Absolute entropies from molecular dynamics simulation trajectories

Heiko Schäfer, Alan E. Mark, and Wilfred F. van Gunsteren
Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH Zentrum, CH-8092 Zürich, Switzerland

(Received 20 March 2000; accepted 21 July 2000)

A heuristic formula for calculating absolute entropies from the covariance matrix of atom-positional fluctuations was extensively tested. Because of its heuristic nature, the results obtained are compared to analytical expressions for an ensemble of harmonic oscillators, for the ideal gas, and to numerical results obtained from the equation of state for the Lennard-Jones fluid as a means of validation of the approximate formula for the entropy. The formula yields rather accurate results. The removal of translational and rotational rigid body motion and the effect of the various fitting procedures involved are discussed for the more realistic system of a β-heptapeptide in solution. © *2000 American Institute of Physics.* [S0021-9606(00)51539-6]

I. INTRODUCTION

Molecular dynamics (MD) simulations are a well-established tool to investigate the stability and behavior of systems ranging from abstract models to complex molecular assemblies of biological interest.[1–5] The energetics of these systems is easily accessible using MD and even relative free energies are now routinely calculated, albeit at a high computational cost. However, the calculation of entropies remains complicated because the entropy depends on the whole phase space of the system of interest.[6] This is true even for entropy differences.

Most methods[7,8] for entropy calculations based on MD simulations have been restricted to the configurational entropy of a molecule as the sampling of the translation and rotation of the molecule was not complete. These methods used a transformation to internal coordinates to separate the slow overall motions from the comparably fast internal motions. Complications due to the handling of constraints were usually neglected. To calculate entropies the probability distribution of each degree of freedom (DOF) must in principle be known. However, it is notoriously difficult to sample probability distributions, especially their tails. A typical approximate solution to this problem is the assumption of a particular functional form for the probability distribution. Karplus and Kushick[7] used a Gaussian form for each internal degree of freedom (bonds, bond angles, dihedrals) and the correlation between them. Di Nola *et al.*[8] tried a combination of direct sampling for each degree of freedom and a Gaussian approximation for the correlation. They found that the inclusion of correlations was important. The Gaussian approximation to the probability distribution of each degree of freedom seemed to be appropriate.

In 1993, Schlitter[9] proposed yet another approximate method that is rationalized using the Gaussian approximation to the probability distribution but elegantly circumvents the need to express the entropy in internal coordinates. Schlitter's formula is of heuristic nature and constitutes an upper bound to the entropy. The approximate formula was rationalized by showing that the correct quantum and classical limits are obtained for the temperature approaching zero or infinity.[9] The numerical tests presented in Ref. 9 were, unfortunately, of limited value. Here, Schlitter's heuristic formula is extensively tested with the aim of its validation. Second, as the method allows us to use Cartesian coordinates and does not restrict us to the configurational entropy, it is used to investigate entropic effects in the reversible folding of a small peptide.[5] We have implemented Schlitter's formula to work with the GROMOS96 simulation package.[10,11]

II. METHOD

Schlitter[9] introduced a very elegant formula to calculate absolute entropies from MD trajectories using the covariance matrix of atom-positional fluctuations. In this section we summarize the heuristic derivation of Schlitter's formula in order to clarify the approximations on which it is based.

The formula is based on a quantum-mechanical treatment of a one-dimensional degree of freedom x with states $|n\rangle, n=0,1,2,\ldots$. The energy of each state is ϵ_n, the mass of x is m, and its mean $\langle x\rangle$ is assumed to be zero. The canonical partition function at temperature T is

$$Z=\sum_n \exp(-\beta\epsilon_n), \tag{1}$$

with $\beta=1/k_BT$, the Boltzmann constant k_B and temperature T. The complete entropy S of the system can be expressed using the probability $p_n=\exp(-\beta\epsilon_n)/Z$ of finding the system in state n as

$$S=-k_B\sum_n p_n \ln p_n . \tag{2}$$

For a given expectation value (ensemble average) of the variance $\langle x^2\rangle=\Sigma_n p_n\langle n|x^2|n\rangle$, the entropy is maximized by varying the probabilities p_n. The variance $\langle x^2\rangle$ and the normalization of the probabilities $\Sigma_n p_n=1$ are taken into account as constraints using Lagrange multipliers.

This results in the demand that the energy of each state must be proportional to the state's variance: $\epsilon_n\propto\langle n|x^2|n\rangle$. A system for which this condition is fulfilled is the simple har-

monic oscillator (SHO). The entropy S_{sho} of a one-dimensional simple harmonic oscillator can be expressed analytically as

$$S \leqslant S_{\mathrm{sho}} = \frac{k_B \alpha}{e^{\alpha}-1} - k_B \ln[1-e^{-\alpha}], \tag{3}$$

where $\alpha = \hbar\omega/k_B T$, $\hbar = h/2\pi$, ω is the frequency of the oscillator, and h is Planck's constant. It was shown by Schlitter[9] that the entropy of the harmonic oscillator S_{sho} is an upper bound for the *true* entropy S of the system.

The frequency ω of the oscillator depends on the quantum-mechanically defined variance $\langle x^2 \rangle$ and must be connected to the classical variance $\langle x^2 \rangle_c$ that can be measured in classical simulations. As a connection, the equipartition theorem

$$m\omega^2 \langle x^2 \rangle_c = k_B T \tag{4}$$

is used. The equipartition theorem holds only in the limit $\hbar\omega \ll k_B T$. This approximation is likely to be good because in a molecule the high-frequency motions for which it fails will contribute very little to the entropy.

