Estimating entropies from molecular dynamics simulations

Christine Peter, Chris Oostenbrink, Arthur van Dorp, and Wilfred F. van Gunsteren[a)]
Laboratorium für Physikalische Chemie, ETH Zürich, CH-8093 Zürich, Switzerland

(Received 28 August 2003; accepted 30 October 2003)

While the determination of free-energy differences by MD simulation has become a standard procedure for which many techniques have been developed, total entropies and entropy differences are still hardly ever computed. An overview of techniques to determine entropy differences is given, and the accuracy and convergence behavior of five methods based on thermodynamic integration and perturbation techniques was evaluated using liquid water as a test system. Reasonably accurate entropy differences are obtained through thermodynamic integration in which many copies of a solute are desolvated. When only one solute molecule is involved, only two methods seem to yield useful results, the calculation of solute–solvent entropy through thermodynamic integration, and the calculation of solvation entropy through the temperature derivative of the corresponding free-energy difference. One-step perturbation methods seem unsuitable to obtain entropy estimates. © *2004 American Institute of Physics.* [DOI: 10.1063/1.1636153]

I. INTRODUCTION

Computing free energies and entropies is one of the main goals and major efforts in the field of MD simulation, since these properties are among both the most sought-after and most elusive properties of a system. Knowledge about free energies is essential to understand the direction of any chemical process as well as the composition of any equilibrium, while knowledge about entropies contributes to our understanding of chemical processes, e.g., of the driving forces of the folding of biomolecules or the binding of ligands. Entropy is the key property to understand chemical phenomena such as hydrophobic interactions. Unfortunately, these two statistical properties are not that easily accessible to simulations, as their evaluation involves a measurement of the extent of phase space available to the system, and so in principle requires an infinitely long simulation to scan the entire space. Nevertheless, a sizable effort has so far been invested in studying free-energy differences, showing that it is, in spite of all sampling problems, possible to obtain reasonable free-energy estimates.[1–7] The reason is, that in order to obtain free-energy differences between two states, the evaluation of the complete partition function is not really necessary, but extensive sampling of the relevant parts where the two states differ, suffices. In contrast to free-energy differences, determining both absolute entropies and entropy differences from MD simulations requires in principle sampling of the complete phase space. Generally, one can distinguish two types of attempts to obtain reasonable estimates for entropies by MD simulations. One type of methods focuses on conformational entropies, where not all, but only internal (conformational) nondiffusive degrees of freedom, for example, in a protein, are considered.[8–12] The other type of method relies on the principles of statistical thermodynamics, where, usually in analogy to the methods to determine free-energy differences, formulas for entropy differences between two states can be derived.[13,14] The first type of methods has been successfully applied to proteins and other biomolecular systems, e.g., to estimate sidechain entropies.[15] A study of the internal entropy of the unfolded and folded conformations of a small peptide showed though, that knowing these conformational entropies is not sufficient to understand the underlying forces that drive peptide and protein folding,[16] but that inclusion of solvent degrees of freedom is necessary. The second type of approach which extends techniques to estimate free-energy differences to entropies, has been applied very rarely so far,[17–21] the reason being that they involve an accurate estimate of an ensemble average that includes the complete Hamiltonian of the system which shows large fluctuations and therefore takes a very long simulation time to converge. Here, several methods to estimate entropy differences are presented and tested for a pure water system, namely by calculation of the excess and hydration entropies of a widely used water model, the simple point charge (SPC) model.[22]

II. THEORY

A. Basic statistical mechanics formulas and notation used

The classical Hamiltonian for a system of N atoms, expressed in their Cartesian coordinates $\mathbf{r}$ and conjugate momenta $\mathbf{p}$ reads

$$\mathcal{H}(\mathbf{p},\mathbf{r}) = E_{\mathrm{kin}}(\mathbf{p}) + E_{\mathrm{pot}}(\mathbf{r}) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + E_{\mathrm{pot}}(\mathbf{r}), \tag{1}$$

