle\langle e^{-i\varphi_l}\rangle,\ k, l = 1, ..., N_d \quad (2)$$

where the angular brackets indicate population-weighted averaging over the N_c conformers, for example, $\langle e^{i\varphi_k} e^{-i\varphi_l}\rangle = \sum_{j=1}^{N_c} p_j e^{i\varphi_k} e^{-i\varphi_l}$ where p_j is the population of conformer j with $\sum_j p_j = 1$, that is, for a conformational ensemble with a uniform distribution of populations $p_j = 1/N_c$. Using Euler's identity, the matrix elements of eq 2 can be expressed as

$$C_{kl} = \mathrm{cov}(\cos\varphi_k, \cos\varphi_l) + \mathrm{cov}(\sin\varphi_k, \sin\varphi_l) - \mathrm{icov}(\cos\varphi_k, \sin\varphi_l) + \mathrm{icov}(\sin\varphi_k, \cos\varphi_l) \quad (3)$$

where $\mathrm{cov}(f, g) = \langle f^* g\rangle - \langle f^*\rangle\langle g\rangle$.

A principal component analysis is then applied to matrix $\mathbf{C}$ by solving the eigenvalue problem $\mathbf{C}|m\rangle = \lambda_m|m\rangle$. The conformational entropy along each eigenmode $|m\rangle$ is calculated in the following way. First, each conformer $|d^{(j)}\rangle$ is projected along eigenmode $|m\rangle$, which yields the complex projection coefficients

$$c_{mj} = \langle m|d^{(j)}\rangle \quad (4)$$

The projection coefficients define the probability distribution of the conformational ensemble along mode m,

$$P_m(z)\, dz = \sum_{j=1}^{N_c} p_j \delta(z - c_{mj})\, dz \quad (5)$$

where $\delta(z - c_{mj}) = \delta[x - \mathrm{Re}(c_{mj})]\, \delta[y - \mathrm{Im}(c_{mj})]$ and where $\delta(x)$ is Dirac's delta function. Because of the finite number of conformers, $P_m(z)$ has a singular shape that is unsuitable for estimating entropies, and a smoothing procedure needs to be applied first.[24] The following procedure is used here: $P_m(z)$ is convoluted with a 2D Gaussian distribution with a standard deviation σ, $(2\pi\sigma^2)^{-1} \exp[-zz^*/(2\sigma^2)]$, which is normalized to ensure that the effective probability is constant. This yields the smoothed probability distribution

$$\tilde{P}_m(z)\, dz = \frac{1}{2\pi\sigma^2} \sum_{j=1}^{N_c} p_j \exp[-(z - c_{mj})(z^* - c_{mj}^*)/(2\sigma^2)]\, dz \quad (6)$$

σ is a smoothing parameter that needs to be calibrated in order to provide quantitative entropies as described below.

The entropy along mode m is then obtained by

$$S_m = -\int \tilde{P}_m(z) \ln \tilde{P}_m(z)\, dz \quad (7)$$

where the integral extends over the Gaussian plane. Here and in the following, Boltzmann's constant k_B is omitted; that is, all entropies are given in units of k_B unless noted otherwise. To correct for the net effect of the finite width σ on the entropy, a reference entropy S_{ref}, which is independent of m, is subtracted from S_m

$$S_{ref} = -\int \tilde{P}_{ref}(z) \ln \tilde{P}_{ref}(z)\, dz \quad (8)$$

where $P_{ref}(z)\, dz = (1/2\pi\sigma^2) \exp[-zz^*/(2\sigma^2)]\, dz$. Thus, the total entropy is obtained as

$$S_{2D} = \sum_{m=1}^{N_d} (S_m - S_{ref}) \quad (9)$$

Note that modes with eigenvalue $\lambda = 0$ do not yield a net contribution to S_{2D}. Because the entropy is computed from a distribution in the complex plane, that is, in two dimensions, it is termed S_{2D}. For the numerical evaluation of S_m