In the generalization of the formula to many degrees of freedom it is simpler to use an approximation to Eq. (3) which was introduced and rationalized by Schlitter,[9]

$$S \leqslant S_{\mathrm{sho}} < S' = \frac{1}{2} k_B \ln\left(1 + \frac{e^2}{\alpha^2}\right) \tag{5}$$

$$= \frac{1}{2} k_B \ln\left(1 + \frac{k_B T e^2}{\hbar^2} m \langle x^2 \rangle_c\right). \tag{6}$$

Here $e = \exp(1)$ is Euler's number.

A. Many degrees of freedom

The generalization to many degrees of freedom is based on the covariance matrix $\boldsymbol{\sigma}$ of the atom-positional fluctuations with the elements

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle, \tag{7}$$

where $x_1, \ldots x_{3N}$ are the Cartesian coordinates of an N particle system. The covariance matrix can also be expressed in mass-weighted coordinates $x_i' = x_i \sqrt{m_i}$ and becomes

$$\boldsymbol{\sigma}' = \mathbf{M}^{1/2} \boldsymbol{\sigma} \mathbf{M}^{1/2} = \mathbf{M}\boldsymbol{\sigma}, \tag{8}$$

where $\mathbf{M} = \mathbf{M}^{1/2}\mathbf{M}^{1/2}$ is the mass matrix of rank $3N$ in which the diagonal elements hold the masses and $m_{ij} = 0$ for $i \neq j$. The last identity holds because $\boldsymbol{\sigma}$ and $\mathbf{M}^{1/2}$ are both symmetric matrices.

The mass-weighted covariance matrix can be diagonalized, giving a new set of *uncorrelated* (the off-diagonal elements are 0) coordinates q_i. For each of these new degrees of freedom the entropy can be calculated with Eq. (6) and the variances $\langle q_{ii}^2 \rangle_c$ which are the diagonal elements of the matrix $\boldsymbol{\sigma}'$ (expressed in the coordinates q_i),

$$S < S' = \frac{1}{2} k_B \sum_{i=1}^{3N} \ln\left[1 + \frac{k_B T e^2}{\hbar^2} \langle q_{ii}^2 \rangle_c\right]$$

$$= \frac{1}{2} k_B \ln\left(\prod_{i=1}^{3N} \left[1 + \frac{k_B T e^2}{\hbar^2} \langle q_{ii}^2 \rangle_c\right]\right). \tag{9}$$

Taking the product of the diagonal elements of a diagonal matrix as in Eq. (9) is equivalent to calculating the determinant of the matrix $\boldsymbol{\sigma}'$. Because the determinant of a matrix is invariant under any orthogonal transformation Eq. (9) can be rewritten as[9]

$$S' = \frac{1}{2} k_B \ln \det\left[\mathbf{1} + \frac{k_B T e^2}{\hbar^2} \mathbf{M}\boldsymbol{\sigma}\right]. \tag{10}$$

Therefore, the transformation to an internal, non-Cartesian set of coordinates is not necessary. To calculate the entropy using Eq. (10), it is only necessary to calculate the covariance matrix [Eq. (7)] from a trajectory. The correct quantum-mechanical limit for high-frequency motion is obtained with Eq. (10), because a covariance matrix whose elements are all vanishingly small will give an entropy of zero.

Very slow motions like the center of mass motion of a larger molecule can lead to a constantly increasing variance and therefore to a convergence problem in the entropy calculations based on Eq. (10). A simple translational fit on the centers of mass of the molecules at the various time points, however, will remove the center of mass motion and lead to more rapidly converging results. The missing entropy contribution (the ideal gas contribution) can be calculated analytically, if required.

B. Approximations

Several approximations are used in the derivation of the formula described above:

(1) Every degree of freedom is treated as a quantum harmonic oscillator,
(2) The equipartition theorem is used to connect the classical variance and the frequency of a quantum harmonic oscillator [Eq. (4)], and
(3) An approximate expression [Eq. (5)] for the entropy of a quantum harmonic oscillator is used.

As stated above, the second approximation is not expected to give rise to any significant errors, as high-frequency motions will contribute little to the entropy. The third approximation can also be shown to be very good over the whole range of molecular frequencies ω.[9]

The first approximation might break down depending on the type of system that is examined. Bond vibrations, angle-bending, and torsional-angle motions in a molecule are probably well described by a harmonic approximation. The motion of a molecule inside a solvent, on the other hand, is probably a bad case for an harmonic approximation. Section IV B examines these effects in detail for the entropy of a Lennard-Jones fluid.

III. STOCHASTIC DYNAMICS OF HARMONIC OSCILLATORS

The simplest system to test Schlitter's formula is an ensemble of independent classical harmonic oscillators. The results can be calculated analytically and checked against the results from simulations.

One hundred completely independent, noninteracting particles of mass 15.994 amu were harmonically restrained

TABLE I. Entropy of an ensemble of 100 independent three-dimensional harmonic oscillators. Values from the stochastic dynamics simulation are compared to analytical results. (Exact method [Eq. (3)] and Schlitter's approximation [Eq. 5].)

Method	S_{sho} J K^{-1} mol^{-1}
SD simulation:	
full covariance matrix	33.449
only diagonal elements	36.897
Analytical calculations:	
Exact using Eq. (3)	36.9782
Schlitter using Eq. (5)	36.9784

TABLE II. Molar entropy of an ideal gas (mass 16 amu) in J K^{-1}mol^{-1} at different temperatures. Exact analytical values obtained using Eq. (A3) are compared to analytical values using Schlitter's approximation, Eq. (A4).

T K	S_{ideal} J K^{-1} mol^{-1}		Error
	exact Eq. (A3)	Schlitter Eq. (A4)	
100	690.774	810.044	17.2%
200	699.419	818.689	17.1%
300	704.476	823.746	16.9%

from leaving their initial position with a force constant of 25 kJ mol^{-1} nm^{-2}. They were given initial velocities according to a Maxwell–Boltzmann distribution at 300 K. A 1 ns stochastic dynamics simulation with a friction coefficient of 10 ps^{-1} at a temperature of 300 K was performed using the GROMOS simulation package.[10,11] The time step was set to 0.01 ps.

