

JCTC

Journal of Chemical Theory and Computation

Absolute Single-Molecule Entropies from Quasi-Harmonic Analysis of Microsecond Molecular Dynamics: Correction Terms and Convergence Properties

Riccardo Baron,^{*,†,‡} Philippe H. Hünenberger,[§] and J. Andrew McCammon^{†,‡,||,⊥}

Department of Chemistry & Biochemistry, Center for Theoretical Biological Physics, Department of Pharmacology, and Howard Hughes Medical Institute, University of California San Diego, La Jolla, California 92093-0365 and Laboratory of Physical Chemistry, ETHZ, ETH Hönggerberg, CH 8093, Zürich, Switzerland

Received July 18, 2009

Abstract: The convergence properties of the absolute single-molecule configurational entropy and the correction terms used to estimate it are investigated using microsecond molecular dynamics simulation of a peptide test system and an improved methodology. The results are compared with previous applications for systems of diverse chemical nature. It is shown that (i) the effect of anharmonicity is small, (ii) the effect of pairwise correlation is typically large, and (iii) the latter affects to a larger extent the entropy estimate of thermodynamic states characterized by a higher motional correlation. The causes of such deviations from a quasi-harmonic behavior are explained. This improved approach provides entropies also for molecular systems undergoing conformational transitions and characterized by highly frustrated energy surfaces, thus not limited to systems sampling a single quasi-harmonic basin. Overall, this study emphasizes the need for extensive phase-space sampling in order to obtain a reliable estimation of entropic contributions.

1. Introduction

Entropy is a key property to understand a wide variety of physical, chemical, and biochemical phenomena. However, the estimation of absolute entropies and entropy differences from computer simulations is a long-standing problem^{1–9} and one of the current challenges in computational chemistry.^{10–15}

The calculation of reliable absolute entropies from molecular dynamics (MD) simulations is intrinsically difficult

* Corresponding author phone: +1-858-534-2913; e-mail: rbaron@mccammon.ucsd.edu.

† Department of Chemistry & Biochemistry, University of California San Diego.

‡ Center for Theoretical Biological Physics, University of California San Diego.

§ ETH Hönggerberg.

|| Department of Pharmacology, University of California San Diego.

⊥ Howard Hughes Medical Institute, University of California San Diego.

because the absolute entropy is a measure of the overall extent of phase space (PS) *accessible* to a molecular system. However, absolute single-molecule entropies can be estimated based on an analytical approximation to the configurational probability distribution corresponding to the PS *accessed* by a simulated system.² The underlying theory, assumptions, approximations, and alternative practical implementations have been recently reviewed.^{10,11} The relationship among quasi-harmonic (QH), essential-mode, and normal-mode analyses has also been investigated.¹¹ For an extensive review of the subject, not limited to the QH approach, see also refs 11 and 16–19 and references therein.

The difference between the *true* entropy of a simulated system and its QH estimate arises from (i) *anharmonicities* (i.e., non-Gaussian behavior) in the probability distributions along individual QH modes and (ii) *correlations* among the probability distributions associated with different QH modes (beyond the pairwise linear correlations accounted for). These effects are neglected in standard QH analysis^{10,11} and (nearly)

always lead to a negative entropy contribution.¹¹ A method to correct for both artifacts was recently described.¹¹ Point ii is of particular relevance when trying to estimate entropy differences between two conformational states of a molecular system because error cancellation cannot be guaranteed a priori.^{11,20} By taking into account correlation effects of increasing order, entropy estimates based on corrected QH analysis aim at capturing the entropy corresponding to the entire PS sampled (see Figure 1 in ref 11). Thus, this approach is not limited to systems sampling single QH basins and allows capturing conformational transitions.

In the present article, we expand the previous study in ref 11. A general formulation is proposed to account for correction terms of increasing order, and its practical implementation and limitations are discussed. We review previous studies employing this novel approach on an array of (bio)molecular systems providing a solid basis for its application and demonstrating the importance of these correction terms in the evaluation of absolute entropy and entropy differences. Using microsecond MD simulation of a test system, we analyze the convergence properties of the absolute single-molecule entropy and of the correction terms used to estimate it. The results emphasize that sufficient PS sampling is required for a reliable estimation of entropic contributions because convergence of both the QH upper bound and the required correction terms should be achieved.

2. Methods

2.1. QH Analysis. QH analysis aims to account for motions in the overall extent of PS *accessible* to a molecular system at thermodynamic equilibrium. It relies on approximating the configurational probability distribution as a multivariate Gaussian, the momenta of which can be estimated, e.g., from molecular dynamics (MD) or Monte Carlo simulations.

More precisely, for a given choice of generalized coordinate system \mathbf{q} (of dimension $M' = 3N$, N being the number of atoms), its input quantity is the covariance matrix \mathbf{C}_q characterizing the atom-positional fluctuations (and their correlations) around an average configuration $\bar{\mathbf{q}}$. Assuming a canonical ensemble and fluctuations resulting from an underlying harmonic potential of the form

$$\tilde{\mathcal{F}}_h(\mathbf{q}) = \frac{1}{2}(\mathbf{q} - \bar{\mathbf{q}}_o)^T \tilde{\mathbf{H}}_q (\mathbf{q} - \bar{\mathbf{q}}_o) \quad (1)$$

where $\tilde{\mathbf{H}}_q$ is an effective Hessian matrix and $\bar{\mathbf{q}}_o$ an effective equilibrium configuration, it follows that¹¹

$$\bar{\mathbf{q}}_o = \bar{\mathbf{q}} \text{ and } \tilde{\mathbf{H}}_q = \beta^{-1} \mathbf{C}_q^{-1} \quad (2)$$

Note that the corresponding harmonic model only strictly produces the correct average configuration $\bar{\mathbf{q}}$ and covariance matrix \mathbf{C}_q for generalized coordinate systems where the mass-metric tensor \mathbf{A}_q is configuration independent.¹¹

In this study, we only consider the specific case of single-molecule entropy (i.e., the entropy of individual distinguishable atoms in a covalently bound molecule) based on MD simulation trajectories. As detailed elsewhere,²¹ *single-molecule entropy* differs from *molecular entropy* in that the

former estimate only accounts for intermolecular correlation in terms of the effect of the solvent on the single-molecule dynamics.

In practice, the QH analysis of an MD trajectory involves the following steps.¹¹

First, the average configuration $\bar{\mathbf{q}}$ and the covariance matrix \mathbf{C}_q in the chosen coordinate system are evaluated as

$$\bar{\mathbf{q}} = \langle \mathbf{q} \rangle \text{ and } \mathbf{C}_q = \langle (\mathbf{q} - \bar{\mathbf{q}}) \otimes (\mathbf{q} - \bar{\mathbf{q}}) \rangle \quad (3)$$

The equilibrium configuration $\bar{\mathbf{q}}_o$ and Hessian matrix $\tilde{\mathbf{H}}_q$ of the effective underlying harmonic model are then defined according to eq 2.

Second, the (symmetric) metric-tensor-weighted covariance matrix is diagonalized

$$\mathbf{V}_q^T \mathbf{A}_q^{1/2} \mathbf{C}_q \mathbf{A}_q^{1/2} \mathbf{V}_q = \mathbf{F}_q \quad (4)$$

where \mathbf{V}_q is a $M \times M$ -dimensional (orthogonal) matrix the columns of which represent the M' components of the eigenvectors $\{\mathbf{v}_{q,m} | m = 1, \dots, M'\}$ (called QH modes) of the metric-tensor-weighted covariance matrix and \mathbf{F}_q is a diagonal matrix containing the corresponding eigenvalues. These eigenvalues are related to the associated angular frequencies of the underlying effective harmonic model as (see eqs 2 and 4)

$$\omega_m = (\beta F_{q,m})^{-1/2}, m = 1, 2, \dots, M \quad (5)$$

The sum of the eigenvalues in \mathbf{F}_q is equal to the total mean-square metric-tensor-weighted fluctuation of the system, i.e.

$$\begin{aligned} \text{Tr}[\mathbf{F}_q] &= \text{Tr}[\mathbf{A}_q^{1/2} \mathbf{C}_q \mathbf{A}_q^{1/2}] \\ &= \langle [\mathbf{A}_q^{1/2}(\mathbf{q} - \bar{\mathbf{q}})] \cdot [\mathbf{A}_q^{1/2}(\mathbf{q} - \bar{\mathbf{q}})] \rangle \end{aligned} \quad (6)$$

so that the eigenvalues can be interpreted as contributions of individual QH modes to this quantity (a larger value indicating a larger contribution to the total fluctuation of the molecule).

