Estimation of absolute solvent and solvation shell entropies via permutation reduction

Friedemann Reinhard and Helmut Grubmüller

Theoretical and Computational Biophysics Department, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

Despite its prominent contribution to the free energy of solvated macromolecules such as proteins or DNA, and although principally contained within molecular dynamics simulations, the entropy of the solvation shell is inaccessible to straightforward application of established entropy estimation methods. The complication is twofold. First, the configurational space density of such systems is too complex for a sufficiently accurate fit. Second, and in contrast to the internal macromolecular dynamics, the configurational space volume explored by the diffusive motion of the solvent molecules is too large to be exhaustively sampled by current simulation techniques. Here, we develop a method to overcome the second problem and to significantly alleviate the first one. We propose to exploit the permutation symmetry of the solvent by transforming the trajectory in a way that renders established estimation methods applicable, such as the quasiharmonic approximation or principal component analysis. Our permutation-reduced approach involves a combinatorial problem, which is solved through its equivalence with the linear assignment problem, for which $O(N^3)$ methods exist. From test simulations of dense Lennard-Jones gases, enhanced convergence and improved entropy estimates are obtained. Moreover, our approach renders diffusive systems accessible to improved fit functions. © *2007 American Institute of Physics.*
[DOI: 10.1063/1.2400220]

I. INTRODUCTION

Biomolecular processes are driven by molecular free energy changes. Many physico-chemical phenomena, such as the hydrophobic effect, emerge from a fine-tuned competition of the two free energy components, entropy and enthalpy. During protein folding, e.g., molecular interactions compete with the associated huge decrease of conformational entropy of the protein.[1] Hydrophobic forces, which drive many biological phenomena such as membrane association and also protein folding, are also governed by entropy changes, but mainly of the solvent.[2] As a fully atomistic description, molecular dynamics (MD) simulations should capture this enthalpy-entropy competition. That this is actually the case has become particularly evident by the successful first-principle MD folding of peptides and proteins of increasing size,[3,4] where the obtained native structure is particularly sensitive to the enthalpy-entropy balance. Similarly, MD has provided in-depth insights of many other complex biomolecular processes.[5–9]

Unexpectedly, however, accurate values for both enthalpy and entropy are notoriously difficult to extract from such simulations, although they are correctly treated and thus, apart from the sampling problem, are fully contained in the simulation. This at first almost paradoxical gap severely inhibits a full causal and thermodynamic understanding of many macromolecular processes.

Free energy *differences* between two states of a system are usually obtained from MD simulations by means of free energy perturbation[10] or thermodynamic integration;[11] absolute values, however, are more difficult to obtain. In both cases accuracy is limited by sampling. For differences between similar systems, reasonable accuracy can be obtained.[12–14] For absolute values, however, the sampling problem is usually prohibitive due to the large volume of phase space that needs to be sampled. Even more so, the separate computation of enthalpy or entropy differences typically suffers from large statistical uncertainties.[15] Accordingly, to estimate these quantities already represents a considerable challenge.

In addition, biological systems usually involve a solute (e.g., a protein) solvated in a solvent (e.g., water). Therefore, an estimation method should include both groups in the analysis and distinguish their contributions. This is an additional challenge which is not met satisfactorily by any method proposed so far.

The quasiharmonic approach[16,17] approximates the phase space density of a system by fitting a multivariate Gaussian density to a MD simulation trajectory using the covariance matrix. From this density estimate, the entropy is calculated analytically. This approach is computationally efficient and often yields good results for the entropy contribution of the solute.[18,19] As we will discuss further below, however, the solvent contribution to the entropy is not accessible to this method. This is particularly problematic since the solvent degrees of freedom are typically correlated with those of the solute and, therefore, the associated entropies are usually not separable.

To compute entropy differences between two states, thermodynamic integration (TI) (Refs. 10 and 11) is the method

of choice. TI is most widely used for the computation of free energy differences and, in principle, includes both solute and solvent. More recently, it has been extended to the computation of entropy differences.[15,20] This variant, however, suffers from severe sampling problems, and is therefore computationally particularly expensive.

A third method, adiabatic switching,[21] is also based on a gradual change of the system's Hamiltonian. In contrast to TI, this change is performed adiabatically, i.e., the pathway is chosen such that the entropy does not change. This change is used to transform the system into a simpler one whose entropy can be calculated analytically. This method has been successfully applied to pure solvent systems; its application to mixed systems, however, is not straightforward,[15] and has not yet been demonstrated.