This setup yields a system of 300 independent, one-dimensional classical harmonic oscillators whose entropy can be calculated analytically using Eq. (3) (exact expression) or Eq. (5) (Schlitter's approximation).

Taking the generated trajectory the entropy can be calculated using the covariance matrix [Eq. (10)]. When using the full covariance matrix the entropy will be slightly lowered because of spurious correlations that arise from the finite numerical accuracy and finite simulation length. These correlations are in this case an artifact and unwelcome when comparing to analytical results for truly independent oscillators with zero correlations. The spurious correlations are omitted by using only the diagonal elements of the covariance matrix in the calculation. The results are summarized in Table I.

The exact result [Eq. (3)] is slightly smaller than the result from Schlitter's approximation [Eq. (5)], showing that the latter approximation is always an upper bound to the exact value. The results of the simulation ignoring spurious correlations agree well with the analytical results.

IV. SIMPLE MANY-PARTICLE SYSTEMS

A. The ideal gas

For the ideal gas it is possible to calculate the entropy analytically, either through classical statistical mechanics or using Schlitter's approximation (see Appendix A).

The molar entropy of an ideal gas with a volume of 22.41/mol was calculated for a range of temperatures using both methods (see Table II). Clearly Schlitter's method gives an upper bound to the exact value. The ideal gas represents the worst case for the harmonic approximation used in Schlitter's formula explaining the relatively large errors. The error is only slightly decreasing with increasing temperature.

B. The Lennard-Jones fluid

The Lennard-Jones fluid is used as a more realistic test system for which numerical values for the entropy are still available for comparison. A system of 256 argon atoms using the parameters of the GROMOS96 force field[10] version 43A1 ($\epsilon/k_B = 119.8$ K, $\sigma = 0.341$ nm) was set up in a cubic box with an edge length of 2.3 nm (reduced density $\rho^* = 0.834$). The system was simulated (time step = 2 fs) at constant volume (volume = 12.2 nm^3) and constant temperatures: $T = 200$, 300, 400 K. At each temperature the systems were simulated for 200 ps and the entropy calculated over the whole 200 ps trajectory. The entropy was calculated ignoring correlations ($S^{sim}_{lj,nc}$) and with all correlations included ($S^{sim}_{lj,cr}$).

The numerical results obtained from the simulations are compared to numerical results which were obtained by calculating the entropy of an ideal gas (S^{exact}_{id}) exactly and afterwards correcting for the interaction between the particles. This correction, the residual entropy of a Lennard-Jones fluid (S^{eos}_{corr}), was calculated numerically using a polynomial fit to the equation of state of the Lennard-Jones fluid.[12]

Table III shows the results. In the case of complete neglect of any correlation between the particles ($S^{sim}_{lj,nc}$), the entropy of the Lennard-Jones fluid is very similar to the entropy of an ideal gas (S^{exact}_{id}). Taking the correlation into account, the entropy ($S^{sim}_{lj,cr}$) is drastically lowered. This again illustrates that for the ideal gas the harmonic approximation is poor.

The difference between the entropy $S^{sim}_{lj,cr}$ from the simulation including all correlations and the corrected entropy S^{exact}_{lj} calculated numerically from the equation of state is also given in Table III and ranges between 6.9% and 3.7%. The difference is significantly smaller than in the case of the ideal gas.

V. ENTROPY CALCULATIONS FOR A β-HEPTAPEPTIDE IN SOLUTION

The ensemble of harmonic oscillators and the Lennard-Jones fluid are highly idealized test systems in which a number of important aspects of more realistic systems are not addressed, e.g., presence of internal degrees of freedom of a molecule. Therefore, the formula was applied to a peptide in solution which has been extensively studied and characterized.[5] A β-heptapeptide in methanol was simulated at different temperatures. The peptide undergoes reversible folding in the simulations and the ratio of folded to unfolded structures decreases with increasing temperature from 49 to 1 at 298 K to 1 to 3 at 360 K. For 298 and 340 K, 200 ns trajectories, and for 350 and 360 K, 50 ns trajectories, were available. The first three simulations were done in a rectan-

TABLE III. Entropy of argon at different temperatures calculated analytically [ideal gas Eq. (A3) and harmonic Eq. (A4)], calculated numerically using a polynomial fit to the equation of state (S_{lj}^{exact}) and calculated numerically [using Eq. (10), nc=no correlations included, cr=correlations included] from MD simulations of the Lennard-Jones fluid. Entropies are in $J\,K^{-1}mol^{-1}$, T^* is the temperature in reduced units.

T [K] (T^*)	200 (1.668)	300 (2.503)	400 (3.337)
Analytical results			
S_{id}^{exact} Eq. (A3) [$J\,K^{-1}mol^{-1}$]	131.42	136.47	140.06
S_{id}^{harm} Eq. (A4) [$J\,K^{-1}mol^{-1}$]	135.82	140.88	144.46
Correction from equation of state			
S_{corr}^{eos} [$J\,K^{-1}mol^{-1}$] ($S_{corr}^{eos}*$)	−22.840 (−2.747)	−20.071 (−2.414)	−18.292 (−2.200)
Numerical result from equation of state			
$S_{lj}^{exact}=S_{id}^{exact}+S_{corr}^{eos}$ [$J\,K^{-1}mol^{-1}$]	108.58	116.399	121.768
Numerical result from simulation			
$S_{lj,nc}^{sim}$ [$J\,K^{-1}mol^{-1}$]	132.08	138.48	142.76
$S_{lj,cr}^{sim}$ [$J\,K^{-1}mol^{-1}$]	101.14	110.79	117.24
Difference in %			
nc (correlations ignored)	21.6	19.0	17.2
cr (correlations included)	6.9	4.8	3.7

gular box containing 962 methanol molecules. The 360 K run involved a truncated octahedron with 1778 methanol molecules.[13]

Figure 1 shows the absolute entropy for all four temperatures. Since the rigid body motion is not removed automatically during the calculation, three different fitting procedures are studied:

(1) No fit, all rigid body motion (translation and overall rotation) is kept,
(2) A translational fit, the centers of mass of the molecules are mapped on top of each other (translation removed), and
(3) A least-squares fit on the positions of all atoms, i.e., rigid body (translation, rotation) motion is removed.