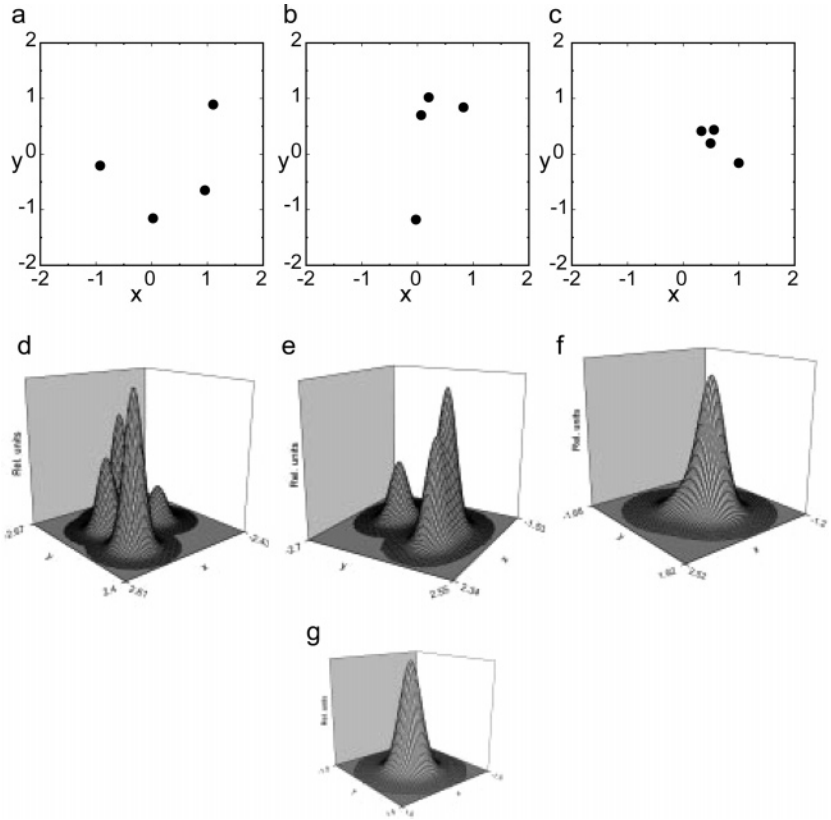

Figure 1. 2D probability distributions and projection coefficients of an ensemble of four conformers belonging to a linear chain molecule consisting of six atoms with three mobile dihedral angles. The dihedral angles of the four conformers are (φ_1, φ_2, φ_3) = (259.9°, 314.9°, 311.4°) for conformer 1, (19.9°, 74.9°, 311.4°) for conformer 2, (259.9°, 74.9°, 71.4°) for conformer 3, and (19.9°, 194.9°, 311.4°) for conformer 4, and their populations are $p_1 = 0.4$, $p_2 = 0.3$, $p_3 = 0.2$, and $p_4 = 0.1$, respectively. The covariance matrix **C** was calculated according to eq 2. Panels a−c show the complex projection coefficients c_{mj} (eq 4) for the three eigenvectors with nonzero eigenvalues $\lambda_1 = 1.243$, $\lambda_2 = 0.758$, and $\lambda_3 = 0.069$. The *x* and *y* axes correspond to the real and imaginary axes, respectively, of the complex plane. Panels d−f show the corresponding probability distributions calculated by the convolution of a 2D Gaussian function (panel g) with a standard deviation $\sigma = 0.3$ with the projection coefficients of panels a−c using eq 6.

and S_{ref}, in eqs 7 and 8, the integrals over the complex plane $z = x + iy$ are replaced by sums over a two-dimensional grid with boundaries at ±5 along *x* and *y* of each projection coefficient and a grid size of $\sigma/10$. Thus, for each mode, more than 10 000 grid points are evaluated. The 2D entropy can be compared with the analytical entropy

$$S_{\mathrm{a}} = -\sum_{j=1}^{N_{\mathrm{c}}} p_j \ln p_j \tag{10}$$

which, for uniform populations, $p_j = 1/N_{\mathrm{c}}$ is equivalent to Boltzmann's relationship $S_{\mathrm{a}} = \ln N_{\mathrm{c}}$.