In summary, despite its importance, estimation of entropy contributions to the free energy and, in particular, those of the solvent, still represents a considerable challenge. Here, we propose a method to fill this gap. By exploiting permutation symmetry we propose to transform the solvent trajectory such that density-based entropy estimation methods like the quasiharmonic approach become applicable.

II. THEORY

Because our method rests on the quasiharmonic approximation, and also to clarify notation, we will first sketch this method. For a comprehensive treatment, the reader is referred to the literature.[16,17] Subsequently, we will describe how permutation symmetry can be exploited to render the solvent accessible to the quasiharmonic approximation.

A. Quasiharmonic approximation

Within the quasiharmonic approximation framework, one assumes that a molecular dynamics trajectory $\{\mathbf{x}_1(t),\ldots,\mathbf{x}_N(t)\}=\{\mathbf{x}(t)\}$ of a classical N-particle system approximates a Boltzmann ensemble which is distributed according to the configuration phase space density,

$$\rho(\mathbf{x}) = \frac{1}{Z} e^{-\beta V(\mathbf{x})},$$

of the system under study. Here, $\mathbf{x}(t)$ denotes the configurational vector of the system at time t, β is the reciprocal thermal energy, $V(\mathbf{x})$ the potential energy (force field), and Z the canonical partition function; the contributions of the momenta can be treated analytically and are, therefore, not considered here. Usually, and also here, the trajectory is one from which rigid body motions of the solute have been removed.

Accordingly, the trajectory is used to reconstruct an approximate density

$$\widetilde{\rho}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^{3N}\det(\mathbf{C})}} e^{-(\mathbf{x}-\langle\mathbf{x}\rangle)^T \mathbf{C}^{-1}(\mathbf{x}-\langle\mathbf{x}\rangle)}$$

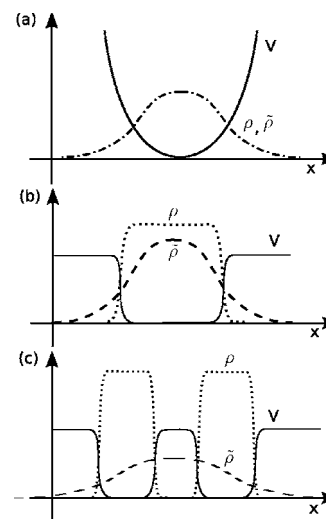from the $3N \times 3N$-dimensional covariance matrix

FIG. 1. Sketch of quasiharmonic density estimates for (a) a protein fluctuating around its average structure; (b) an ideal gas in a box; (c) a solvent in a box. Shown are potential energy (solid lines), resulting configurational densities ρ (dotted), and derived Gaussian density estimates $\widetilde{\rho}$ (dashed).

$$C_{ij} = \langle(x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle)\rangle,$$

where $\langle\ldots\rangle$ denotes an average over the trajectory. For harmonic $V(\mathbf{x})$, this approximation is exact; hence, the approach is referred to as quasiharmonic approximation.[22]

The approximate density $\widetilde{\rho}$ serves to obtain an analytical estimate $\widetilde{S}$ for the (configurational) entropy S,[16]

$$S \approx \widetilde{S} = \tfrac{1}{2}k_B\{3N + \ln[(2\pi)^{3N}\det(\mathbf{C})]\}. \tag{1}$$

More recently, a quantum-mechanical correction has been proposed[17] which additionally is numerically more stable,

$$\widetilde{S} = \frac{1}{2}k_B\sum_{i=1}^{3N} \ln\left[1 + \frac{k_B T e^2}{\hbar^2}\sigma_i^2\right]. \tag{2}$$

Here, σ_i^2 are the $3N$ eigenvalues of the mass-weighted covariance matrix.

For later reference we note that for indistinguishable particles, a correction $k_B \ln N!$ (Gibbs factor) has to be applied. Due to their chemical bonding, the atoms of a protein molecule are usually considered distinguishable, and no Gibbs factor is applied.