Third, the simulated trajectory is projected onto the QH modes, i.e., one considers the transformed coordinates \mathbf{b}_q defined as

$$\mathbf{b}_q = \mathbf{V}_q^T \mathbf{A}_q^{1/2} (\mathbf{q} - \bar{\mathbf{q}}) \quad (7)$$

These so-called QH coordinates satisfy the properties¹¹

$$\langle \mathbf{b}_q \rangle = \mathbf{0} \text{ and } \langle \mathbf{b}_q \otimes \mathbf{b}_q \rangle = \mathbf{V}_q^T \mathbf{A}_q^{1/2} \mathbf{C}_q \mathbf{A}_q^{1/2} \mathbf{V}_q = \mathbf{F}_q \quad (8)$$

Because \mathbf{F}_q is diagonal, eq 8 enforces that the individual components $\{\mathbf{b}_{q,m} | m = 1, \dots, M'\}$ of the QH coordinates \mathbf{b}_q are pairwise linearly uncorrelated, which, however, does not imply the absence of higher order (i.e., pairwise supralinear and higher order) correlations.

We previously motivated the choice of a Cartesian vs internal coordinate system.¹¹ If a Cartesian coordinate system \mathbf{r} is employed^{6,8} (after removal of the overall translational and rotational motion from the sampled trajectory²²), the mass-metric tensor \mathbf{A}_r is identical to the mass matrix \mathbf{M} (thus configuration independent, so that eq 2 is exactly satisfied).

In this case, the QH analysis relies on the diagonalization of the mass-weighted Cartesian covariance matrix, i.e.

$$\mathbf{D}_r = \mathbf{M}^{1/2} \mathbf{C}_r \mathbf{M}^{1/2} \quad (9)$$

in place of $\mathbf{A}_q^{1/2} \mathbf{C}_q \mathbf{A}_q^{1/2}$ in eq 4.

In the absence of geometric constraints, the corresponding eigenvalue matrix \mathbf{E}_r contains $3N - 6$ nonzero and 6 vanishing elements. If N_c geometrical constraints are present in the system (e.g., bond-length constraints), these will map to an identical number of zero eigenvalues (see Appendix A in ref 23 for a derivation in the mathematically similar context of essential-mode analysis). Thus, the number of QH modes with nonzero eigenvalues is $M = 3N - N_c - 6$, where $M' = 3N$. When using a generalized coordinate system excluding overall translation and rotation variables, one has $M' = M = 3N - N_c - 6$. Note that the QH coordinates have units of $\text{mass}^{1/2} \times \text{length}$.¹¹

2.2. Entropies and Correction Terms. Single-molecule entropies can be obtained as follows.¹¹ In terms of QH coordinates, the configurational probability distribution associated with the effective harmonic model of eq 2 corresponds to that of M independent harmonic oscillators. Thus, the associated entropy S_o can be calculated analytically. Assuming a canonical ensemble and a configuration-independent mass-weighted metric tensor, this leads to¹¹

$$S_o = \sum_{m=1}^M s(\omega) \cdot ((\beta E_{r,m})^{-1/2}) \quad (10)$$

where $s(\omega)$ is the canonical entropy of a one-dimensional harmonic oscillator with angular frequency ω . The classical expression $s_{cl,o}(\omega)$ and the quantum-mechanical expression $s_{qm,o}(\omega)$ for this quantity are

$$s_{cl,o}(\omega) = k_B(1 - \ln \beta \hbar \omega) \quad (11)$$

and

$$s_{qm,o}(\omega) = k_B \left[\frac{\beta \hbar \omega}{e^{\beta \hbar \omega} - 1} - \ln(1 - e^{-\beta \hbar \omega}) \right] \quad (12)$$

where $\hbar = h(2\pi)^{-1}$ is the reduced Planck's constant, leading to eq 10 to corresponding total estimates $S_{cl,o}$ and $S_{qm,o}$, respectively.

In practice, even if the underlying trajectory was generated at the classical level, the QH entropy must be evaluated using the quantum-mechanical oscillator formula because in the high-frequency limit the classical entropy of a one-dimensional harmonic oscillator diverges to the unphysical limit of $-\infty$ rather than to the physical limit of zero.^{8,11} However, the QH entropy estimate $S_{qm,o}$ is not the absolute configurational entropy of a single molecule but an upper bound for this quantity due to the presence of QH mode anharmonicities and correlations not accounted for in the effective harmonic model of eq 2. Corresponding correction terms can be formulated exactly at the classical level using an approach previously described¹¹ and briefly summarized below.

In the canonical ensemble, assuming a configuration-independent mass-metric tensor, the exact classical single-molecule entropy reads¹¹

$$S_{cl} = -k_B \left[\frac{M}{2} \left(1 - \ln \frac{\beta \hbar^2}{2\pi} \right) - \int d\mathbf{b}_q p(\mathbf{b}_q) \ln p(\mathbf{b}_q) \right] \quad (13)$$

where $p(\mathbf{b}_q)$ is the probability distribution in the M -dimensional space of the QH coordinates \mathbf{b}_q (eq 7). This expression can be compared with the approximate (classical) QH estimate $S_{cl,o}$ based on eqs 10 and 11, i.e.

$$S_{cl,o} = -k_B \sum_{m=1}^M \left(1 - \frac{1}{2} \ln \frac{\beta \hbar^2}{F_{q,m}} \right) \quad (14)$$

A series of increasingly accurate estimates $\{S_{cl,K} | K = 0, 1, \dots, M\}$ may now be formulated as

$$S_{cl,K} = S_{cl,o} - k_B \left[\frac{K}{2M} C(K, M) \sum_{m=1}^M (1 + \ln 2\pi F_{q,m}) + \sum_{c=1}^{C(K,M)} \int d\mathbf{b}_q^{(c)} p^{(c)}(\mathbf{b}_q^{(c)}) \ln p^{(c)}(\mathbf{b}_q^{(c)}) \right] \quad (15)$$

where c denotes a combination of K QH modes, $C(K, M) = [(M - K)!K!]/M!$ for $K > 0$ along with $C(0, M) = 0$ represents the total number of possible combinations c of K modes among the M QH modes and $p^{(c)}(\mathbf{b}_q^{(c)})$ is the K -dimensional probability distribution in the subspace of the QH coordinates $\mathbf{b}_q^{(c)}$ within \mathbf{b}_q that are involved in a combination c . The derivation of this equation is given in the Appendix in the Supporting Information.