$E_{\mathrm{kin}}(\mathbf{p})$ and $E_{\mathrm{pot}}(\mathbf{r})$ being the kinetic and the potential energy of the system. The partition function in the canonical ensemble is given by

$$Q(N,V,T) = \frac{\int\int \exp(-\mathcal{H}(\mathbf{p},\mathbf{r})/k_B T)\,\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}}{h^{3N} N!}, \tag{2}$$

a)Author to whom correspondence should be addressed.
Phone: +41 1 632 5501; Fax: +41 1 632 1039.
Electronic mail: wfvgn@igc.phys.chem.ethz.ch

where V is the volume of the system, T is the absolute temperature, k_B is Boltzmann's constant, and h is Planck's constant, respectively. The factorial should only be present when the N particles are indistinguishable. In analogy, the partition function of the isothermal–isobaric (NPT) ensemble can be written

$$Q(N,p,T)=\frac{\int\int\int \exp(-(\mathcal{H}(\mathbf{p},\mathbf{r})+pV)/k_BT)\,\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}\mathrm{d}V}{h^{3N}N!}, \quad (3)$$

the only difference being a pV term in addition to the Hamiltonian and an additional integration over volume.

The phase-space probability $\pi(\mathbf{p},\mathbf{r})$ to find the system in a given state (in the canonical ensemble) with configuration $\mathbf{r}$ and conjugate momenta $\mathbf{p}$ is given by

$$\pi(\mathbf{p},\mathbf{r})_{NVT}=\frac{\exp(-\mathcal{H}(\mathbf{p},\mathbf{r})/k_BT)}{\int\int \exp(-\mathcal{H}(\mathbf{p},\mathbf{r})/k_BT)\,\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}}=\frac{\exp(-\mathcal{H}(\mathbf{p},\mathbf{r})/k_BT)}{h^{3N}N!Q(N,V,T)} \quad (4)$$

and is used to define the ensemble average of a property X through

$$\langle X\rangle_{NVT}=\int\int X(\mathbf{p},\mathbf{r})\,\pi(\mathbf{p},\mathbf{r})\,\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}. \quad (5)$$

The key quantity of the canonical ensemble, the Helmholtz free energy, can thus be expressed as an ensemble average

$$\begin{aligned}A(N,V,T)&=-k_BT\ln Q(N,V,T)\\&=-k_BT\ln\left\{\frac{\int\int \exp(-\mathcal{H}(\mathbf{p},\mathbf{r})/k_BT)\,\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}}{h^{3N}N!}\right\}\\&=+k_BT\ln\langle\exp(+\mathcal{H}/k_BT)\rangle. \quad (6)\end{aligned}$$

In the NPT ensemble the analogous quantity is the Gibbs free energy

$$G(N,p,T)=A(N,V,T)+pV=-k_BT\ln Q(N,p,T). \quad (7)$$

In the statistical mechanical context, the entropy can be formulated through two fundamental expressions, first as temperature derivative of the free energy (A or G)

$$S=-\left(\frac{\partial A}{\partial T}\right)_{N,V}=-\left(\frac{\partial G}{\partial T}\right)_{N,p}, \quad (8)$$

and second through the difference of free energy (A or G) and energy U (or enthalpy H)

$$\begin{aligned}S(N,V,T)&=k_B\ln Q(N,V,T)+\frac{\int\int\mathcal{H}(\mathbf{p},\mathbf{r})\exp(-\mathcal{H}(\mathbf{p},\mathbf{r})/k_BT)\,\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}}{h^{3N}N!Q(N,V,T)T}\\&=k_B\ln Q(N,V,T)+\frac{\langle\mathcal{H}\rangle_{NVT}}{T}=\frac{-A+U}{T} \quad (9)\end{aligned}$$

or

$$S(N,p,T)=k_B\ln Q(N,p,T)+\frac{\langle\mathcal{H}\rangle_{NpT}}{T}=\frac{-G+H}{T}. \quad (10)$$

In the following, we will focus on the canonical ensemble and give only the key equations for the NPT ensemble or brief indications how to obtain these.

B. Determining free-energy differences

There exists various methods to determine the free-energy difference between two states a and b of a system. We concentrate on two basic approaches, thermodynamic integration[13] and perturbation,[14] which we wish to apply to the determination of entropies.