By construction, the harmonic approximation is accurate as long as ρ resembles a Gaussian distribution. Solvated biological macromolecules often adopt a well-defined average structure, and the atomic fluctuations around their average structure are small (though exceptions exist). Therefore, their configurational space density is compact, and the Gaussian density approximation is reasonably good [Fig. 1(a)]. The quasiharmonic approximation is also applicable to an ideal gas in a finite box.[19,23] This is unexpected because the associated rectangular phase space density [Fig. 1(b)] is far from Gaussian-shaped.

However, for noncompact densities the harmonic approximation breaks down [Fig. 1(c)]. In this situation, the Gaussian density approximation is maximal at the "hole" in the center of Fig. 1(c), where the actual density vanishes.

This sketch captures the situation for solvents. Here, the diverging potential for overlapping solvent atoms or molecules creates, via the Boltzmann factor, holes in the high-dimensional configurational space density ρ of the system. Even larger holes arise from repulsive interactions of the solvent molecules with a macromolecular solute such as a protein. This is one of the reasons why the solvent contribution to the entropy is usually inaccessible to the harmonic approximation.

B. Permutation symmetry

Motivated by this observation, and aiming at removal of the holes, we propose to "localize" the solvent degrees of freedom by exploiting the solvent's permutation symmetry. Specifically, we will permute the indices of the solvent molecules for each individual trajectory frame $\mathbf{x}(t)$ such as to minimize the error introduced by the Gaussian density approximation. This process will be referred to as "relabeling." Due to its permutation symmetry, relabeling of the molecules leaves the physics of the system invariant, but translates the $3N$-dimensional configurational vectors $\mathbf{x}$ of the trajectory and, therefore, enables us to tailor its "shape."

Assuming that the approximation error decreases with increasing compactness of the trajectory, we therefore propose to relabel each frame i of the trajectory such as to minimize its $3N$-dimensional radius of gyration,

$$|\pi_i \cdot \mathbf{x}(t_i) - \langle \mathbf{x} \rangle|^2 \overset{!}{=} \min, \tag{3}$$

where π_i is a permutation operator such that $\pi_i \cdot \mathbf{x}(t_i)$ denotes the relabeled configuration at time t_i, and $\langle \mathbf{x} \rangle$ is the average configuration. Note that the latter depends on the particular choice of all optimal permutations π_i, which renders the above equation recursive and precludes straightforward optimization. We circumvent this problem by replacing the permutation-dependent average configuration with an appropriately chosen reference configuration $\mathbf{x}_0$. This trick reduces the coupled optimization problem, Eq. (3), to the independent optimization problems for each trajectory frame i of finding a permutation π_i which minimizes the distance to the reference configuration, $|\pi_i \cdot \mathbf{x} - \mathbf{x}_0|$.

This problem is a special case of the linear assignment problem (LAP), for which efficient algorithms have been developed.[24,25] We will refer to this procedure—relabeling the trajectory according to the optimal permutations—as "permutation reduction," and to an entropy estimate derived from the permutation-reduced trajectory as "PRPCA" estimate.