All the fits are performed with reference to the first configuration of the analyzed trajectory.

The steplike increase in the entropy in Fig. 1 is indica-

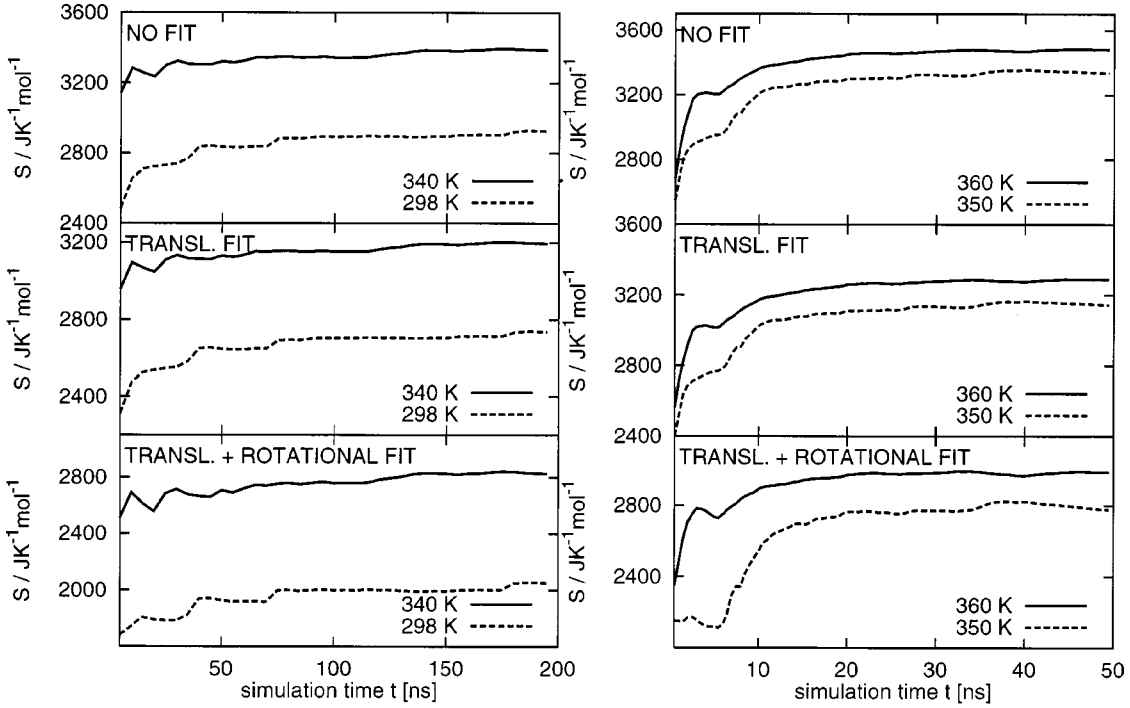

FIG. 1. Entropy of the β-heptapeptide at different temperatures and using different fitting procedures. Upper panels: no fitting is performed; middle panels: the centers of mass of the trajectory configurations are superimposed (no overall translation); lower panels: in addition, a rotational least-squares fit on the atom positions is carried out for the trajectory configurations (no overall translation and rotation). On the left are the 200 ns simulations at 298 and 340 K, on the right the 50 ns simulations at 350 and 360 K.

TABLE IV. Diffusion coefficient D; translational contribution to the entropy S^{trans} for the β-heptapeptide and entropy (S) and enthalpy (H) change upon folding: comparison between results from the simulation $S^{\text{trans}}_{\text{sim}}$ and analytical values $S^{\text{trans}}_{\text{id}}$. V is the average volume of the simulation box.

T K	V nm^3	D nm^{2}ns^{-1}	$S^{\text{trans}}_{\text{sim}}$ J K^{-1}mol^{-1}	$S^{\text{trans}}_{\text{id}}$ J K^{-1}mol^{-1}	$S^{\text{fold}}-S^{\text{unfold}}$ J K^{-1}mol^{-1}	$H^{\text{fold}}-H^{\text{unfold}}$ kJ mol^{-1}
298	61.665	0.7	189.94	186.77	−527	−71
340	66.072	1.7	192.27	188.99	−608	−87
350	66.925	1.6	191.69	189.46	−768	−56
360	124.340	1.9	195.62	194.96	−873	−92

tive of the exploration of phase space by the peptide: each jump opens up a new region of phase space. All simulations were started in the folded (helical) configuration except for the simulation at 360 K. This difference can be seen when comparing 350 and 360 K (Fig. 1, bottom panel at right-hand side): while the folded peptide (350 K) stays in its native conformation for 7 ns, the extended structure (360 K) experiences a quick increase in entropy because it explores phase space more quickly. On the other hand at 340 K (Fig. 1, bottom panel at left-hand side) the peptide unfolds within 2 ns.[13]

A. Rigid body motion: Translation

By using the fitting procedures described above, a separation of the rigid body motion from the internal motion can be attempted. The translation of the molecules can be separated exactly by a translational fit on the center of mass. The difference between the entropy calculated without fitting and the entropy calculated with the translational fit yields the translational entropy contribution S^{trans}. The translational entropy increases monotonically with the volume that is sampled [see Eq. (A3)] and levels off if the volume is finite. When applying periodic boundary conditions in a simulation, the effective volume is finite since the molecule is set back into the box when it crosses the (periodic) box edge.