3. Results

The entropy estimator described above is first tested for the RIS model of simple polymer chains for which the exact conformational entropy is known. In this model, the polymer is represented as a linear chain molecule consisting of N_{a} atoms with constant bond angles of 109.5° defined by consecutive atom triples. Each dihedral angle φ_k, which is defined by four consecutive atoms, occupies either $N_{\mathrm{r}} = 3$ or $N_{\mathrm{r}} = 2$ or rotameric states corresponding to a jump angle $\Delta\varphi$ of 120° and 180°, respectively. For each dihedral angle, the value of the first rotamer is either 0° (i.e., the four atoms defining the dihedral angle lie in the same plane forming a "cis" geometry) or it is chosen randomly between 0° and 360°. Excluded volume effects are considered by excluding any conformer for which one or more interatomic distances are shorter than the bond length *b*. For random values of the first rotamers, the total number of sterically allowed conformers may vary for different ensembles with the same number of dihedral angles.

An example of how discrete sets of projection coefficients are converted into continuous probability distributions is given in Figure 1. The projection coefficients are defined in eq 4, and the probability distributions used to evaluate the entropy are given in eqs 5−9. The figure shows the probability distributions for the three largest modes of a linear chain molecule consisting of $N_{\mathrm{a}} = 6$ atoms and $N_{\mathrm{d}} = 3$ dihedral angles with $N_{\mathrm{r}} = 3$. The generated ensemble consists of four conformers with conformer populations $p_1 = 0.4$, $p_2 = 0.3$, $p_3 = 0.2$, and $p_4 = 0.1$. Panels a−c display the projection coefficients for the three largest modes with $\lambda_1 = 1.243$, $\lambda_2 = 0.758$, and $\lambda_3 = 0.069$. Panels d−f show the corresponding probability distributions $\tilde{P}_m(z)$ after con-

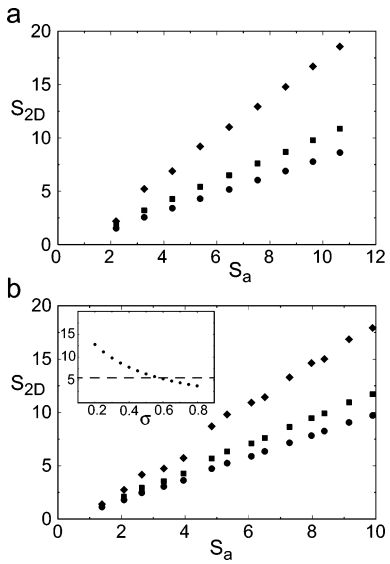

Figure 2. 2D entropy (eq 9) calculated for uniformly populated ensembles with $N_d = 2-10$ dihedral angles and $N_r = 3$ rotamers (panel a) and ensembles with $N_d = 2-14$ dihedral angles and $N_r = 2$ rotamers (panel b) vs the analytical entropy $S_a = \ln N_c$ (eq 10). The width σ was set to 0.3 (diamonds), 0.5 (squares), and 0.6 (circles). The insert in panel b shows the dependence of S_{2D} on σ for eight flexible dihedral angles (filled circles), and the horizontal line denotes S_a.

volution with the 2D Gaussian function with $\sigma = 0.3$ (see eq 6) of Panel g.

3.1. Uncorrelated Dihedral Angles. The effect of the Gaussian width σ on the 2D entropy S_{2D} is shown in Figure 2 for $N_r = 3$ and $N_r = 2$ as a function of the number of dihedral angles and in comparison to the analytical entropies S_a. σ is set to 0.3, 0.5, and 0.6. For a flip angle of $\Delta\varphi = 120°$ and $N_r = 3$ (panel a), the best agreement between S_{2D} and S_a is obtained for $\sigma = 0.5$, whereas for $N_r = 2$ (Panel b), the best agreement is obtained for $\sigma = 0.6$. The optimal value for σ shows a moderate dependence on the underlying rotameric jump model. The smaller the jump angle, the smaller is the optimal σ value because discrimination between the different rotameric states in the probability distribution $\tilde{P}_m(z)$ requires a narrower 2D Gaussian convolution function. The slight scatter in Figure 2 (as well as in Figure 3) is due to the random character of the first rotamer of each dihedral angle. A constant rotamer offset generally leads to smoother behavior.