Figure 2 illustrates this approach for a simulation of 216 water molecules in a cubic box. Shown are superpositions of the trajectories of all oxygen atoms; the atom numbers are color coded. Figure 2(a) shows the trajectory as obtained from the simulation. Clearly visible is the free self-diffusion of the molecules, such that each color covers the whole volume. Figure 2(b) displays the permutation-reduced trajectory. As intended, the diffusive motion is converted into localized fluctuations of the water molecules around their reference positions. In this respect, the water molecules collectively
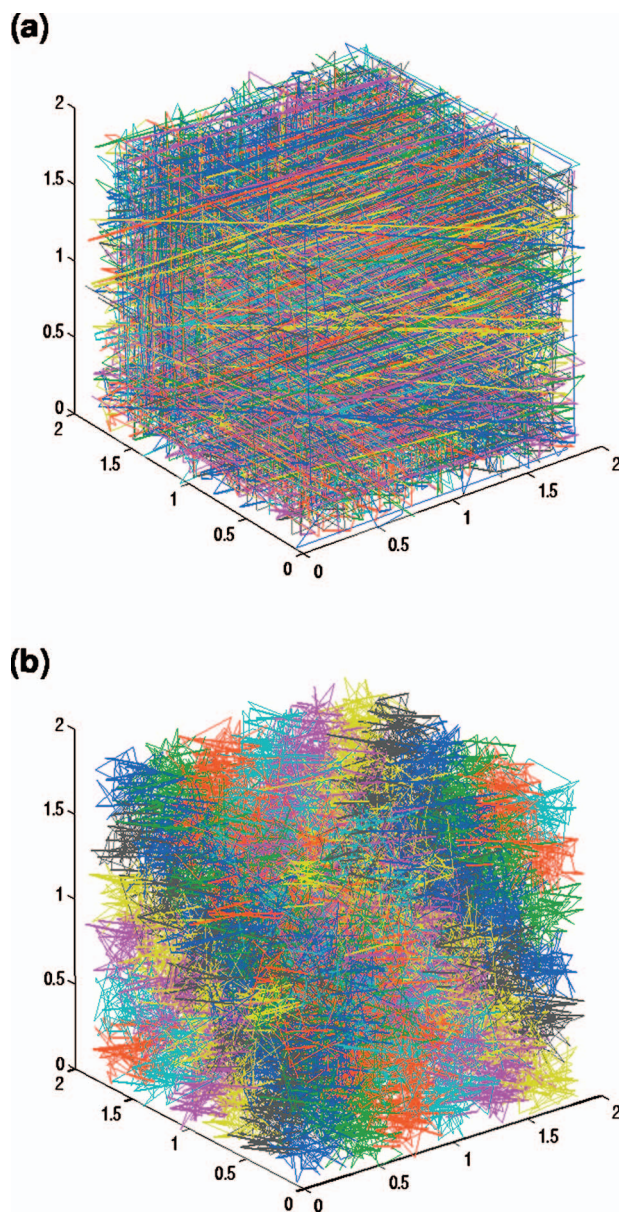

FIG. 2. (Color) Molecular dynamics simulation of 216 water molecules within a cubic box. Shown are trajectories of the oxygen atoms without (a) and with (b) permutation reduction applied. In the first case, all molecules diffuse freely within the box volume. After permutation reduction, each molecule remains close to its reference position because it is relabeled otherwise. Note that the physics of the system remains unchanged.

behave similar to a protein fluctuating around its well-defined average structure; in this sense, the solvent has been "proteinized."

By construction, permutation reduction projects the trajectory into one of $N!$ identical subvolumes of the configurational space. Accordingly, the underlying configurational space density ρ is projected onto a reduced density ρ_0, which vanishes everywhere outside the subvolume accessible to the relabeled trajectory. This reduced density ρ_0 is by construction much more compact than the original density ρ. Accordingly, the Gaussian approximation should allow for an improved estimate of the density ρ_0 as well as of its entropy S_0. The original density is easily recovered by applying all $N!$ permutations

$$\rho(\mathbf{x}) = \frac{1}{N!} \sum_{\pi} \rho_0(\pi^{-1} \cdot \mathbf{x}),$$

where the factor $N!$ accounts for proper normalization. It follows that $S = S_0 + k_B \ln(N!)$. We note that for indistinguishable particles, $S = S_0$.

III. METHODS

We implemented the outlined approach using the LAP algorithm of Jonker and Volgenant,[26] which solves the LAP with $O(N^3)$ efficiency. This implementation is available electronically.[27]

To assess the accuracy of our approach, we chose a gas of Lennard-Jones particles as a first simple test system. Similar test systems have also been considered previously.[18,19] The parameters c_{12} and c_6 of the Lennard-Jones interaction potential,

$$V(r) = \frac{c_{12}}{r^{12}} - \frac{c_6}{r^6},$$

were scaled by a dimensionless parameter λ such that $c_{12} = \lambda \times 9.8493 \times 10^{-6}$ nm^{12} kJ mol^{-1} K^{-1}, $c_6 = \lambda \times 6.2644 \times 10^{-3}$ nm^6 kJ mol^{-1} K^{-1}. The case $\lambda = 1$ corresponds to argon.[19,28] We simulated 100 atoms of mass 39.948 amu within a box of constant volume $(2.2$ nm$)^3$ with periodic boundary conditions. The system was kept at $T = 300$ K by coupling to a temperature bath with a 0.1 ps relaxation time.[29] Lennard-Jones interactions were cut off at $r = 1.0$ nm. All simulations were started from an energy-minimized structure, and a 2 fs time step was used for the numerical integration of the equations of motion.