It is easily verified that $S_{cl,K=0} = S_{cl,o}$ (eq 14, i.e., the uncorrected classical QH entropy) and $S_{cl,K=M} = S_{cl}$ (eq 13, i.e., the exact classical entropy). Substituting the classical estimate $S_{cl,o}$ by the corresponding quantum-mechanical estimate $S_{qm,o}$ (eqs 10 with 12) into eq 15 and introducing successive correction terms defined as

$$\Delta S_{cl,K} = S_{cl,K} - S_{cl,(K-1)} \quad (16)$$

leads to a (classically) corrected QH entropy estimate

$$\tilde{S}^{ctd} = S_{qm,o} + \sum_{K=1}^M \Delta S_{cl,K} \quad (17)$$

The successive correction terms of eq 17 involve integrals over the probability distributions $p^{(c)}(\mathbf{b}_q^{(c)})$ in eq 15 with increasing dimensionality K . Note that these terms are all individually negative (or vanishing). The first correction term $\Delta S_{cl,1}$ involves one-dimensional (1D) integrals and accounts for anharmonicities in the individual QH modes. The second correction term $\Delta S_{cl,2}$ involves two-dimensional (2D) integrals and accounts for pairwise (supralinear) correlations between the QH modes. For simplicity, these two terms will be renamed ΔS_{cl}^{ah} and ΔS_{cl}^{pc} , respectively, to match the notation used in other studies.^{11,20,24-28}

The following higher order correction terms account for correlations among QH modes beyond the pairwise ones. Although the classical QH entropy estimate $S_{cl,o}$ usually represents a poor approximation to its quantum-mechanical

counterpart $S_{qm,o}$, the evaluation of the correction terms at the classical level remains accurate because anharmonicities and correlations principally affect the low-frequency QH modes for which the classical approximation holds.¹¹

The successive correction terms in the series of eq 17 are increasingly difficult to evaluate because both (i) the number of terms $C(K,M)$ involved in the evaluation of $\Delta S_{cl,K}$ and (ii) the sparseness in the required multiple-mode probability distributions $p^{(c)}(\mathbf{b}_q^{(c)})$ increase exponentially with K . For this reason, their evaluation is restricted in practice to the first two terms and implies an intrinsic uncertainty on the final estimate compared to the *true* single-molecule entropy (i.e., persisting in the limit of infinite sampling).

However, note that, in a different context, alternative approximate formulations to estimate terms of increasing order mutual information have been proposed and seem to suggest that the first two correction terms in eq 17 are indeed dominant.^{29–31} No study heretofore investigated the convergence properties of these terms along a simulation trajectory.

Following from eqs 16 and 17, the expressions for the first two correction terms are

$$\Delta S_{cl}^{ah} = -k_B \left[\frac{1}{2} \sum_{m=1}^M (1 + \ln 2\pi F_{q,m}) + \sum_{m=1}^M \int db_{q,m} p^{(m)}(b_{q,m}) \ln p^{(m)}(b_{q,m}) \right] \quad (18)$$

and

$$\Delta S_{cl}^{pc} = -k_B \left[\frac{M-1}{2} \sum_{m=1}^M (1 + \ln 2\pi F_{q,m}) + \sum_{m=1}^M \sum_{n=m+1}^M \iint db_{q,m} db_{q,n} p^{(m,n)}(b_{q,m}, b_{q,n}) \ln p^{(m,n)}(b_{q,m}, b_{q,n}) \right] - \Delta S_{cl}^{ah} \quad (19)$$

leading to the corrected absolute single-molecule entropy estimate

$$S^{ctd} = S_{qm,o} + \Delta S_{cl}^{crc} = S_{qm,o} + \Delta S_{cl}^{ah} + \Delta S_{cl}^{pc} \quad (20)$$

The relative magnitudes f_{cl}^{crc} , f_{cl}^{ah} , and f_{cl}^{pc} of the correction terms ΔS_{cl}^{crc} , ΔS_{cl}^{ah} , and ΔS_{cl}^{pc} with respect to the QH entropy upper-bound $S_{qm,o}$ (expressed in percent), i.e.

$$f_{cl}^{crc} = \frac{100\Delta S_{cl}^{crc}}{S_{qm,o}}, f_{cl}^{ah} = \frac{100\Delta S_{cl}^{ah}}{S_{qm,o}} \text{ and } f_{cl}^{pc} = \frac{100\Delta S_{cl}^{pc}}{S_{qm,o}} \quad (21)$$

may then serve as a measure for the importance of the aforementioned corrections.

In practice, the 1D and 2D integrals involved in eqs 18 and 19 are evaluated numerically in the form of sums over corresponding histograms. It is reasonable to choose the bin width along a given QH mode in proportion to the width (first moment) of the probability distribution along this mode with proportionality factors κ_1 and κ_2 for 1D and 2D integrals, respectively. However, κ_1 and κ_2 values must be selected carefully in order to keep both finite-sampling and binning errors to a minimum, i.e., to ensure the independence of the

results on these two parameters.¹¹ For this reason, we monitored the dependence of such numerical integrals on the width of histogram bins for increasing periods of time, as described in section 3.4.

Note, finally, that the absolute single-molecule entropies so far discussed exclude roto-translational contributions. In principle, a translational entropy contribution can be included using the quantum-mechanical expression of the Sackur–Tetrode equation for a specified standard state of the pressure (molecule in the gas phase) or of the concentration (molecule in solution). Similarly, the rotational entropy contribution could be included using the appropriate quantum-mechanical expression (e.g., rigid-rotor approximation, based on the average inertia tensor of the molecule^{32–34}). However, these two contributions are likely to be highly coupled with each other and with \tilde{S}^{ctd} , i.e., they are not strictly additive, and their rigorous treatment is therefore still challenging. A recent study reported on relatively small effects of motional correlation on changes of reorientational entropy using selected QH modes from a 1.5 ns simulation of the ubiquitin protein.³⁵ In the present article, single-molecule configurational entropies refer to entropies excluding roto-translational effects.

2.3. Computational Details. A 1.1 μ s long MD simulation of the cc β peptide (CH₃-CO-S-I-R-E-L-E-A-R-I-R-E-L-E-L-R-I-COO⁻) at 300 K was performed with the AMBER 9 software,³⁶ the AMBER 99SB parameter set,³⁷ and the compatible TIP3P water model.³⁸ The simulation was initialized from the α -helical configuration based on a X-ray model structure of the cc β coiled coil (PDB ID 1s9z).³⁹ Trajectory snapshots were saved every 10 ps for analysis. The simulation setup and trajectory analyses are detailed elsewhere.⁴⁰ Backbone atom-positional root-mean-square deviations (RMSD) from the initial folded structure and radius of gyration (RGYR) were calculated using all C $^{\alpha}$ atoms.

Independent QH analyses were performed for 22 increasingly long segments of the simulation (differing in length by 50 ns) by calculation of the solute all-atom mass-weighted covariance matrix \mathbf{D}_r (eq 9) in Cartesian coordinates after least-squares fit superposition²² of all configurations onto the initial structure to eliminate overall translation and rotation and diagonalization (eq 4 with $\mathbf{A}_q^{1/2} \mathbf{C}_q \mathbf{A}_q^{1/2} = \mathbf{D}_r$). A total of 534 ($M = 3 \times 297 - 351 - 6$) modes associated with nonvanishing eigenvalues were considered. After determination of the QH modes (columns of the matrix \mathbf{V}_r in eq 4; sorted in order of decreasing eigenvalues, i.e., increasing ω_m frequency in eq 5), the trajectory was projected in this basis set to obtain the time series of the corresponding QH coordinates \mathbf{b}_r (eq 7). This first part of the analysis was performed using the S_correction program as implemented in the gromos++ module of the GROMOS05 software⁴¹ for biomolecular simulation.

The QH entropy upper bound, $S_{qm,o}$ (eq 10 with eq 12), the corrections for mode anharmonicity, ΔS_{cl}^{ah} (eq 18), and pairwise supralinear mode correlation, ΔS_{cl}^{pc} (eq 19), their sum, ΔS_{cl}^{crc} (eq 20), the improved absolute single-molecule entropy, S^{ctd} (eq 20), the relative terms f_{cl}^{crc} , f_{cl}^{ah} , and f_{cl}^{pc} (eq

21), and the sum of the eigenvalues, $Tr[\mathbf{E}_r]$ (eq 6), were then calculated for each of the 22 trajectory segments.