3.2. Strongly Correlated Dihedral Angles. An essential criterion for the usefulness of an entropy estimator is that it faithfully takes into account the presence of correlations and anticorrelations between degrees of freedom. The behavior of S_{2D} was tested in this regard by generating ensembles with a reduced number of effective degrees of freedom by adding an increasing number of rotameric "mutual exclusivity constraints". Each such constraint precludes the simultaneous presence of two rotamers. For example, a constraint can impose that rotamer 1 of dihedral angle 5 is mutually exclusive with rotamer 3 of dihedral angle 7. Such constraints were randomly generated and successively applied to a 10-dihedral-angle ensemble with $N_r = 3$. The original ensemble consisting of 43 040 conformers that obey the excluded

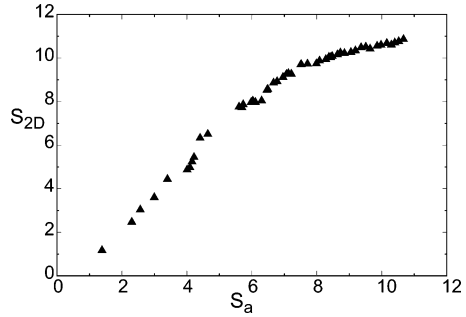

Figure 3. Reduction of entropies S_{2D} and S_a for an increasing number of mutually exclusive pairs of dihedral angles for a 10-dihedral-angle chain with $N_r = 3$. The total number of conformers is gradually reduced from 43 040 conformers in the absence of correlations (except for excluded volume effects) to 4 conformers upon introduction of an increasing number of pairwise correlations.

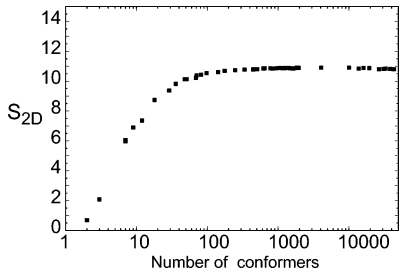

Figure 4. The effect of incomplete sampling on S_{2D} for a 10-dihedral-angle chain with $N_r = 3$. The initial ensemble of 43 161 allowed conformers was gradually reduced by excluding an increasing number of randomly selected conformers. S_{2D} with $\sigma = 0.5$ is plotted as a function of ensemble size N_c.

volume effect was thereby gradually reduced to an ensemble of only four conformers after adding up to 50 of these mutual exclusivity constraints. In Figure 3, for each ensemble, S_{2D} is plotted versus S_a. Overall, it shows a good equivalence between the two measures, although for small ensembles, S_{2D} tends to slightly overestimate the actual entropy. For ensembles constructed with other random sets of mutual exclusivity constraints, very similar relationships between S_a and S_{2D} are found.

3.3. Undersampling. In many practical (bio-)polymer applications, conformational space cannot be exhaustively searched, and a representative subset of conformational space is sampled instead. A good entropy estimator should allow extrapolation to the exact entropy from a relatively small subset of conformers. To test S_{2D} with respect to this property, a conformational ensemble is generated for the 10-dihedral-angle chain with $N_r = 3$. S_{2D} is then calculated for randomly chosen subsets of structures ranging between 2 and 43 161 conformers. A plot of S_{2D} versus the number of conformers N_c is given in Figure 4. It shows that S_{2D} converges rapidly toward the analytical entropy $S_a = 10.67$. A very good estimate of $S = 10.60$ is already obtained for 142 conformers, which accounts for less than 0.4% of all conformers. Since conformers are eliminated randomly, spurious correlations among dihedrals are mainly introduced in the limit of small numbers of conformers. This is in contrast to Figure 3 where significant correlations between

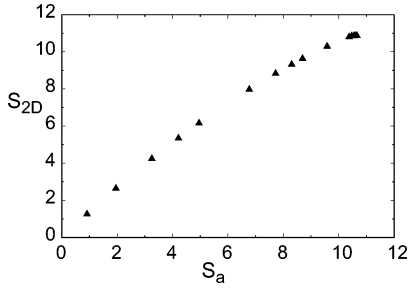

Figure 5. Correlation between S_{2D} and S_a for an ensemble of 10-dihedral-angle chains with $N_r = 3$ with conformer populations that are increasingly biased toward compact conformers as assessed by their radius of gyration: $p_k = (r_g - r_{g,min} + \Delta r_g)^{-1}$ (eq 11). Δr_g is varied between 10^{-6} and 1.0. The total number of allowed conformers is 43 082.

dihedral angles exist for any number of conformers. The convergence does not depend significantly on the value of σ for the range of interest here ($\sigma = 0.4-0.7$).