Two sets of simulations were performed. The first set comprised eighteen 50 ns trajectories. For each trajectory, a fixed λ between 0 and 1 was chosen. Note that $\lambda = 0$ describes an ideal gas. For the second set of simulations, the interaction of the particles was held fixed at $\lambda = 1$, and an additional particle—mimicking a solute—was inserted at the center of the box. This particle was restrained by a harmonic potential with a spring constant of 10^6 kJ mol^{-1} nm^{-2}. For its interaction with the other particles, a Lennard-Jones potential with the above parameters was chosen, with fixed λ ranging from $\lambda = 0$ to $\lambda = 100$.

For all trajectories, the 100 solvent atoms were permutation reduced. To characterize the influence of the reference structure $\mathbf{x}_0$ on the entropy estimate, three different types of reference structures were considered, the minimized initial structure, a simple cubic lattice of argon atoms, lattice constant $d = 0.55$ nm, and the last frame of the respective simulation.

Covariance analyses were carried out for both the original and the relabeled trajectories. From the obtained covariance matrices, the entropy was estimated via Eq. (2). Excess entropies were computed by subtracting the ideal gas value obtained with the respective estimation method. For proper comparison, the $N!$ correction was used for the PRPCA estimate, i.e., the particles were treated as distinguishable.

To obtain accurate entropy reference values, for each of the considered systems the excess entropy to the ideal gas
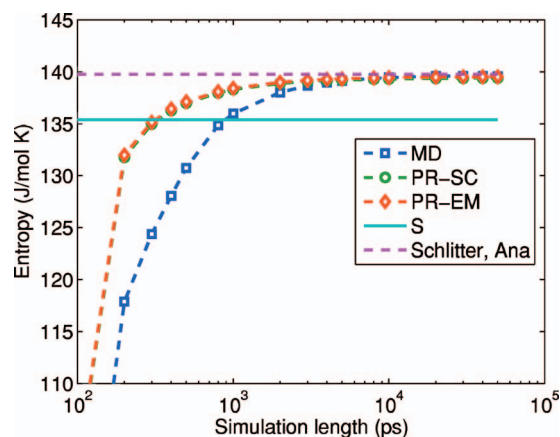

FIG. 3. (Color) Convergence of an ideal gas simulation. Shown are entropy estimates for different portions of the original trajectory (MD) as well as of the relabeled trajectories with the simple cubic structure (PR-SC) and the energy-minimized structure (PR-EM) as reference structures. The horizontal lines show the analytical results for the ideal gas entropy (S) and its Gaussian approximation (Schlitter, Ana method). Note that The PR-SC and PR-EM curves match so closely that they can hardly be distinguished in the figure at the chosen scale.

was calculated by thermodynamic integration,[11,15,20] which was possible for the simple test system used here. The computations were performed by switching the interaction parameter in 100 steps from 0 to λ. For each step, a 200 ps simulation was performed. A soft core potential, as implemented in GROMACS, was used with $\alpha = 1.5$ and $\sigma = 0.3$.

To check for convergence, forward and backward thermodynamic integration was performed for both simulation sets. As the worst case, the largest interaction strengths were considered ($\lambda = 1$ for the first and $\lambda = 100$ for the second set, respectively). Only small differences below 0.12 J mol^{-1} K^{-1} for the first and 0.04 J mol^{-1} K^{-1} for the second series were observed. Hence, sufficient convergence is assumed.

All simulations were carried out using GROMACS 3.2.1.[30] For the thermodynamic integration calculations, the code was slightly modified to output the derivatives of the Hamiltonian required for the entropy thermodynamic integration.[15]

IV. RESULTS AND DISCUSSION

A. Convergence of the PRPCA entropy estimate

To test for convergence, the ideal gas case ($\lambda = 0$) of the first simulation set was considered. For this case, analytic reference values for the entropy as well as the Gaussian entropy estimate, Eq. (2), are available for comparison,[23] shown as horizontal lines in Fig. 3.

Figure 3 shows entropies from the Schlitter method for trajectory portions of different lengths. As can be seen, the entropy estimates from the relabeled portions (PR-SC, circles and PR-EM, diamonds) converge about three times faster than the estimates from the original trajectory portions (MD, squares). This improvement was expected and is because the relabeled trajectory is restrained to the much smaller subvolume of the phase space, which, for given trajectory length, is sampled more completely.