In the thermodynamic integration (TI) method, the Hamiltonian $\mathcal{H}(\mathbf{p},\mathbf{r})$ is made a function of a coupling parameter λ, $\mathcal{H}(\mathbf{p},\mathbf{r},\lambda)$. Thus, the free energy also becomes λ dependent and the free-energy difference between two states a and b, characterized by λ_a and λ_b, respectively, can be expressed through the following integral:

$$\Delta A^{\mathrm{TI}}_{ba}=A(\lambda_b)-A(\lambda_a)=\int_{\lambda_a}^{\lambda_b}\frac{\mathrm{d}A}{\mathrm{d}\lambda}\mathrm{d}\lambda, \quad (11)$$

where the derivative of the free energy with respect to λ can be written as

$$\begin{aligned}\frac{\mathrm{d}A}{\mathrm{d}\lambda}&=\int\int\frac{\partial\mathcal{H}(\mathbf{p},\mathbf{r},\lambda)}{\partial\lambda}\times\frac{\exp(-\mathcal{H}(\mathbf{p},\mathbf{r},\lambda)/k_BT)}{\int\int\exp(-\mathcal{H}(\mathbf{p}',\mathbf{r}',\lambda)/k_BT)\,\mathrm{d}\mathbf{p}'\mathrm{d}\mathbf{r}'}\,\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}\\&=\int\int\frac{\partial\mathcal{H}(\mathbf{p},\mathbf{r},\lambda)}{\partial\lambda}\,\pi(\mathbf{p},\mathbf{r},\lambda)\,\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}=\left\langle\frac{\partial\mathcal{H}(\lambda)}{\partial\lambda}\right\rangle_\lambda. \quad (12)\end{aligned}$$

The original perturbation (PT) formula was derived by Zwanzig[14]

$$\begin{aligned}\Delta A^{\mathrm{PT}}_{ba}&=A(b)-A(a)\\&=-k_BT\ln\left(\frac{Q(b)}{Q(a)}\right)\\&=-k_BT\ln\left(\int\int\exp(-(\mathcal{H}(\mathbf{p},\mathbf{r},b)-\mathcal{H}(\mathbf{p},\mathbf{r},a))/k_BT)\,\pi(\mathbf{p},\mathbf{r},a)\,\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}\right)\\&=-k_BT\ln(\langle\exp(-(\mathcal{H}(b)-\mathcal{H}(a))/k_BT)\rangle_a), \quad (13)\end{aligned}$$

and, in the terminology of coupling parameter approach, corresponds to making use of the numerical derivative of the free energy with respect to λ. The PT method can be applied in two ways: as in TI, one can make use of artifical intermediate states between states a and b (multistep perturbation[23]), or one can perform a single simulation of one of the endstates[24,25] or of an artificial reference state,[26–28] from which extrapolations to the (other) endstates are carried out (one-step perturbation). The Gibbs free energy G is computed in analogy to A when using NPT simulations.

C. Determining entropy differences

The most obvious, ''naive'' approach to estimate the entropy difference of two states a and b is based on Eq. (9) and uses the free-energy difference, e.g., determined through TI

or PT, and the difference of the energies (or enthalpies) of the two states (i.e., the endpoints of the TI or PT pathways)

$$\Delta S_{ba}^{\mathrm{end}}=\frac{\Delta U_{ba}-\Delta A_{ba}}{T}=\frac{\Delta H_{ba}-\Delta G_{ba}}{T}. \tag{14}$$

A second approach makes use of Eq. (8), which leads to the finite-difference approximation to compute the entropy at temperature T via the free energies (A or G) at $T+\Delta T$ and $T-\Delta T$,

$$\Delta S_{ba}^{\Delta T}=-\frac{\Delta A_{ba}(T+\Delta T)-\Delta A_{ba}(T-\Delta T)}{2\Delta T}. \tag{15}$$

This method assumes a constant difference in the heat capacity $\Delta c_{v,ba}$ (or $\Delta c_{p,ba}$ for NPT) over the temperature range $2\Delta T$.