Note that this analysis is computationally intensive because, as discussed in section 3.3, each of the 22 ΔS_{cl}^{ah} values requires the estimation of 534 1D integrals, while each of the 22 ΔS_{cl}^{pc} values requires the estimation of 142 311 2D integrals (eq 15). In addition, for each of these integrals the optimized proportionality factors κ_1 and κ_2 were determined based on multiple integral calculations for an accurate numerical integration (see section 3.4). As an indication of the actual computational cost, using an Intel Xeon X5450 3.0 GHz, dedicated software, and the above procedure, each 1D or 2D integral can be estimated with an average CPU time of 0.07 or 0.21 s, respectively, from 50 ns trajectory windows. Overall, the analyses presented in this work require a CPU time that sums up to ~ 6 months.

3. Results and Discussion

3.1. Review of Previous Studies. The key findings of previous studies concerning the uncorrected QH upper bound, $S_{qm,o}$ (eq 10 with eq 12), the improved absolute single-molecule entropy S^{ctd} (eq 20), and the relative magnitude of the cumulative correction term f_{cl}^{cfc} (eq 21) are summarized graphically in Figure 1.

These results span systems with different chemical nature: 2 β -peptides in methanol,¹¹ the 11 disaccharides of glucose in water,²⁸ the dipalmitoylphosphatidylcholine (DPPC) lipid in a hydrated bilayer,²⁴ the W191G mutant cavity and its gating loop within cytochrome *c* peroxidase in water,²⁰ the H-ras lipopeptide anchor in water or inserted into a model lipid membrane,²⁷ and the *cc\beta*-peptide in water (this study). In some cases, the QH analysis was also performed separately for different chemical environments or conformational states of the molecule, which permits estimating relative entropies, thereby quantifying the impact of the correction terms on the thermodynamic process of interest. These processes include reversible peptide folding (ref 11 and this study), conformational changes in carbohydrates²⁸ and lipids,²⁴ lipopeptide insertion in a model membrane bilayer,²⁷ ligand binding to a protein cavity,²⁰ and protein-loop gating.²⁰ The results presented in Figure 1 are scaled by the number, N , of atoms to allow for a comparison among molecules of varying size (the raw data is available as Supporting Information, Table S1).

Some clear qualitative trends are evident, although a direct comparison among these studies is not possible due to the different MD time scales and physicochemical conditions. In all systems the cumulative correction term ΔS_{cl}^{cfc} (eq 20) is generally sizable, demonstrating an overall large deviation from a QH behavior as evaluated up to the pairwise supralinear level. The corresponding relative magnitudes, f_{cl}^{cfc} , display values from 9% to 73% of the QH upper-bound value $S_{qm,o}$ (Figure 1). In detail, these important cumulative terms result from the sum of correction terms for mode anharmonicity (ΔS_{cl}^{ah} ; eq 18) that are always relatively small (up to 3% of the upper-bound value $S_{qm,o}$) and for pairwise supralinear mode correlation (ΔS_{cl}^{pc} ; eq 19) that are always dominant. The latter correction term has a magnitude that

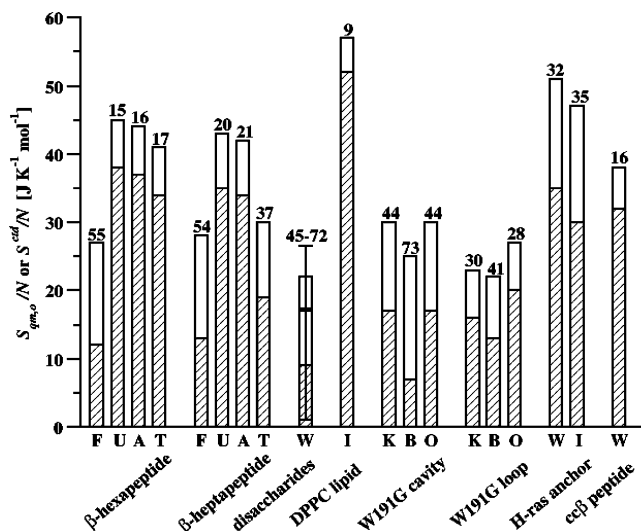


Figure 1. Summary of previous studies investigating the improved absolute single-molecule entropy. The QH entropy upper-bound $S_{qm,o}$ (eq 10 with eq 12; empty bars) and the improved entropy estimate S^{ctd} (eq 20; hatched areas) are displayed scaled by the number N of system particles. The relative (%) values of the cumulative correction term f_{cl}^{cfc} (eq 21; bar labels) are reported as a measure of the importance of the deviation from the QH approximation. From left to right: two β -peptides in methanol at high temperature (F, folded; U, unfolded; A, all; T, 298 instead of 340 K),¹¹ the 11 glucose-based disaccharides (W, free in water),²⁸ dipalmitoylphosphatidylcholine, DPPC (I, inserted in a hydrated bilayer),²⁴ the cavity and its gating loop of the W191G mutant of cytochrome *c* peroxidase (K, bound to a K⁺ ion with closed gating loop; B, bound to 2-amino-5-methylthiazole with closed gating loop; O, bound to a K⁺ ion with open gating loop),²⁰ the H-ras lipopeptide anchor,²⁷ and the *cc\beta*-peptide (this study). For the disaccharides,²⁸ corresponding mean entropy values are displayed (a vertical bar represents the range of values). See Supporting Information for details.

depends on the physical nature of the molecular motional correlation experienced by the molecular system in a given thermodynamic state.

The largest relative corrections, f_{cl}^{cfc} , are expected and found for intrinsically more ordered systems (Figure 1). This can be explained by considering that restricted flexibility is typically promoted by inter- and/or intramolecular interactions, simultaneously inducing increased motional correlation. For example, the ligand-bound state of the W191G protein cavity²⁰ displays the largest f_{cl}^{cfc} value (73%), i.e., the thermodynamic ensemble involving the largest motional correlations and lowest entropy content among those studied. On the other end of the spectrum and in line with this qualitative picture, the smallest f_{cl}^{cfc} values were reported for the DPPC lipid in a bilayer (9%),²⁴ i.e., the ensemble characterized by the highest molecular flexibility and thus the lowest motional correlations. Interestingly, the 11 disaccharides of glucose in water²⁸ display high variability and always large f_{cl}^{cfc} values (45–72%). This behavior can be explained considering that these molecules involve a reduced number of degrees of freedom overall and the linkage between rather stiff glucose rings is the torsion defining major conformational changes.^{25,26}