The good agreement of this estimate with the analytical result is unexpected. Although it has been pointed out
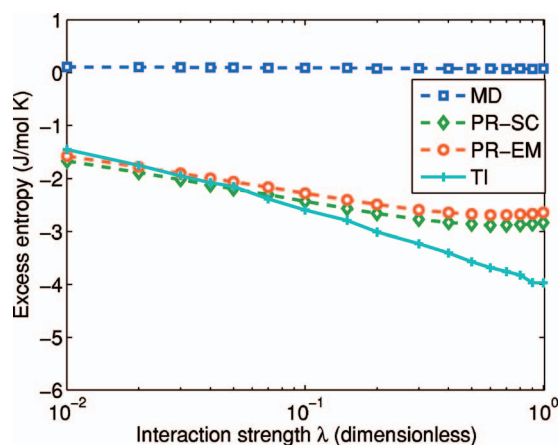
FIG. 4. (Color) Excess entropy of a Lennard-Jones gas, estimated from the original trajectory (MD) as well as from the relabeled trajectories with the simple cubic (PR-SC) and the energy-minimized structure (PR-EM) as reference structures. Values computed by thermodynamic integration (TI) provide a reference.

recently[23] that a Gaussian fit to a constant density distribution within a hypercube yields good entropy estimates, it is far from clear that a Gaussian fit function should remain an appropriate choice for the simplexlike subvolume sampled by the relabeled trajectory. In any case, this is an encouraging result.

All entropy estimates converge within 10 ns; therefore, sampling errors are considered negligible for the 50 ns test simulations described below.

B. Lennard-Jones gas

In real solvents, interactions give rise to correlations between molecules and thereby reduce the entropy. To analyze how well this effect is captured by our method, we studied a system of Lennard-Jones atoms for different interaction strengths (the first simulation set). With a density of $9.4/nm^3$ or $0.623 \ g/cm^3$, which approaches that of liquids, this system should represent a sufficiently tough test case.

Figure 4 compares two PRPCA estimates (PR-SC, PR-EM) with estimates obtained from Gaussian fits to the original trajectory (MD). As a reference, the thermodynamic integration result is provided (solid line). Note that the *excess* entropy is shown here. For this difference between two entropies, the quasiharmonic approximation result is no strict upper limit. As can be seen from the thermodynamic integration curve, the entropy of the gas decreases for increasing interaction strength, due to increasing correlations between the particles. This effect is completely missed by the entropy estimate derived from the unrelabeled trajectory, and even shows a slight increase. This failure has also been observed by others.[19]

Much better estimates are obtained from the relabeled trajectory. In particular, the entropy decrease with increasing interaction strength is now fully captured. Furthermore, quantitative agreement with the thermodynamic integration reference is obtained for weak to medium interaction strengths ($\lambda < 0.1$). However, for strong interactions, deviations remain.
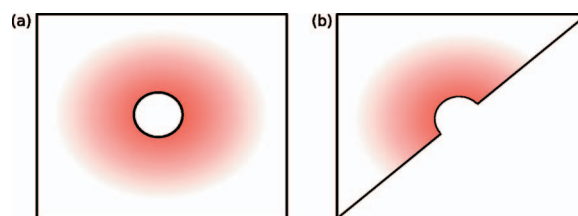

FIG. 5. (Color) Sketch of the configurational space density for an unrelabeled (a) and a relabeled trajectory (b).

Closer analysis suggests that the effect of the repulsive core is well described by our method. In contrast, the attractive part of the interaction—which becomes dominant for large λ—is described less accurately by the Gaussian approximation. Nevertheless, also for this critical interaction regime, a significantly improved estimate is achieved by exploiting the permutation symmetry.

Figure 5 illustrates this situation. The full configurational space density [Fig. 5(a)] exhibits holes due to repulsive-core interactions, which undermine the Gaussian fit. As can be shown, the centers of all these holes are located at the surface of the permutation-reduced subvolumes [Fig. 5(b)], which renders Gaussian fits more accurate.

C. Solvation of a particle in a Lennard-Jones gas

The second set of simulations aimed at characterizing the accuracy of our entropy estimate for a macromolecular solute, here mimicked by a strongly interacting Lennard-Jones particle. Figure 6 shows the excess entropies with respect to the pure "solvent" for the four methods discussed above. Additionally, a relabeled trajectory with the *last* frame of the 50 ns simulations as reference frame $\mathbf{x}_0$ is considered.