3.4. Soft Correlations: Radius of Gyration. In Figure 5, the behavior of S_{2D} is tested for ensembles whose conformer populations are biased toward conformers with a compact structure as reflected in a small radius of gyration. The conformer populations are given by

$$p_k = c(r_{g,k} - r_{g,min} + \Delta r_g)^{-1} \qquad (11)$$

where $r_{g,k}$ is the radius of gyration of conformer k computed as $r_{g,k}^2 = N_a^{-1} \sum_{j=1}^{N_a} \Delta r_{k,j}^2$, where $\Delta r_{k,j}$ is the distance of atom j to the center of mass of conformer k and c is a normalization constant. $r_{g,min}$ is the minimal radius of gyration of the ensemble, and Δr_g is an offset. The larger Δr_g, the more uniformly distributed are the probabilities, whereas for a small Δr_g, the most compact conformer dominates the ensemble.

In Figure 5, S_{2D} is compared with S_a for the 10-dihedral-angle conformational ensemble with $N_r = 3$ using the conformer probabilities of eq 11 with offset Δr_g varied between 10^{-6} (low entropy) and 1.0 (high entropy). The total number of conformers is 43 082, and the smallest and largest radii of gyration are 2.015 and 2.936, respectively. The good correlation between S_{2D} and S_a reflects the sensitive response of S_{2D} with respect to dihedral angle correlations underlying the global geometric property of compactness.

3.5. Soft Correlations: Gaussian Interaction Energies. A different method to introduce correlation effects between dihedral angles uses a pairwise energy potential function $E_{iu,jv}$, which denotes the energy between the uth rotamer of dihedral angle i and the vth rotamer of dihedral angle j. Using a Gaussian energy distribution

$$p(E_{iu,jv}) = \frac{1}{(2\pi\sigma_E^2)^{1/2}} e^{-(E_{iu,jv}-E_0)^2/(2\sigma_E^2)} \qquad (12)$$

where E_0 is an energy offset and σ_E is the standard deviation; the total energy of conformer k is given by $E_k = \sum_{i<j} \sum_{uv} E_{iu,jv}$, where the second sum goes over the rotamers occupied by conformer k; the relative population of a conformer k is given by $p_k = c \exp[-E_k/(k_B T)]$, where T is the absolute temperature. In Figure 6, S_{2D} is compared with S_a for the

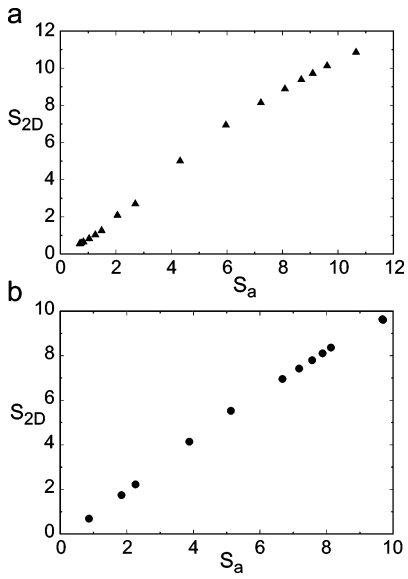

Figure 6. Comparison between S_{2D} and S_a for a Gaussian pairwise energy function with $k_B T$ varying between 0.1 and 1000. (Panel a) A 10-dihedral-angle chain with $N_r = 3$ and $\sigma = 0.5$. (Panel b) A 15-dihedral-angle chain with $N_r = 2$ and $\sigma = 0.6$.

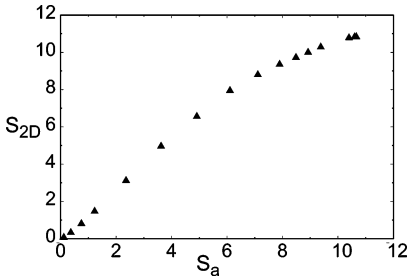

Figure 7. Comparison between S_{2D} and S_a for an ensemble of 10-dihedral-angle chains with $N_r = 3$ and $\sigma = 0.5$. Each atom carries a Coulomb charge of +1 or −1 with the total molecular charge being neutral. The entropies are calculated for Boltzmann distributions with different temperatures $k_B T$ ranging from 0.02 to 100.0.