Other techniques to determine entropy differences can be derived from the methods to compute free-energy differences [using Eq. (8)]. In order to obtain a TI formulation for ΔS_{ba}, the entropy needs to be expressed as a function of the coupling parameter λ,

$$\begin{aligned}S(\lambda)&=-\left(\frac{\partial(-k_BT\ln Q(N,V,T,\lambda)}{\partial T}\right)_{N,V}\\&=k_B\ln Q(N,V,T,\lambda)\\&\quad+k_BT\frac{\partial}{\partial T}\ln\left(\frac{\int\int\exp(-\mathcal{H}(\mathbf{p},\mathbf{r},\lambda)/k_BT)\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}}{h^{3N}N!}\right)\\&=k_B\ln Q(N,V,T,\lambda)\\&\quad+\frac{\int\int\mathcal{H}(\mathbf{p},\mathbf{r},\lambda)\exp(-\mathcal{H}(\mathbf{p},\mathbf{r},\lambda)/k_BT)\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}}{T\int\int\exp(-\mathcal{H}(\mathbf{p},\mathbf{r},\lambda)/k_BT)\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}}.\end{aligned} \tag{16}$$

Differentiated with respect to λ this gives

$$\begin{aligned}\left(\frac{\mathrm{d}S}{\mathrm{d}\lambda}\right)_{N,V,T}&=\frac{1}{k_BT^2}\left\{\left\langle\frac{\partial\mathcal{H}(\lambda)}{\partial\lambda}\right\rangle_\lambda\langle\mathcal{H}(\lambda)\rangle_\lambda\right.\\&\quad\left.-\left\langle\frac{\partial\mathcal{H}(\lambda)}{\partial\lambda}\mathcal{H}(\lambda)\right\rangle_\lambda\right\},\end{aligned} \tag{17}$$

thus, the entropy difference reads

$$\Delta S_{ba}^{\mathrm{TI}}=\frac{1}{k_BT^2}\int_{\lambda_a}^{\lambda_b}\left\{\left\langle\frac{\partial\mathcal{H}}{\partial\lambda}\right\rangle_\lambda\langle\mathcal{H}\rangle_\lambda-\left\langle\frac{\partial\mathcal{H}}{\partial\lambda}\mathcal{H}\right\rangle_\lambda\right\}\mathrm{d}\lambda. \tag{18}$$

The corresponding perturbation analog is

$$\begin{aligned}\Delta S_{ba}^{\mathrm{PT}}&=S(b)-S(a)\\&=-\left(\frac{\partial\Delta A_{ba}^{\mathrm{PT}}}{\partial T}\right)_{N,V}\\&=k_B\ln\langle\exp(-(\mathcal{H}(b)-\mathcal{H}(a))/k_BT)\rangle_a\\&\quad+\frac{1}{T}\frac{\langle\exp(-(\mathcal{H}(b)-\mathcal{H}(a))/k_BT)\cdot\mathcal{H}(b)\rangle_a}{\langle\exp(-(\mathcal{H}(b)-\mathcal{H}(a))/k_BT)\rangle_a}\\&\quad-\frac{1}{T}\langle\mathcal{H}(a)\rangle_a.\end{aligned} \tag{19}$$

Note, that equations analogous to Eqs. (17) and (19) can be derived for the NPT ensemble. They contain additional pV terms in any ensemble average linear in the Hamiltonian. Thus, any ensemble average $\langle\mathcal{H}\rangle$ becomes $\langle\mathcal{H}+pV\rangle$, and $\langle\mathcal{H}\cdot X\rangle$ becomes $\langle[\mathcal{H}+pV]\cdot X\rangle$. Generally, these pV terms do not contribute much to the ensemble-average differences and are therefore neglected throughout this study.

D. Splitting the Hamiltonian: Solute–solvent entropy

Alternative formulas to compute entropy differences (both for TI and PT) can be derived by distinguishing parts of the Hamiltonian that are and that are not dependent on the coupling parameter λ (that do or do not differ for the two states a and b). Without loss of generality, we may think of a Hamiltonian that distinguishes between solute (u) and solvent (v) degrees of freedom, in which the intra-solute and the solute–solvent interactions are dependent on λ, while the intrasolvent interactions are λ independent

$$\mathcal{H}(\mathbf{p},\mathbf{r},\lambda)=\mathcal{H}_{uu}(\mathbf{p},\mathbf{r},\lambda)+\mathcal{H}_{uv}(\mathbf{p},\mathbf{r},\lambda)+\mathcal{H}_{vv}(\mathbf{p},\mathbf{r}). \tag{20}$$