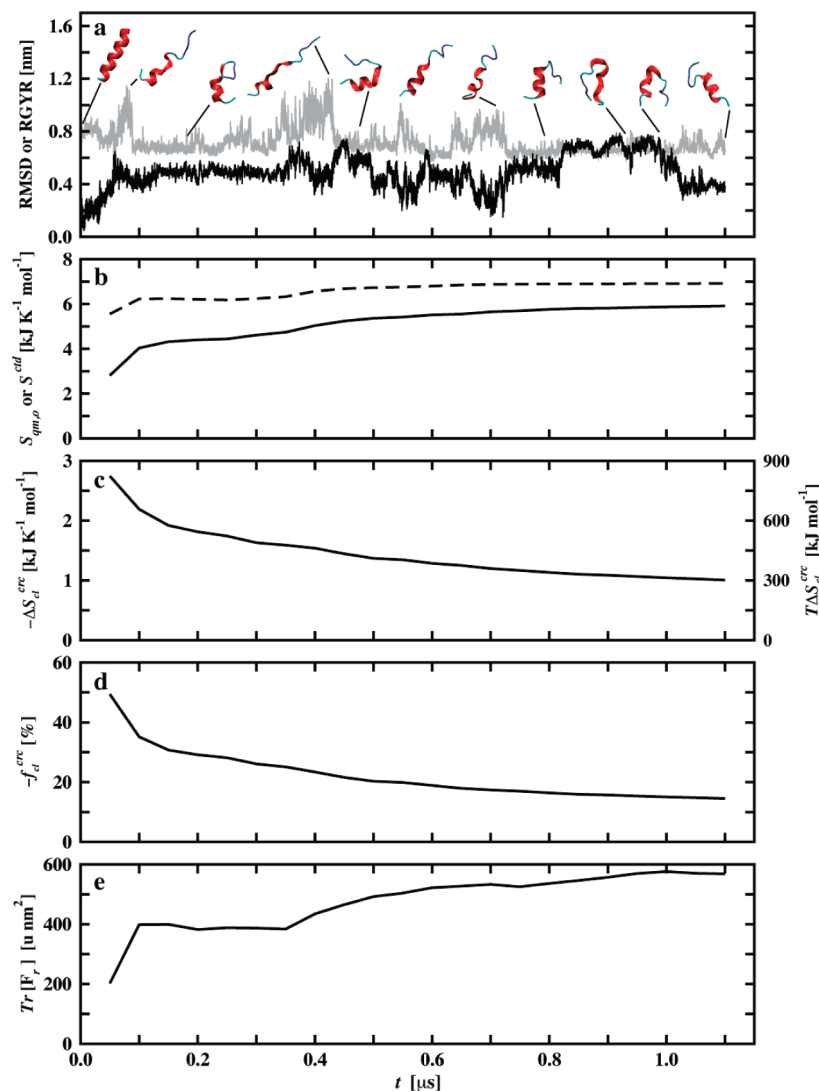


Figure 2. $cc\beta$ peptide dynamics on the microseconds time scale and entropy convergence. (a) The backbone atom-positional root-mean-square deviation (rmsd; black) from the initial helical fold and of the backbone radius of gyration (RGYR; gray) are shown along the time, t . The cartoon representations highlight example configurations (oriented with the $\text{CH}_3\text{-CO}$ terminus down). (b) Build-up curves of the QH entropy upper bound $S_{qm,o}$ (eqs 10 with 12; dashed line) and of the improved absolute single-molecule entropy S^{ctd} (eq 20; solid line). Convergence of (c) the cumulative correction term $\Delta S_d^{c,c}$ (eq 20) and its contribution to the free energy $T\Delta S_d^{c,c}$, (d) its relative value $f_d^{c,c}$ (eq 21), and (e) the sum of the eigenvalues $Tr[\mathbf{E}_r]$ (eq 6).

These qualitative trends are also in agreement with the observation that entropy is the measure of PS sampling for a molecular system. The QH upper bound, $S_{qm,o}$, and the improved absolute single-molecule entropy, S^{ctd} , are estimated based on the PS that has been *accessed* during a MD simulation of finite time scale, i.e., only a fraction of the PS *accessible* to the system. These time scales ranged from 50 ns (W191G mutant²⁰ showing the lowest entropy) to 25.6 μs (concatenated trajectory of the DPPC lipid,²⁴ showing the largest entropy). However, this hampers a quantitative comparison of $S_{qm,o}$, S^{ctd} , and $\Delta S_{cl}^{c,c}$ values among previous studies.

Prompted by these observations, the dependence of these quantities on the extent of accessed PS was assessed on the microsecond time scale for a peptide test system.

3.2. Convergence of the QH Analysis. The $cc\beta$ peptide in water was chosen as a test system to investigate entropy convergence properties because of its small size and broad

PS accessibility. Figure 2a shows the time series of the backbone atom-positional root-mean-square deviation (RMSD) from the folded structure and of the backbone radius of gyration (RGYR) along a 1.1 μs of MD simulation. The peptide undergoes several reversible folding/unfolding events and samples a variety of unfolded configurations and compact folds.⁴⁰

The probability distributions $p^{(m)}(\mathbf{b}_{r,m})$ of the transformed QH coordinates \mathbf{b}_r (eq 7) along selected QH modes ($m = 1, 2, 6, 10, 50,$ and 500) are shown in Figure 3a. The reference Gaussian functions with identical variances and vanishing averages are also represented. The actual distributions become increasingly narrow and similar to the Gaussian functions for higher m indices, i.e., the corresponding QH modes become increasingly stiff and harmonic. However, the distributions along the lowest frequency modes (e.g., Figure 3a, $m = 1$ or 2) differ significantly from Gaussian functions and evidently result from the superposition of

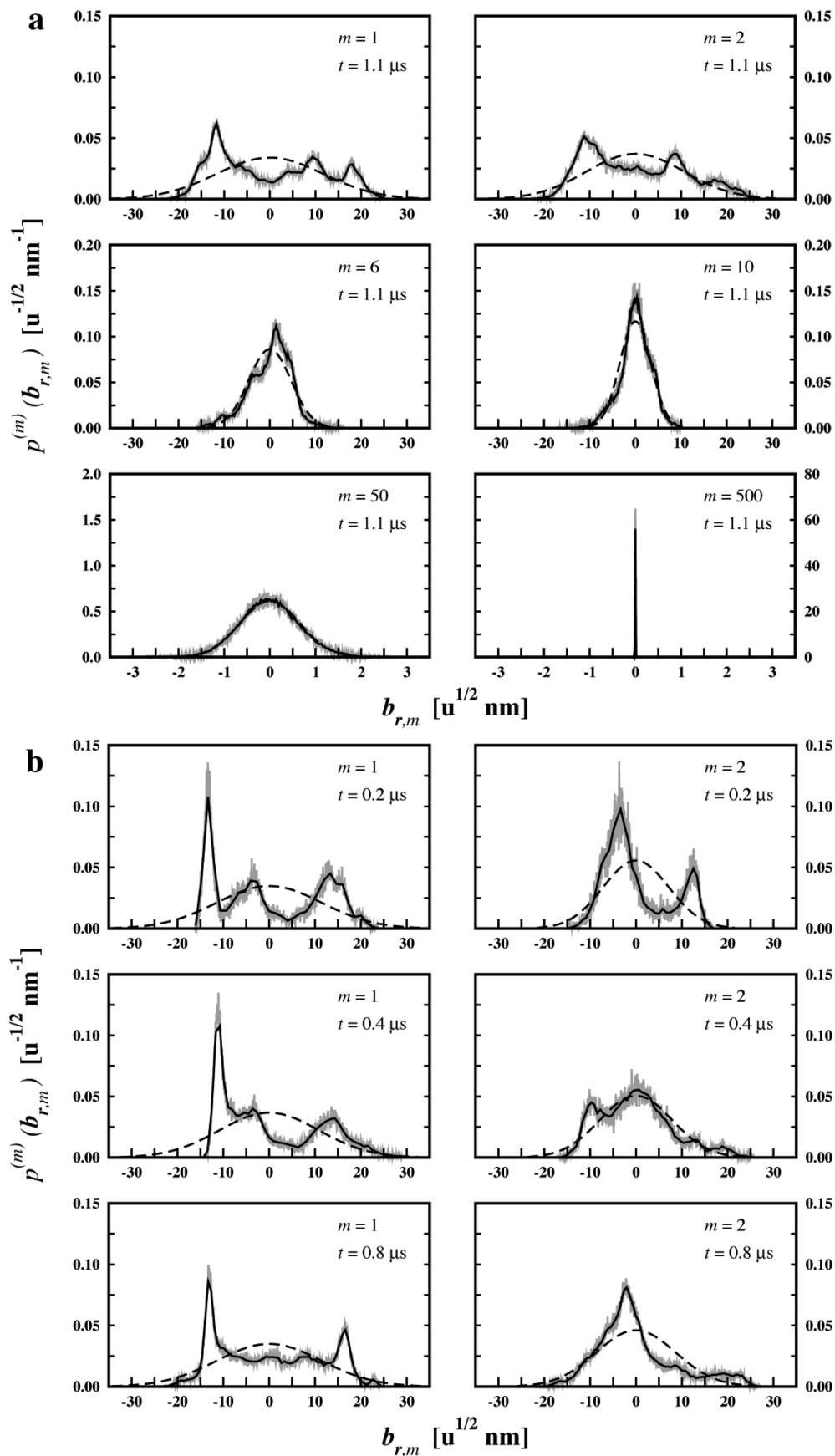


Figure 3. Probability distributions along selected components of the QH coordinate b for the $cc\beta$ peptide. The actual distributions (gray line) are displayed together with the corresponding Gaussians, i.e., $p'_{o,m}(b_{r,m}) = (2\pi F_{r,m})^{-1/2} e^{-1/2 F_{r,m}^{-1} b_{r,m}^2}$ (dashed lines) for (a) increasing component indices, m , and (b) increasing periods of time, t . The distributions employed for optimal numerical integration of the actual distributions (eq 18) are also displayed (solid lines). All probability distributions are normalized. Note that different scaling may be employed for graphical purposes. The letter “u” stands for atomic mass unit.