Table IV compares the results of the simulation, $S^{\text{trans}}_{\text{sim}}$, and the analytical results, $S^{\text{trans}}_{\text{id}}$. The volume of the simulation box V is also given. $S^{\text{trans}}_{\text{id}}$ is the ideal gas contribution to the absolute entropy of the β-heptapeptide (see Appendix A). Both values agree very well, which indicates that the complete volume of the box was sampled in the simulation. The results from the simulation $S^{\text{trans}}_{\text{sim}}$ are consistently larger than the analytical results because Schlitter's approximation always gives an upper bound to the true entropy. The rather complete sampling of the box was confirmed by projecting the position of the center of mass of the β-heptapeptide from the whole trajectory onto the three orthogonal box axes (Fig. 2). Only results for the three temperatures that were simulated in a rectangular box are shown; the fourth, simulated in a truncated octahedron, is omitted. All three temperatures show a flat profile for all three axes, confirming complete sampling of the simulation box.

With such extensive sampling a very reliable estimate of the diffusion constant D can be made. The diffusion constant

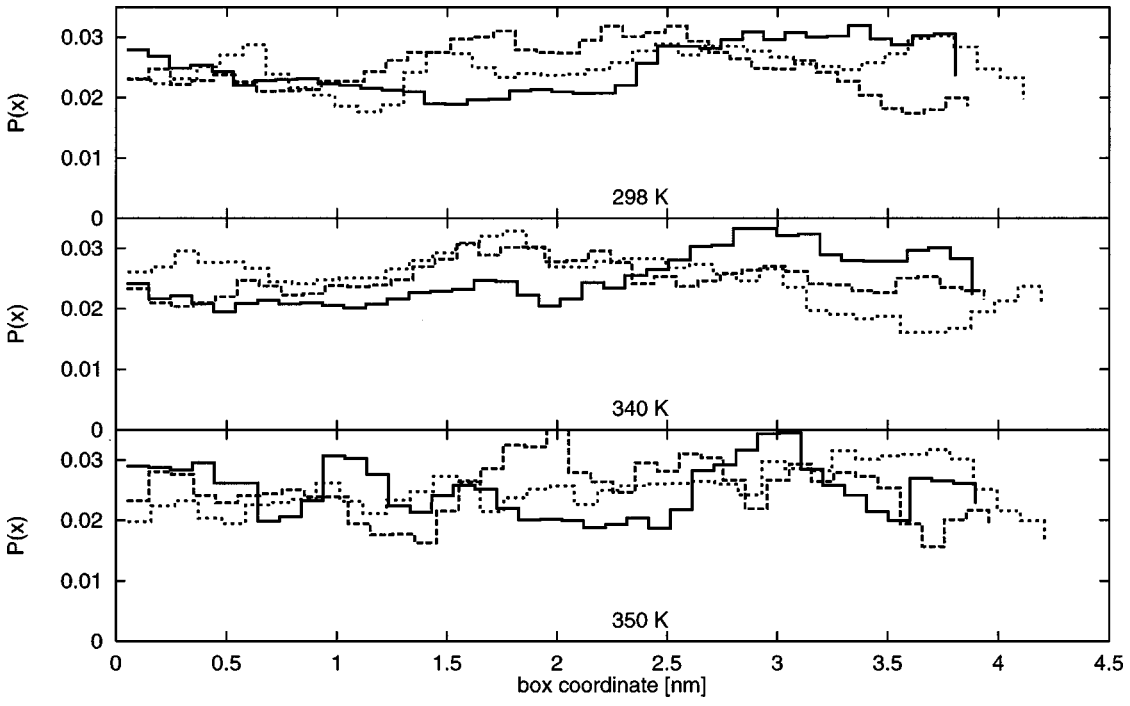

FIG. 2. Distribution $P(x)$ of the center of mass of the β-heptapeptide projected onto the x, y, and z axes of the simulation box.

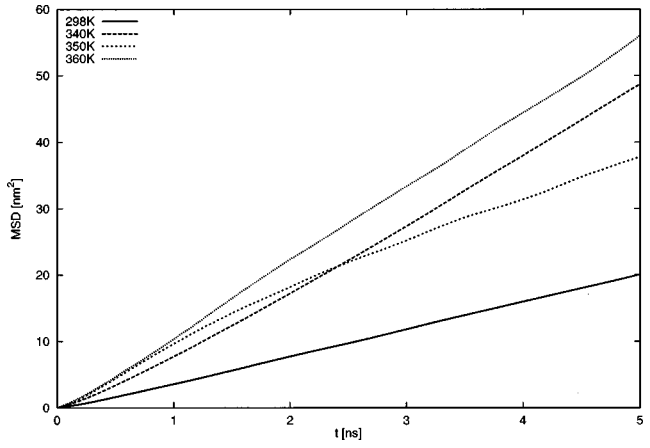

FIG. 3. Mean-square displacement (MSD) of the center of mass of the β-heptapeptide as a function of time.

was calculated from the mean-square displacement of the center of mass using the Einstein relation,[1]

$$2tD = \tfrac{1}{3}\langle|\mathbf{r}(t)-\mathbf{r}(0)|^2\rangle_t . \tag{11}$$

Here $\mathbf{r}_i(t)$ is the position of the center of mass at time t. The mean-square displacement is shown in Fig. 3. It shows the expected linear relation at short times where the statistics is best. Using a linear regression in the linear regime at short time scales the diffusion constants for all four temperatures were calculated, see Table IV.