We then find for the free energy (TI),

$$\Delta A_{ba}^{\mathrm{TI}}=\int_{\lambda_a}^{\lambda_b}\frac{\mathrm{d}A}{\mathrm{d}\lambda}\mathrm{d}\lambda=\int_{\lambda_a}^{\lambda_b}\left\langle\frac{\partial\mathcal{H}_{uu}}{\partial\lambda}+\frac{\partial\mathcal{H}_{uv}}{\partial\lambda}\right\rangle_\lambda\mathrm{d}\lambda, \tag{21}$$

or (PT)

$$\begin{aligned}\Delta A_{ba}^{\mathrm{PT}}=-k_BT\ln(\langle\exp(-(\mathcal{H}_{uu}(b)+\mathcal{H}_{uv}(b)-\mathcal{H}_{uu}(a)\\-\mathcal{H}_{uv}(a))/k_BT)\rangle_a).\end{aligned} \tag{22}$$

For ΔS_{ba}, Eq. (18), (TI) becomes

$$\begin{aligned}\Delta S_{ba}^{\mathrm{TI}}&=\frac{1}{k_BT^2}\int_{\lambda_a}^{\lambda_b}\left\{\left\langle\frac{\partial(\mathcal{H}_{uu}+\mathcal{H}_{uv})}{\partial\lambda}\right\rangle_\lambda\langle\mathcal{H}_{uu}+\mathcal{H}_{uv}+\mathcal{H}_{vv}\rangle_\lambda\right.\\&\quad\left.-\left\langle\frac{\partial(\mathcal{H}_{uu}+\mathcal{H}_{uv})}{\partial\lambda}[\mathcal{H}_{uu}+\mathcal{H}_{uv}+\mathcal{H}_{vv}]\right\rangle_\lambda\right\}\mathrm{d}\lambda\\&=\frac{1}{k_BT^2}\int_{\lambda_a}^{\lambda_b}\left\{\left\langle\frac{\partial(\mathcal{H}_{uu}+\mathcal{H}_{uv})}{\partial\lambda}\right\rangle_\lambda\langle\mathcal{H}_{uu}+\mathcal{H}_{uv}\rangle_\lambda\right.\\&\quad\left.-\left\langle\frac{\partial(\mathcal{H}_{uu}+\mathcal{H}_{uv})}{\partial\lambda}[\mathcal{H}_{uu}+\mathcal{H}_{uv}]\right\rangle_\lambda\right\}\mathrm{d}\lambda\\&\quad+\frac{1}{k_BT^2}\int_{\lambda_a}^{\lambda_b}\left\{\left\langle\frac{\partial(\mathcal{H}_{uu}+\mathcal{H}_{uv})}{\partial\lambda}\right\rangle_\lambda\langle\mathcal{H}_{vv}\rangle_\lambda\right.\\&\quad\left.-\left\langle\frac{\partial(\mathcal{H}_{uu}+\mathcal{H}_{uv})}{\partial\lambda}\mathcal{H}_{vv}\right\rangle_\lambda\right\}\mathrm{d}\lambda.\end{aligned} \tag{23}$$

The last integral can be simplified using

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\langle\mathcal{H}_{vv}\rangle_\lambda=\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\int\int\mathcal{H}_{vv}(\mathbf{p},\mathbf{r})\exp(-\mathcal{H}(\mathbf{p},\mathbf{r},\lambda)/k_BT)\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}}{\int\int\exp(-\mathcal{H}(\mathbf{p},\mathbf{r},\lambda)/k_BT)\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}}$$
$$=\frac{1}{k_BT}\left\{\left\langle\frac{\partial(\mathcal{H}_{uu}+\mathcal{H}_{uv})}{\partial\lambda}\right\rangle_\lambda\langle\mathcal{H}_{vv}\rangle_\lambda-\left\langle\frac{\partial(\mathcal{H}_{uu}+\mathcal{H}_{uv})}{\partial\lambda}\mathcal{H}_{vv}\right\rangle_\lambda\right\}. \quad (24)$$