multiple off-center Gaussian-like distributions. A similar observation was previously reported in the context of two β -peptides in methanol for which two main subensembles of folded and unfolded configurations could be disentangled based on the lowest frequency modes (Figures 3–5 in ref 11). In the present case, the most pronounced peaks for the $cc\beta$ peptide arise from folded configurations (see Figure 3a, $m = 1$ and 2, leftmost peak).

The time dependence of these results was investigated as summarized in Figure 3b for the two coordinates \mathbf{b}_r with lowest frequencies, i.e., those contributing the most to the total mean-square metric-tensor-weighted fluctuation of the system (eq 6). The corresponding distributions vary significantly with the extent of PS sampling, as revealed by averaging over the first 0.2, 0.4, and 0.8 μs periods, or over the entire 1.1 μs ensemble (cf. Figure 3b vs Figure 3a for $m = 1$ and 2). Increasing the simulation time results into broader distributions due to the larger extent of PS sampled. The intensity of the leftmost peak, corresponding to the contribution of the folded configurations, clearly reduces along the simulation initialized from the $cc\beta$ helical fold and evolving through a broad range of heterogeneous configurations (Figure 1a). The data indicate that convergence of the probability distributions associated with the low-frequency QH coordinates in \mathbf{b}_r requires sampling times longer than 1 μs , considering that such differences persist when comparing results from the first 800 ns period with the whole 1.1 μs simulation.

3.3. Entropy Convergence. The entropy convergence for the $cc\beta$ peptide in water as a function of sampling time is illustrated in Figure 2b. The upper-bound curve obtained by application of the uncorrected QH formula ($S_{qm,o}$; eq 10 with eq 12) is compared to the build-up curve of the improved absolute single-molecule entropy (S^{ctd} ; eq 20). Both curves require periods of several hundred nanoseconds to reach a first plateau. Interestingly, the QH upper bound reaches convergence noticeably faster than the improved absolute single-molecule entropy. In detail, ~ 0.3 or ~ 0.7 μs is needed to sample 90% or 99% of the final $S_{qm,o}$ estimate of $6922 \text{ J K}^{-1} \text{ mol}^{-1}$, while larger sampling times of ~ 0.5 or ~ 1.0 μs are needed to sample 90% or 99% of the final S^{ctd} estimate of $5916 \text{ J K}^{-1} \text{ mol}^{-1}$ (see also Supporting Information, Table S2). These results clearly demonstrate that the convergence of the $S_{qm,o}$ upper bound does not imply the convergence of the absolute single-molecule entropy, S^{ctd} . In fact, the first quantity relies on the convergence of linear motional correlations only (following from the definition of linearly independent QH modes in eq 7 with eq 8). Instead, as described below, the second quantity requires in addition the convergence of supralinear motional correlations. For this reason, S^{ctd} seems to represent a better indicator of convergence (compared to $S_{qm,o}$) for the absolute single-molecule entropy.

The convergence behavior of the cumulative correction term, ΔS_{cl}^{cre} (eq 20), was also monitored (Figure 2c). This entropy term describes the overall deviation from the QH model due to mode anharmonicity (ΔS_{cl}^{ah} , eq 18; $4 \text{ J K}^{-1} \text{ mol}^{-1}$ after 1.1 μs) and correlation (ΔS_{cl}^{pc} , eq 19; $1002 \text{ J K}^{-1} \text{ mol}^{-1}$ after 1.1 μs) effects, associated with all unique

combinations of modes m and n . Its convergence behavior can be used as well as a measure of uncertainty on the entropy estimate. For comparison with previous studies (Figure 1), the time evolution of the corresponding relative contribution f_{cl}^{cre} (eq 21) to the uncorrected $S_{qm,o}$ estimate (eq 10 with eq 12) is also displayed (Figure 2d). In all studies, the dominant part of this correction arises from QH pairwise (supralinear) mode correlations (f_{cl}^{ah} values are $< 0.05\%$; see also Supporting Information, Tables S1 and S2).

Importantly, it is found that the magnitude of the correction term ΔS_{cl}^{cre} monotonically decreases from an initial value of $-2745 \text{ J K}^{-1} \text{ mol}^{-1}$ (first 0.5 μs sampling) to a final value of $-1006 \text{ J K}^{-1} \text{ mol}^{-1}$ (entire 1.1 μs sampling), showing an initial convergence behavior. This suggests that limited PS sampling results in both the underestimation of $S_{qm,o}$ and the overestimation of ΔS_{cl}^{cre} (predominantly through its ΔS_{cl}^{pc} component), both artifacts leading to an underestimation of the final absolute single-molecule entropy S^{ctd} .

This result can be explained considering that motional correlations are larger for a molecular system sampling a confined part of PS as opposed to sampling of a multiple-minima landscape. For the $cc\beta$ test system these results demonstrate that a limited PS sampling leads to the overestimation of corresponding motional correlations, thus of ΔS_{cl}^{pc} values with respect to that expected for a canonical ensemble of the same system at thermodynamic equilibrium.

The mass-weighted root-mean-squared fluctuation, i.e., the sum of the eigenvalues of the mass-weighted covariance matrix, $Tr[\mathbf{E}_r]$ (eq 9), was also monitored along time as an independent measure of convergence (Figure 2e). In terms of $Tr[\mathbf{E}_r]$, we note that a first plateau region is reached after ~ 1.1 μs (Figure 2e), in line with the S^{ctd} values (Figure 2b). This observation confirms as well that the $cc\beta$ peptide is not trapped in a few local minima. Instead, it explores new configurations even after several hundreds of nanoseconds (Figure 2a).

Three important general points are worth noting.

First, the magnitude of the cumulative correction term ΔS_{cl}^{cre} is large. This is evident when the term is expressed in the form of its contribution to the system free energy, $T\Delta S_{cl}^{cre}$ (Figure 2c, right axis). The resulting value (302 kJ mol^{-1} based on 1.1 μs) is about an order of magnitude larger than the free energy changes of typical (bio)chemical processes. Thus, although partial cancellation of this term can be expected for entropy differences between two different molecular environments or conformational states, small differences will still lead to large free-energy contributions (of sign and magnitude difficult to be predicted a priori). The importance of this correction for reliable entropy calculations is therefore evident. In addition, this result suggests that time convergence of the entropy estimate should be taken into account as well when comparing the efficiency and accuracy of alternative computational approaches.

Second, we stress that all M per-mode contributions need to be included for an accurate estimation of eqs 10 and 18 because modes with large m indices (high frequencies) also contribute to $S_{qm,o}$ (data not shown; see Figure 8 in ref 11 for a similar analysis). This marks a difference with what is typically observed for the contribution of a reduced number

of essential modes to the total system fluctuation.²³ Due to the similar mathematical formalism,¹¹ this argument can be easily demonstrated as well for the calculation of entropies from normal-mode analysis for systems sampling one local PS minimum.