10-dihedral-angle chain with $N_r = 3$ and $\sigma = 0.5$ (panel a) and the 15-dihedral-angle chain with and $N_r = 2$ and $\sigma = 0.6$ (panel b). For all calculations, E_0 is set to 10 and σ_E to 2, and $k_B T$ is varied from 0.1 (low entropy) to 1000 (high entropy). As can be seen from Figure 6, the correspondence between S_{2D} and S_a is good in both cases.

3.6. Soft Correlations: Electrostatic Energies. To test an alternative correlation mechanism, a distance-dependent interaction energy is used in the form of a Coulomb potential. For each atom, a charge of $q_i = 1$ or $q_i = -1$ was randomly assigned with the condition that the total charge of the molecule is zero. The corresponding Coulomb interaction takes the form $E = \sum_{i<j}(q_i q_j / r_{ij})$. The comparison between S_{2D} and S_a for an ensemble of 10-dihedral-angle chains with $N_r = 3$ and $\sigma = 0.5$ is depicted in Figure 7 with the temperature $k_B T$ varying between 0.02 (low entropy) and 100.0 (high entropy). Again, a good correspondence between the approximate and exact entropy measures can be seen, reflecting the sensitivity of S_{2D} to the correlation effects caused by the pairwise atomic Coulomb energy terms.

Table 1. S_{2D} of Ubiquitin in Units of J/mol/K

protein part	number of dihedrals	S_{2D} (correlated)	S_{2D} (uncorrelated)	ΔS_{2D}[a]
whole protein	312	483.01	561.08	78.07
backbone	151	129.73	134.65	4.91
side chains	161	372.63	426.44	53.81
backbone + side chains[b]	312	502.36[c]	561.09	58.72

[a] $\Delta S_{2D} = S_{2D}$ (uncorrelated) $- S_{2D}$ (correlated). [b] Arithmetic sums of backbone and side-chain entropies. [c] Includes all correlations, except correlations between backbone and side chains.

3.7. Application to Native Ubiquitin. The 2D entropy measure is applied to an ensemble of snapshots of the globular protein ubiquitin generated by molecular dynamics simulation. Despite the branched character of proteins, the entropy estimator is expected to deliver meaningful results with respect to the effect of dihedral angle correlations on the conformational entropy. The simulation was performed in a box with 2909 explicit water molecules at 300 K using CHARMM,[25] with details to be found elsewhere.[26,27] From a 5 ns trajectory, 1000 snapshots were extracted at a 5 ps time increment. The trajectory shows a stable behavior with a root-mean-square deviation of all heavy atoms (backbone and side chains) around 2 Å.[27] Since correlation times of some rotameric side-chain jump motions fall well into the nanosecond range, the total side-chain entropy is not fully converged for a simulation of this length. From each snapshot, all 151 mobile backbone φ, ψ dihedral angles were extracted, as well as the 161 mobile side-chain torsion angles, which amounts to a total of 312 dihedral angles. The 2D entropy is calculated as outlined in the Methods section using a standard deviation $\sigma = 0.5$ for the Gaussian convolution function. The results are summarized in Table 1. For the total conformational entropy S_{2D}, a value of 483.01 J/mol/K is obtained. The importance of correlation effects for S_{2D} is assessed by calculating S_{2D} after setting all off-diagonal elements to zero in covariance matrix **C**. This leads to an entropy increase of 78.07 J/mol/K. Backbone−backbone correlations contribute 4.91 J/mol/K, whereas side-chain−side-chain correlations contribute 53.81 J/ mol/K. The difference between 53.81 + 4.91 = 58.72 J/mol/K and 78.07 J/mol/K is −19.35 J/mol/ K, which reflects the entropy loss due to correlations between side-chain and backbone dihedral angles. As a consequence of dihedral angle correlations, the conformational entropy of the whole protein is reduced by 16.2%, for the backbone by 3.8% and for the side chains by 14.4%. Thus, motional side-chain−side-chain correlations are the dominant contributor to the conformational entropy loss in native ubiquitin.