Thus, Eq. (23) can be reformulated

$$\Delta S_{ba}^{\mathrm{TI,all}}=\frac{1}{k_BT^2}\int_{\lambda_a}^{\lambda_b}\left\{\left\langle\frac{\partial(\mathcal{H}_{uu}+\mathcal{H}_{uv})}{\partial\lambda}\right\rangle_\lambda\langle\mathcal{H}_{uu}+\mathcal{H}_{uv}\rangle_\lambda-\left\langle\frac{\partial(\mathcal{H}_{uu}+\mathcal{H}_{uv})}{\partial\lambda}[\mathcal{H}_{uu}+\mathcal{H}_{uv}]\right\rangle_\lambda\right\}\mathrm{d}\lambda+\frac{1}{T}\int_{\lambda_a}^{\lambda_b}\frac{\mathrm{d}}{\mathrm{d}\lambda}\langle\mathcal{H}_{vv}\rangle_\lambda\mathrm{d}\lambda$$
$$=\Delta S_{ba}^{\mathrm{TI},uuv}+\frac{1}{T}(\langle\mathcal{H}_{vv}\rangle_{\lambda_b}-\langle\mathcal{H}_{vv}\rangle_{\lambda_a})$$
$$=\Delta S_{ba}^{\mathrm{TI},uuv}+\frac{\Delta U^{\mathrm{end},vv}}{T}, \quad (25)$$

where the notation $\Delta S_{ba}^{\mathrm{TI},uuv}$ was introduced to indicate that the first integral refers to the solute–solute and solute–solvent entropy difference. Physically, this result shows that the entropy difference is determined through the difference in the solute–solute plus the solute–solvent entropy and the solvent–solvent energy, since the contributions of solvent–solvent entropy and solute–solvent energy cancel exactly [Eq. (24), see also Ref. 29]. A detailed analysis of these types of entropy–enthalpy compensation is given in Ref. 30.

Splitting the Hamiltonian, Eq. (19) (PT) can be reformulated too,

$$\Delta S_{ba}^{\mathrm{PT}}=k_B\ln\langle\exp(-(\mathcal{H}_{uu}(b)+\mathcal{H}_{uv}(b)-\mathcal{H}_{uu}(a)-\mathcal{H}_{uv}(a))/k_BT)\rangle_a+\frac{1}{T}\frac{\langle\exp(-(\mathcal{H}_{uu}(b)+\mathcal{H}_{uv}(b)-\mathcal{H}_{uu}(a)-\mathcal{H}_{uv}(a))/k_BT)\cdot[\mathcal{H}_{uu}(b)+\mathcal{H}_{uv}(b)]\rangle_a}{\langle\exp(-(\mathcal{H}_{uu}(b)+\mathcal{H}_{uv}(b)-\mathcal{H}_{uu}(a)-\mathcal{H}_{uv}(a))/k_BT)\rangle_a}-\frac{1}{T}\langle\mathcal{H}_{uu}(a)+\mathcal{H}_{uv}(a)\rangle_a$$
$$+\frac{1}{T}\frac{\langle\exp(-(\mathcal{H}_{uu}(b)+\mathcal{H}_{uv}(b)-\mathcal{H}_{uu}(a)-\mathcal{H}_{uv}(a))/k_BT)\cdot\mathcal{H}_{vv}(b)\rangle_a}{\langle\exp(-(\mathcal{H}_{uu}(b)+\mathcal{H}_{uv}(b)-\mathcal{H}_{uu}(a)-\mathcal{H}_{uv}(a))/k_BT)\rangle_a}-\frac{1}{T}\langle\mathcal{H}_{vv}(a)\rangle_a$$
$$=\Delta S_{ba}^{\mathrm{PT},uuv}+\frac{1}{T}(\langle\mathcal{H}_{vv}(b)\rangle_b-\langle\mathcal{H}_{vv}(a)\rangle_a), \quad (26)$$

where the last line again gives a physical explanation of the different terms that contribute to $\Delta S_{ba}^{\mathrm{PT}}$, namely, combining the first three lines yields an estimate for the solute–solute and solute–solvent entropy difference, whereas the next two lines refer to the difference in the solvent–solvent energy.