B. Rigid body motion: Rotation

In contrast to the separation of the translational motion, the separation of the overall rotation from the internal motion is not unambiguous for flexible molecules. Overall rotation and internal motion are highly coupled and a rotational least-squares fit will only make sense for more rigid molecules.

Two different sets of atoms were used in the least-squares fit to check the dependence of the results on the fitting procedure:

(1) All 64 atoms of the molecule were used with equal weights.
(2) Only four atoms in the central residue (29-4N, 31-4CB, 32-4CG, 33-4CA) of the peptide were used.

The results are compared in Fig. 4. Naively one would expect a decrease in entropy with an increase in the number of atoms used in the fit. In the case of 298 K this is true; the entropy calculated after performing a least-squares fit for all atoms is smaller compared to the entropy where only four atoms were used in the fit. For the other temperatures the entropy is lower in the beginning for the all-atom fit but higher afterward. At 298 K the peptide spends nearly all of the time (98%) in the folded configuration[5] and the least-squares fit makes sense: By using more atoms in the fitting, the entropy is lowered. With increasing temperature the peptide unfolds, sampling more diverse configurations that cannot be fitted easily onto the first configuration in the trajectory. This can increase the entropy compared to a fit based on a few core atoms. For example, fitting all the atoms of an extended structure on to the rather spherical folded structure of a short helix might artificially induce rotation and thus induce considerable spurious internal motion.

There is no correct, unambiguous way to separate the overall rotation from the internal motion and Fig. 4 gives an estimate of 50 to 80 J K^{-1} mol^{-1} of the uncertainty in the entropy calculation due to arbitrariness of the choice of atoms in the fitting procedure.

C. Rotational correlation times

To investigate to what extent the overall rotation of the β-heptapeptide is sampled and to estimate the rotational cor-

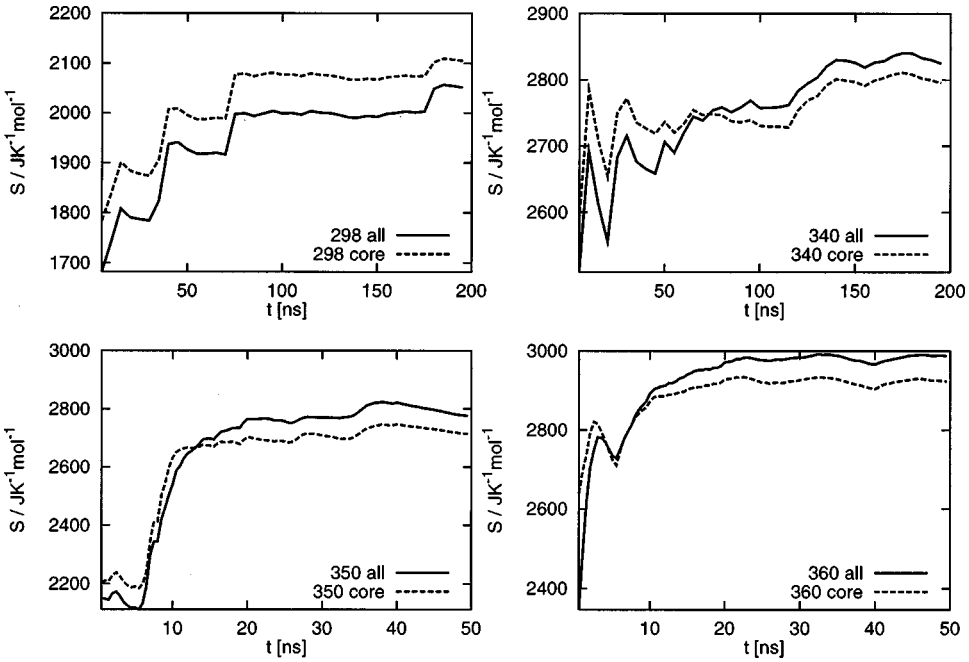

FIG. 4. Entropy of the β-heptapeptide at all four temperatures after a translational and rotational least-squares fit using all atoms (dashed line) and using only four atoms in the core of the peptide (solid line).

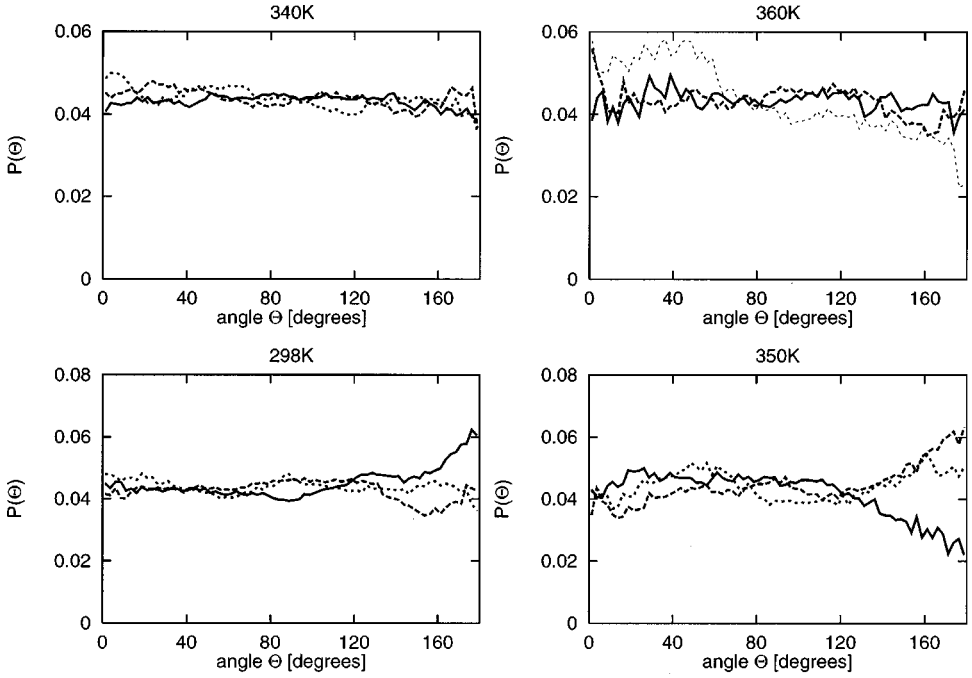

FIG. 5. Distribution $P(\Theta)$ of the angles Θ between the axis of rotation which corresponds to the largest moment of inertia of the β-heptapeptide and the x, y, and z axes of the simulation box.