Similar to the previous system, increasing interaction strengths induce increasing correlations and thereby decrease the entropy. The PRPCA estimate captures also this effect, which is missed by the estimate from the original trajectory.
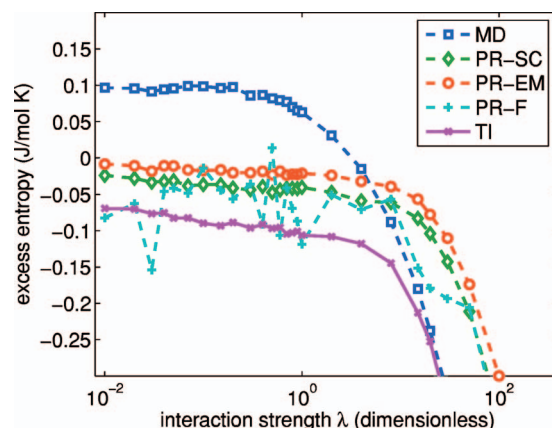

FIG. 6. (Color) Solvation of a Lennard-Jones particle solvated in a Lennard-Jones system. The interaction strength between the solvated particle and the solvent is varied from weak to very strong. Entropy estimates are obtained from the original trajectory (MD) as well as from relabeled trajectories with the simple cubic (PR-SC) and the energy minimized configuration (PR-EM), respectively, as reference structures. In an additional set of relabeled trajectories, the respective last trajectory frame we used as reference structure (PR-F). Thermodynamic integration values (TI) are given as reference.

This similar behavior is nontrivial, because the sketch in Fig. 5 cannot be applied here in a straightforward manner due to the large radius of the solvated particle.

For large interaction strengths v, self-diffusion of the first solvation shell is drastically reduced, similar to the known behavior of a protein hydration shell. As a result, an accelerated decrease of the TI entropy is seen. An even more pronounced decrease is seen for the estimate from the original trajectory. Because this estimate has been rigorously shown to provide an upper bound for any configurational space density,[17] and assuming increasing deviation from a Gaussian of the density at hand for increasing λ, the agreement with the TI entropy is likely to be due to insufficient sampling. In this case, it would be purely accidental. Indeed, insufficient convergence is seen for different trajectory lengths (data not shown). In contrast, much faster convergence is observed for the relabeled trajectory. This effect is particularly pronounced for large λ. Accordingly, the PRPCA estimates capture the marked decrease seen for the TI entropies, albeit with decreasing accuracy.

Finally, the influence of the choice of the reference structure $\mathbf{x}_0$ remains to be assessed.

To this end, we consider the entropy estimates from the relabeled trajectory which used the last frame as reference (PR-F in Fig. 6). The apparent scatter is due to the fact that here, in contrast to the other estimates, each entropy estimate rests on a *different* reference structure. Accordingly, this scatter of about $\sigma = 0.05$ J mol^{-1} K^{-1} half-width measures the uncertainty introduced by the choice of the reference structure. Interestingly, the deviation of the PRPCA estimate from the TI reference falls into this range. Optimization of the reference structure, therefore, might provide further improved entropy estimates.

V. CONCLUSION

We have presented a entropy estimation method particularly tailored for diffusive systems, such as the solvent contribution to the entropy of solvated macromolecules, e.g., proteins. As test cases, we have applied this method to a gas of 100 Lennard-Jones particles and to the solution of a strongly interacting Lennard-Jones particle in this gas. This system was chosen because its entropy is accessible to thermodynamic integration, such that a sufficiently accurate reference value can be provided.

In both test cases, our PRPCA method enhanced convergence and significantly improved the entropy estimate compared to straightforward application of the conventional PCA approach. There remain deviations from the exact result as well as from more accurate methods such as thermodynamic integration or hypothetical scanning Monte Carlo.[31] These deviations, however, have to be attributed to the chosen quasiharmonic (Gaussian) approximation for an inherently non-Gaussian configurational space density. Thus, we are now in the favorable situation that construction of a better fit function *alone* will push the accuracy of our method to the level reached by statistical approaches, which was not possible before.

The intractable sampling problem, common to all competitive methods, is thus significantly alleviated and transformed into a fitting problem that should be addressable with drastically reduced computational effort, especially in view of the fact that the characteristic shape of the permutation-reduced configurational space density should allow one to take advantage of specialized fit functions. Our PRPCA method thus opens the route to further systematic entropy estimate improvements without the computational burden of exhaustive sampling, the main obstacle for TI and Monte Carlo-type methods.