Third, the analysis of the leading correction ΔS_{eff}^f in terms of all $C(2,M) = [(M-2)!]M!$ unique pair combinations reveals that not only modes with low indexes (high amplitudes, low frequencies) contribute substantially. Thus, all pairs of QH modes need to be considered in eqs 15 and 19. This requirement arises from the observation that high correlations can be present among modes with either large or small m and n indices (low or high frequencies; Figure 12 in ref 11). Interestingly, this behavior was observed for highly flexible systems (e.g., the *ccβ* peptide of this study or the reversibly folding β -peptides in ref 11) but not for more rigid systems confined to a local PS sampling (e.g., the W191G cavity in ref 20, unpublished results). Whether the latter result depends on a limited PS accessed or on the physical nature of QH-mode correlation remains to be addressed.

3.4. Numerical Integration of the Correction Terms.

The analysis presented in this work relies on the numerical integration of the actual probability distributions $p^{(m)}(\mathbf{b}_{r,m})$ and $p^{(m,n)}(\mathbf{b}_{r,m}, \mathbf{b}_{r,n})$ evaluated based on the MD trajectory (eqs 18 and 19). Two alternative procedures were described to estimate these 1D and 2D integrals with optimal (nonarbitrary) histogram bin widths, as detailed in Appendix C of ref 11. In this study, optimal parameters κ_1° and κ_2° were chosen as the midpoint between the intersections of a horizontal line with the limiting lines for too *fine* and too *coarse* integration at the optimal value of the 1D or 2D integrals in the graph showing these values as a function of $\ln \kappa_1$ or $\ln \kappa_2$, as summarized in Figure 4.

Figure 4a shows the values of the 1D integrals for a sample set of eigenvectors ($m = 1, 50, \text{ and } 500$), evaluated numerically using different values of κ_1 . Both limiting lines are shown, together with the optimal κ_1° values. Values approaching these limiting curves are incorrect because they show a dependence of the evaluated integral on the bin size. However, for each curve, a clear plateau defines the range of κ_1 values for which the integration result is essentially independent of the bin size. Finite-sampling artifacts affect the integration with the smallest values of κ_1 , while coarse-binning artifacts affect the integration with the largest values. Note that 1D integrals may be individually negative or positive.

Figure 4b shows the values of the 2D integrals for a sample group of eigenvector pairs ($m,n = 1,2; 1,100; 1,500$), evaluated numerically with different values of κ_2 , together with the corresponding limiting lines and the optimal κ_2° values. Here, the plateau regions are narrower and the value of κ_2 has to be chosen more carefully. Note that 2D integrals are always negative when estimated using the optimal κ_2° values, but incorrect positive values would be obtained based on too small κ_2 values.

The dependence of these curves on the simulation time was also monitored (Figure 4). All 1D and 2D curves show a coarse-integration limit that is essentially independent of the MD period considered. Yet, they also show that the fine-

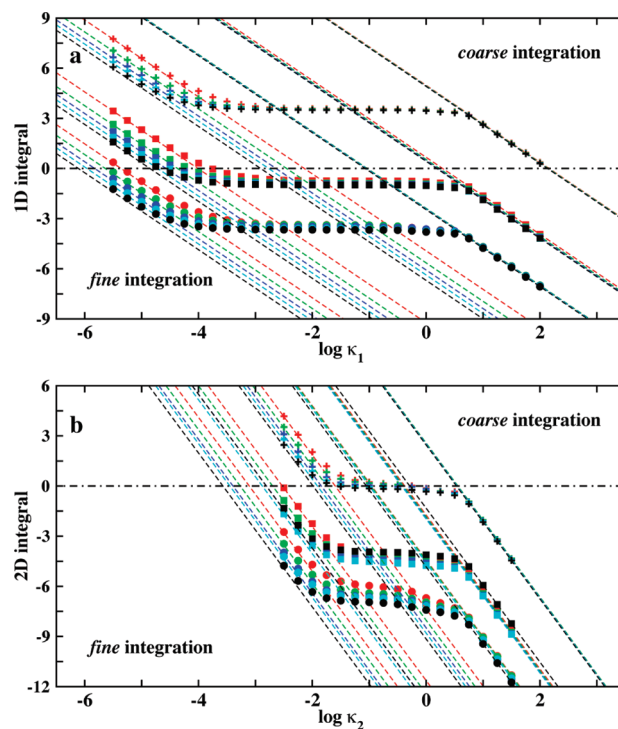


Figure 4. Dependence of the numerical integration of probability distributions on the width of histogram bins for increasing periods of time, t . (a) Integrals over the 1D distributions involved in eq 18 are shown for eigenvectors 1 (circles), 50 (squares), and 500 (crosses). (b) Integrals over the 2D distributions involved in eq 19 are shown for eigenvector pairs 1,2 (circles), 1,100 (squares), and 1,500 (crosses). The results are displayed for the whole ensemble (1.1 μs ; black) or $t = 200$ (red), 400 (green), 600 (blue), 800 (cyan) ns as a function of $\ln \kappa_1$ (a) or $\ln \kappa_2$ (b), where κ is the ratio of the bin width along each dimension to the corresponding distribution width. The middle point between a pair of limiting lines for too fine (left side) and too coarse (right side) numerical integrations (dashed lines) defines optimal κ_1° or κ_2° values. A black dot-dashed reference line is drawn at zero.

integration-limiting curves shift to lower κ_1 and κ_2 values upon increasing the simulation time, thus reducing the dependence on the integration bin size. This effect is more pronounced for the 2D integrals because they require more data points than 1D integrals.

Overall, these results demonstrate that the procedure employed in this work allows estimating both 1D and 2D integrals of eqs 18 and 19 in a nonarbitrary way as a function of the simulation time. The presence of plateau regions independent of the integration bin size for all MD periods considered shows that the observed change of integral values along the simulation time largely depends on the extent of PS sampled yet not on the numerical procedure employed. A similar analysis performed in the case of corresponding 3D integrals (triplewise combinations of QH modes) revealed that indeed no such behavior can be achieved, although using a 1.1 μs trajectory. In practice, eq 15 can be only estimated for the first two terms owing to finite sampling artifacts and data sparseness.

4. Conclusion

The theory and practical implementation of an approach recently proposed¹¹ to estimate improved configurational entropies from quasi-harmonic analysis of molecular dynamics simulations are briefly reviewed. It involves the calculation of correction terms of increasingly high order to account for deviations from the quasi-harmonic approximation in frustrated molecular systems. The convergence properties of the absolute single-molecule entropy are critically investigated using microsecond molecular dynamics simulation of the *ccβ* peptide in water. Prompted by the comparison of the results with previous studies addressing mode anharmonicity and correlation effects, the convergence behavior of individual quasi-harmonic modes, of the absolute single-molecule entropy, and of the correction terms for anharmonicity and pairwise (supralinear) correlations are analyzed. Our data provide a number of new insights to tackle the challenge of accurate entropy estimation by computer simulation.

In line with a previous study,¹¹ the probability distributions associated with components of the quasi-harmonic coordinates only deviate significantly from Gaussian functions for the first few components, resembling the behavior observed in the different context of a single-atom displacement for an α -helical peptide⁴² and of essential modes for protein dynamics.²³ For these components, the probability distributions result from a superposition of clearly distinguishable contributions from the folded and unfolded ensembles. However, it is shown that the components of these eigenvectors converge slowly ($>1 \mu\text{s}$), consistent with the observation that the *ccβ* peptide steadily explores new configurations.