relation function, the tensor of inertia $\mathbf{I}$ was calculated for each configuration in the trajectory. The elements of the inertia tensor $\mathbf{I}$ are given by

$$I_{ij}=\sum_{k=1}^{N} m_k(\mathbf{x}^{(k)^2}\delta_{ij}-x_i^{(k)}x_j^{(k)}). \tag{12}$$

Here $x_i^{(k)}$ is component i of vector $\mathbf{x}^{(k)}$, which is the vector from the center of mass to atom k and m_k is the mass of atom k. $\mathbf{I}$ is thus a three-dimensional quadratic symmetric matrix. Diagonalization of $\mathbf{I}$ yields the three moments of inertia (eigenvalues) and the three principal axes of rotation (eigenvectors).

Figure 5 shows the distribution of the angles Θ between the axis of rotation which corresponds to the largest moment of inertia and the three edges of the simulation box. The results were weighted with $1/\sin\Theta$. There are only two possible orientations for an axis to be parallel to a particular edge of the simulation box, while there is a multitude of orientations for an axis to be perpendicular to a particular edge. The rather flat profiles in Fig. 5 confirm a complete sampling of the overall rotation of the molecule.

The rotational correlation times can be estimated from the rotational correlation functions

$$C^{(l)}(t)=\langle P_l[\mathbf{u}_i(\tau+t)\cdot\mathbf{u}_i(\tau)]\rangle, \tag{13}$$

where $\mathbf{u}_i(t)$ is a unit vector along one of the three axes of rotation at time t and $P_l(x)$ is the lth Legendre polynomial. $C^{(1)}(t)$ and $C^{(2)}(t)$ are shown in Fig. 6 for the principal axis of rotation which corresponds to the largest moment of inertia and for all four temperatures. The figure shows that the rotational correlation time decreases with higher temperature. The 350 and 360 K simulations are very similar, which is also evident from all other results. The logarithmic plot shows a linear regime from which the correlation times can be calculated using a linear regression. The correlation times τ_1 from P_1 vary between 80 and 175 ps, while the correlation times τ_2 from P_2 vary between 50 and 130 ps.

VI. ENTROPY OF FOLDING

The configurations of the MD trajectory were identified as folded or unfolded conformations using an atom-positional root-mean-square deviation (RMSD) criterion for the backbone atoms of residues 2-6: a peptide configuration with an RMSD with respect to the helical NMR model structure smaller than 0.1 nm is considered to be folded, a configuration with this RMSD larger than 0.15 nm as unfolded. Using only folded configurations in the entropy calculation yields the entropy of the folded state; the unfolded configurations yield the unfolded state. The entropy change of the peptide upon folding is given in Table IV. The loss in entropy upon folding of the peptide is substantial and increases with increasing temperature. The enthalpy change of the peptide upon folding is also given in Table IV. It is negative, but not sufficiently negative to compensate for the loss of peptide entropy upon folding. The free energy change of the peptide is positive for all four temperatures. Yet, according to the relative population of folded versus unfolded conformations present in the MD trajectory, the free energy of folding of the complete system, i.e., peptide and solvent, is negative at 298 K and slightly positive (a few kJ mol^{-1}) at the other temperatures. This means that the change in free energy of the peptide alone upon folding cannot explain the observed folding behavior. An increase in the entropy of the solvent and a loss in peptide–solvent correlation seem to contribute non-negligibly to the folding process. This underlines the important role of the solvent in peptide folding.

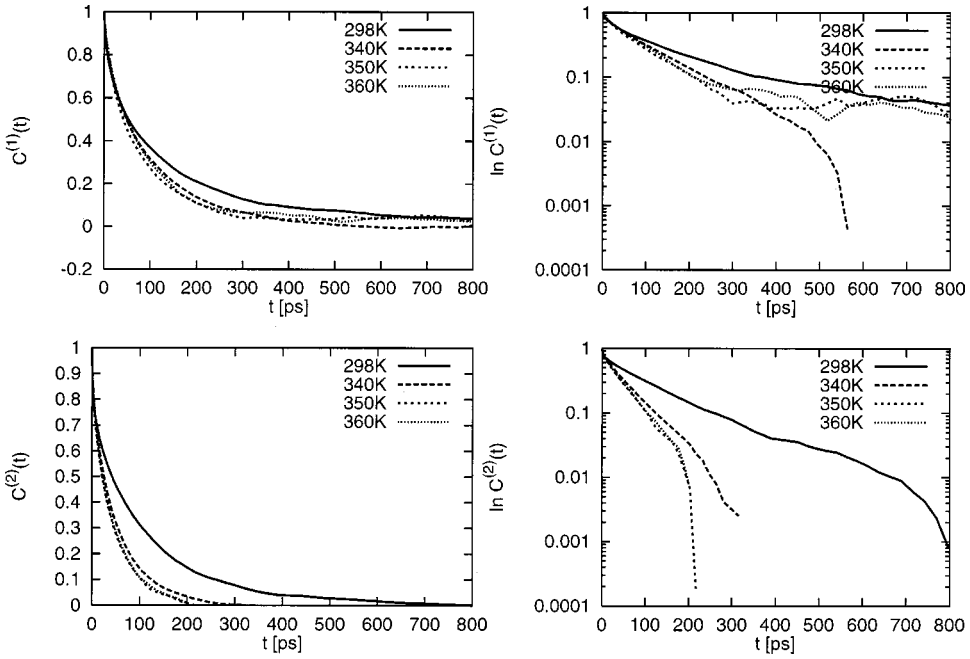

FIG. 6. Rotational correlation functions (1st and 2nd Legendre polynomial) of the first eigenvector of the inertia tensor of the β-heptapeptide. On the right the logarithmic plot shows the linear regime at small time scales.