Similar developments will enable us to apply the method to more complex solvents such as water. While the permutation reduction will be fully applicable and will provide similarly enhanced sampling, the fit function here will have to be extended to additionally include the rotational and intramolecular degrees of freedom. Finally, by transforming the solvent entropy calculation into the quasiharmonic framework, the solvation entropy of solvated macromolecules such as proteins will become accessible.

An implementation of the PRPCA method is available electronically.[32]

[1] K. A. Dill, Protein Sci. **8**, 1166 (1999).
[2] N. T. Southall, K. A. Dill, and A. D. J. Haymet, J. Phys. Chem. B **106**, 521 (2002).
[3] Y. M. Rhee, E. J. Sorin, G. Jayachandran, E. Lindahl, and V. S. Pande, Proc. Natl. Acad. Sci. U.S.A. **101**, 6456 (2004).
[4] C. D. Snow, N. Nguyen, V. S. Pande, and M. Gruebele, Nature **420**, 102 (2002).
[5] B. L. de Groot and H. Grubmüller, Science **294**, 2353 (2001).
[6] T. Hansson, C. Oostenbrink, and W. F. van Gunsteren, Curr. Opin. Struct. Biol. **12**, 190 (2002).
[7] M. Karplus and J. A. McCammon, Nat. Struct. Biol. **9**, 646 (2002).
[8] J. Norberg and L. Nilsson, Q. Rev. Biophys. **36**, 257 (2003).
[9] W. F. van Gunsteren, D. Bakowies, R. Baron *et al.*, Angew. Chem., Int. Ed. **45**, 4064 (2006).
[10] Sterling Chemistry Laboratory, J. Chem. Phys. **22**, 1420 (1954).
[11] J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).
[12] T. Rodinger and R. Pomes, Curr. Opin. Struct. Biol. **15**, 164 (2005).
[13] T. Simonson, G. Archontis, and M. Karplus, Acc. Chem. Res. **35**, 430 (2002).
[14] W. F. van Gunsteren, X. Daura, and A. E. Mark, Helv. Chim. Acta **85**, 3113 (2002).
[15] C. Peter, C. Oostenbrink, A. van Dorp, and W. F. van Gunsteren, J. Chem. Phys. **120**, 2652 (2004).
[16] M. Karplus and J. N. Kushick, Macromolecules **14**, 325 (1981).
[17] J. Schlitter, Chem. Phys. Lett. **215**, 617 (1993).
[18] I. Andricioaei and M. Karplus, J. Chem. Phys. **115**, 6289 (2001).
[19] H. Schäfer, A. E. Mark, and W. F. van Gunsteren, J. Chem. Phys. **113**, 7809 (2000).
[20] W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, *Computer Simulations of Biomolecular Systems* (ESCOM Science Publishers, Leiden, 1993).
[21] M. Watanabe and W. P. Reinhardt, Phys. Rev. Lett. **65**, 3301 (1990).
[22] R. M. Levy, A. R. Srinivasan, W. K. Olson, and J. A. McCammon, Biopolymers **23**, 1099 (1984).
[23] R. Böckmann, "Molekulare Dynamik von Proteinen," Ph.D. thesis (Cuvillier Verlag, Göttingen, 2002).
[24] R. E. Burkard, Discrete Appl. Math. **123**, 257 (2002).
[25] R. E. Burkard and E. Cela, "Linear Assignment Problems and Extensions," Technical Report, 1998, Bericht No. 127, SFB Optimierung und Kontrolle.

[26] R. Jonker and A. Volgenant, Computing **38**, 325 (1987).
[27] MagicLogic, http://www.magiclogic.com/assignment.html.
[28] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide* (Biomos B.V., Zürich, Groningen, 1996).

[29] H. J. C. Berendsen, J. P. M. Potsma, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[30] E. Lindahl, B. Hess, and D. van der Spoel, J. Mol. Model. **7**, 306 (2001).
[31] R. P. White, J. Chem. Phys. **119**, 12096 (2003).
[32] F. Reinhard, J. Haas, O. Lange, and H. Grubmüller, http://www.mpibpc.mpg.de/groups/grubmueller/start/software/g_permute/.