In line with previous studies,^{11,20,25,26,28} the entropic contribution of anharmonicity is small while the pairwise (supralinear) correlation correction to the entropy is large. The deviation from the quasi-harmonic assumption affects more significantly conformational states dominated by high motional correlation. Using microsecond molecular dynamics simulation of a peptide test system we show that limited phase-space sampling results in an overestimation of correlation effects, and we discuss its implications for entropy estimation.

This study demonstrates that the convergence of the quasi-harmonic upper-bound entropy with simulation time does not imply the convergence of the system absolute single-molecule entropy. As a consequence, our study also suggests that the convergence of the absolute single-molecule entropy rather than that of the quasi-harmonic upper bound should be preferably monitored. Because the cumulative correction term accounting for both mode anharmonicity and pairwise (supralinear) correlation effects converges slowly and monotonically decreases, previous studies based on shorter time scales may have, in some cases, partly overestimated this correction term, thus leading to underestimated absolute entropy estimates.

Overall, the present study emphasizes the need of sufficient phase-space sampling to estimate entropic contributions from computer simulations. Ideally, only thermodynamic ensembles at equilibrium should be considered to this end, i.e., full phase-space sampling obtained from simulations on time

scales of several microseconds. In practice, we suggest that enhanced sampling techniques^{28,43} and/or concatenated copies of independent simulation trajectories^{21,24} will be useful tools to alleviate these problems in the future if properly combined with the correction terms used herein.²⁸ This strategy will open the possibility to include as well correlation effects of higher order than the pairwise (supralinear) explicitly considered in this study. A bright future opens for the estimation of accurate thermodynamic properties for biomolecular systems using chemical theory and computation.

Acknowledgment. We thank the members of the McCammon group for useful discussions. This work was supported by the National Science Foundation grant PHY-0822283 and Center for Theoretical Biological Physics for the computing resources, the National Institutes of Health and National Science Foundation to R.B. and J.A.M., and the Howard Hughes Medical Institute and the National Biomedical Computing Resource to J.A.M.

Supporting Information Available: Summary of the single-molecule absolute entropy for systems of different chemical nature, Table S1; entropy and correction values along microsecond molecular dynamics of the *ccβ* peptide in water, Table S2; derivation of eq 15, Appendix. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) Gö, N.; Scheraga, H. A. *J. Chem. Phys.* **1969**, *51*, 4751.
- (2) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325.
- (3) Di Nola, A.; Berendsen, H. J. C.; Edholm, O. *Macromolecules* **1984**, *17*, 2044.
- (4) Edholm, O.; Berendsen, H. J. C. *Mol. Phys.* **1984**, *51*, 1011.
- (5) Rojas, O. L.; Levy, R. M.; Szabo, A. *J. Chem. Phys.* **1986**, *85*, 1037.
- (6) Schlitter, J. *Chem. Phys. Lett.* **1993**, *215*, 617.
- (7) Schafer, H.; Mark, A. E.; van Gunsteren, W. F. *J. Chem. Phys.* **2000**, *113*, 7809.
- (8) Andricioaei, I.; Karplus, M. *J. Chem. Phys.* **2001**, *115*, 6289.
- (9) Reinhardt, W. P.; Miller, M. A.; Amon, L. M. *Acc. Chem. Res.* **2001**, *34*, 607.
- (10) Chang, C. E.; Chen, W.; Gilson, M. K. *J. Chem. Theory Comput.* **2005**, *1*, 1017.
- (11) Baron, R.; van Gunsteren, W. F.; Hünenberger, P. H. *Trends Phys. Chem.* **2006**, *11*, 87.
- (12) Wang, J.; Bruschweiler, R. *J. Chem. Theory Comput.* **2006**, *2*, 18.
- (13) Chang, C. E.; Chen, W.; Gilson, M. K. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1534.
- (14) Killian, B. J.; Yundenfreund Kravitz, J.; Gilson, M. K. *J. Chem. Phys.* **2007**, *127*, 24107.
- (15) Hnizdo, V.; Tan, J.; Killian, B. J.; Gilson, M. K. *J. Comput. Chem.* **2008**, *29*, 1605.
- (16) Carlsson, J.; Åqvist, J. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5385.

- (17) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glatli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F.; Yu, H. B. *Angew. Chem., Int. Ed.* **2006**, *45*, 4064.
- (18) Meirovitch, H. *Curr. Opin. Struct. Biol.* **2007**, *17*, 181.
- (19) Meirovitch, H.; Cheluvajala, S.; White, R. P. *Curr. Protein Pept. Sci.* **2009**, *10*, 229.
- (20) Baron, R.; McCammon, J. A. *ChemPhysChem* **2008**, *9*, 983.
- (21) Baron, R.; de Vries, A. H.; Hünenberger, P. H.; van Gunsteren, W. F. *J. Phys. Chem. B* **2006**, *110*, 8464.
- (22) McLachlan, A. D. *J. Mol. Biol.* **1979**, *128*, 49.
- (23) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins* **1993**, *17*, 412.
- (24) Baron, R.; de Vries, A. H.; Hünenberger, P. H.; van Gunsteren, W. F. *J. Phys. Chem. B* **2006**, *110*, 15602.
- (25) Pereira, C. S.; Kony, D.; Baron, R.; Müller, M.; van Gunsteren, W. F.; Hünenberger, P. H. *Biophys. J.* **2006**, *90*, 4337.
- (26) Pereira, C. S.; Kony, D.; Baron, R.; Müller, M.; van Gunsteren, W. F.; Hünenberger, P. H. *Biophys. J.* **2007**, *93*, 706.
- (27) Gorfe, A. A.; Baron, R.; McCammon, J. A. *Biophys. J.* **2008**, *95*, 3269.
- (28) Peric-Hassler, L.; Hansen, H. S.; Baron, R.; Hünenberger, P. H. Manuscript in preparation.
- (29) Somani, S.; Killian, B. J.; Gilson, M. K. *J. Chem. Phys.* **2009**, *130*, 134102.
- (30) Matsuda, H. *Phys. Rev. E* **2000**, *62*, 3096.
- (31) Numata, J.; Wan, M.; Knapp, E. W. *Genome Inform.* **2007**, *18*, 192.
- (32) Carlsson, J.; Åqvist, J. *J. Phys. Chem. B* **2005**, *109*, 6448.
- (33) Darian, E.; Hnizdo, V.; Fedorowicz, A.; Singh, H.; Demchuk, E. *J. Comput. Chem.* **2005**, *26*, 651.
- (34) Lu, B. Z.; Wong, C. F. *Biopolymers* **2005**, *79*, 277.
- (35) Prompers, J. J.; Bruschiweiler, R. *J. Phys. Chem. B* **2000**, *104*, 11416.
- (36) Case, D. A.; Darden, T.; Cheatham, T. III; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Merz, K.; Pearlman, D.; Crowley, M.; Walker, R.; Zhang, W.; Wang, B.; Hayik, A.; Roiberg, A.; Seabra, G.; Wong, K.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Morgan, J.; Hornak, V.; Cui, G.; Beroza, P.; Matthews, D.; Schffmeister, C.; Ross, W.; Kollman, P. *AMBER 9*; University of California: San Francisco 2006.
- (37) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.
- (38) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (39) Kammerer, R. A.; Kostrewa, D.; Zurdo, J.; Detken, A.; Garcia-Echeverria, C.; Green, J. D.; Müller, S. A.; Meier, B. H.; Winkler, F. K.; Dobson, C. M.; Steinmetz, M. O. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 4435.
- (40) Baron, R.; de Oliveira, C. A. F.; McCammon, J. A. Manuscript in preparation.
- (41) Christen, M.; Hünenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. *J. Comput. Chem.* **2005**, *26*, 1719.
- (42) Perahia, D.; Levy, R. M.; Karplus, M. *Biopolymers* **1990**, *29*, 645.
- (43) Minh, D. D.; Hamelberg, D.; McCammon, J. A. *J. Chem. Phys.* **2007**, *127*, 154105.

CT900373Z