VII. CONCLUSIONS

The formula proposed by Schlitter provides an elegant way to calculate entropies from molecular dynamics trajectories. It has been shown to give generally good results and only the harmonic approximation is critical. In the case of the ensemble of harmonic oscillators this approximation is fulfilled and the simulation results are in perfect agreement with the analytical results. The ideal gas represents the worst case for the harmonic approximation and Schlitter's formula yields errors around 17% for typical temperatures that are used in simulations. These are still acceptable errors considering the quality and convergence problems of the calculation of any entropic quantity. For interacting particles the harmonic approximation does better, even for the Lennard-Jones fluid where the potential is rather anharmonic. The error is reduced to around 5%.

In the case of the β-heptapeptide different contributions to the entropy could be calculated by using different overall translational, rotational fitting schemes. It could be shown that with sufficient sampling Schlitter's method is not restricted to conformational entropies but that translational and rotational contributions can also be calculated. The effect of the fitting procedure on the entropy cannot be neglected and a consistent approach should be used to make the results comparable. The possibility of neglecting the correlation between atoms makes the investigation of the cooperative nature of movements in molecules possible. For systems of the size of small peptides, entropy calculations are viable using Schlitter's method and will help to give new insights into their behavior.

The entropy and enthalpy change of the peptide upon folding are both negative at all four temperatures considered, and lead to a positive free energy change of the peptide upon folding. Since the peptide is predominantly folded in the simulation at room temperature, this result means that changes in the solvent entropy and peptide–solvent correlation cannot be ignored when explaining peptide folding.

ACKNOWLEDGMENT

We would like to thank Xavier Daura for providing the trajectories of the heptapeptide simulations and Emile Apol for helpful scripts and discussions about the equation of state of the Lennard-Jones liquid.

APPENDIX A: ENTROPY OF AN IDEAL GAS

The translational part of the single-particle partition function q is given by

$$q_{\text{trans}}=\left(\frac{2\pi m k_B T}{h^2}\right)^{3/2} V, \tag{A1}$$

where m is the mass and V is the volume of the box in which the particle moves. If we take an ensemble of N noninteracting distinguishable particles the canonical partition function is $Q=q^N$. The entropy in the canonical ensemble is given by

$$S=\frac{E}{T}+k_B \ln Q=k_B T\left(\frac{\partial \ln Q}{\partial T}\right)_V+k_B \ln Q, \tag{A2}$$

where E is the energy difference between T and absolute zero temperature. The molar entropy for an ideal gas is therefore given by

$$S_{\text{ideal}}=\frac{3}{2}R+R \ln\left[\left(\frac{2\pi m k_B T}{h^2}\right)^{3/2} V\right]. \tag{A3}$$

It is also possible to calculate the entropy of an ideal gas using Schlitter's approximation Eq. (5). First the variance $\langle x^2\rangle_c$ must be calculated. For the ideal gas the distribution of positions of a particle in a box should be uniform in each of the three dimensions. Therefore the normalized probability $P(x)$ of finding a particle at position x is a constant: $P(x)$

$=L^{-1}$, assuming a cubic box of length L. Assuming a mean $\langle x\rangle=0$ we get the variance $\langle x^2\rangle_c=L^2/12$. The covariance matrix is given by $\boldsymbol{\sigma}=\mathbf{1}\langle x^2\rangle_c$ and Eq. (10) simplifies to

$$S'=\frac{3Nk_B}{2}\ln\left(1+\frac{L^2m_ik_BTe^2}{12\hbar^2}\right). \tag{A4}$$

[1] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford Science Publications, 1989).
[2] W. F. van Gunsteren and H. J. C. Berendsen, Angew. Chem. Int. Ed. Engl. **92**, 992 (1990).
[3] M. Karplus and G. A. Petsko, Nature (London) **347**, 631 (1990).
[4] D. P. Tieleman, S. J. Marrink, and H. J. C. Berendsen, Biochim. Acta **1331**, 235 (1997).
[5] X. Daura, W. F. van Gunsteren, and A. E. Mark, Proteins **34**, 269 (1999).
[6] A. Mark, Encycl. Comput. Chemi. **2**, 1070 (1998).
[7] M. Karplus and J. N. Kushick, Macromolecules **14**, 325 (1981).
[8] A. Di Nola, H. J. C. Berendsen, and O. Edholm, Macromolecules **17**, 2044 (1984).
[9] J. Schlitter, Chem. Phys. Lett. **215**, 617 (1993).
[10] 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* (Vdf Hochschulverlag, Zürich, 1996), ISBN 3 7281 2422 2.
[11] W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, and W. F. van Gunsteren, J. Phys. Chem. A **103**, 3596 (1999).
[12] J. K. Johnson, J. A. Zollweg, and K. E. Gubbins, Mol. Phys. **78**, 591 (1993).
[13] X. Daura, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark, J. Mol. Biol. **280**, 925 (1998).