4. Discussion

To gain useful insight into the thermodynamic properties of macromolecules from computer simulations (i) efficient sampling of conformational space and (ii) effective conversion of that information into thermodynamic quantities is required. The entropy measure, S_{2D}, introduced here represents a simple and robust estimator of the entropy associated with rotameric transitions of dihedral angles of an ensemble of conformers. Since dihedral angles are determined modulo 2π, there is an ambiguity in defining the dihedral angle average and its variance. This difficulty is avoided here by representing dihedral angles as complex numbers on the unit circle. Correlation effects between the dihedral angles are taken into account up to second order in terms of covariances and followed by a principal component analysis. Since this involves diagonalization of a complex $N_d \times N_d$ matrix, the method is efficient for systems with a small to moderately large number of dihedral angles. A continuous probability distribution is constructed for each eigenmode by convolution with a 2D Gaussian function with width σ. For an optimal choice of σ, information on the rotameric jump angles is required. This information can be obtained from the molecular force field or from the conformers themselves. The estimator is tested and calibrated on polymer chain models for which exact conformational entropies can be calculated for reference. The method provides good entropy estimates in the absence and presence of different types of correlation effects even when only a small fraction of all conformers is randomly sampled. S_{2D} focuses on the non-Gaussian dihedral angle distributions, reflecting primarily interconversion between different rotameric states. For these processes, the role of dihedral angle correlations is found in ubiquitin to be on the order of 16%. This contribution is dominated by motional correlations between side chains. Because of the finite width of σ and because in eq 9 the reference entropy S_{ref} is subtracted, dihedral angle variations that are significantly smaller than σ are not manifested in S_{2D}. σ can be viewed as a measure for the intrinsic structural uncertainty of a single conformer and thereby acts as a motional filter for the entropy evaluation: high-frequency vibrations and other small amplitude motions are not included in S_{2D} because their fluctuations are typically well-below the $\sigma = 0.5$ threshold. Such contributions can be evaluated using a normal-mode analysis[28−30] or quasiharmonic analysis applied to segments of MD or MC trajectories.[3−16]

Acknowledgment. This work was supported by NSF Grant MCB-0211512.

References

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

(2) Hagler, A. T.; Stern, P. S.; Sharon, R.; Becker, J. M.; Naider, F. *J. Am. Chem. Soc.* **1979**, *101*, 6842−6852.

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

(4) Cheatham, T. E.; Srinivasan, J.; Case, D. A.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1998**, *16*, 265−280.

(5) Wrabl, J. O.; Shortle, D.; Woolf, T. B. *Proteins* **2000**, *38*, 123−133.

(6) Kuhn, B.; Kollman, P. A. *J. Med. Chem.* **2000**, *43*, 3786−3791.

(7) Schäfer, H.; Smith, L. J.; Mark, A. E.; van Gunsteren, W. F. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 215−224.

(8) Gohlke, H.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 238−250.

(9) Cheluvaraja, S.; Meirovitch, H. *J. Chem. Phys.* **2005**, 122, 054903.

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

(11) Karplus, M.; Ichiye, T.; Pettitt, B. M. *Biophys. J.* **1987**, *52*, 1083−1085.

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

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

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

(15) Schäfer, H.; Daura, X.; Mark, A. E.; van Gunsteren, W. F. *Proteins* **2001**, *43*, 45−56.

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

(17) Carlsson, J.; Aqvist, J. *J. Phys. Chem. B* **2005**, 109, 6448−6456.

(18) Prompers, J. J.; Brüschweiler, R. *J. Phys. Chem. B* **2000**, *104*, 11416−11424.

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

(20) Edholm, O.; Berendsen, H. J. C. *Mol. Phys.* **1984**, *51*, 1011−1028.

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

(22) Hnizdo, V.; Fedorowicz, A.; Sing, H.; Demchuk, E. *J. Comput. Chem.* **2003**, *24*, 1172−1183.

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

(24) In the quasiharmonic approximation,[3,10] for example, the δ peak distribution is "smoothed" by approximating it by a multivariate Gaussian function for which the entropy can be calculated analytically.

(25) Brooks, R. B.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(26) Lienin, S. F.; Bremi, T.; Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1998**, *120*, 9870−9879.

(27) Prompers, J. J.; Scheurer, C.; Brüschweiler, R. *J. Mol. Biol.* **2001**, *305*, 1085−1097.

(28) Noguti, T.; Go, N. *J. Phys. Soc. Jpn.* **1983**, *52*, 3283−3288.

(29) Brooks, B. R.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 6571−6575.

(30) Levitt, M.; Sander, C.; Stern, P. S. *J. Mol. Biol.* **1985**, *181*, 423−447.

CT050118B