Concluding this section, it should be mentioned that, since the kinetic energy part of the Hamiltonian [Eq. (1)] can be separated in the ensemble averages, in practice only the potential energy $E_{\mathrm{pot}}(\mathbf{r})$ is considered instead of $\mathcal{H}(\mathbf{p},\mathbf{r})$. This implies that the internal (total) energy U is replaced by the internal potential energy U_{pot}.

E. Other methods to compute entropy

Obviously, other methods to determine free-energy differences can in principle also be extended to obtain entropy differences. One method that can be mentioned here as example is particle insertion,[19] the application of which is limited to the calculation of solvation free energies (and entropies) of atoms and small molecules.

For completeness, other techniques to determine entropy should be mentioned: the adiabatic switching method[31] can be used to obtain absolute entropies and has been succesfully applied to determine excess entropies of pure liquids,[32] however, a generalization to other (mixed) systems seems to be difficult, if not impossible. An acceptance ratio method to determine the excess entropy of pure liquids has been developed and successfully tested for water.[33] A number of methods to determine configurational entropies based on a harmonic approximation of the internal degrees of freedom has been developed for nondiffusive systems.[8–12]

III. COMPUTATIONAL DETAILS

All simulations were performed using the GROMOS96 package of programs.[34,35] They were all based on an equilibrated cubic, periodic simulation box containing 1000 SPC (Ref. 22) molecules (initial box length: 3.132 nm). The simulation temperature was kept constant by weakly coupling to a temperature bath with a relaxation time of 0.1 ps.[36] For NPT simulations, the pressure was maintained at 1 atm by also applying the weak coupling algorithm with a relaxation time of 0.5 ps and an isothermal compressibility of 76.24×10^{-5} $(\mathrm{kJ\,mol^{-1}\,nm^{-3}})^{-1}$. For the nonbonded interactions, a twin-range method with cutoff radii of 0.8 and 1.4 nm was used.[34,35] Outside the longer cutoff radius a reaction field correction[37] was applied with a relative dielectric permittivity of 78.5. The integration time step was 2 fs, the

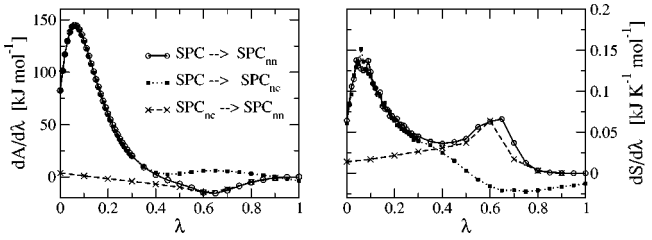

FIG. 1. Thermodynamic integration perturbing all (1000) molecules. Simulation length per λ point: 100 ps; SPC: full nonbonded interactions, SPC_{nn}: no nonbonded interactions, SPC_{nc}: no Coulomb interactions. $dA/d\lambda$: Eq. (12). $dS/d\lambda$: Eq. (17).

pairlist for pairs within the inner cutoff and the energies and forces for pairs between the inner and outer cutoff radii were updated every 10 fs.

In a thermodynamic cycle three types of SPC molecules were considered: SPC molecules with normal interactions (denoted as SPC), SPC molecules without Coulombic interactions (denoted as SPC_{nc}), and SPC molecules without nonbonded interactions (SPC_{nn}). Two types of transitions between these molecules were performed: (i) perturbing the interactions of all molecules (for the transition SPC $\rightarrow SPC_{nn}$ this gives the excess free energy and entropy), and (ii) perturbing only one molecule, which yields the hydration free energy and entropy of the molecule of interest.

Two methods to determine free-energy differences and to test various approaches to compute entropies were applied, thermodynamic integration and one-step perturbation. Multistep perturbation had been briefly tested and performed essentially like TI. It should be noted though, that perturbing all molecules can only sensibly be carried out using thermodynamic integration (with constant volume).

In the TI calculations the states *a* and *b* were connected using λ points 0.1 apart in the range [0,1]. In particular cases more λ points were inserted to obtain a more accurate reference free energy or entropy value (see Figs. 1 and 2). All TI calculations were performed using a quadratic dependence of the potential energy on the coupling parameter λ and a soft-core interaction on the perturbed atoms (see below). At each λ point, an initial equilibration of 50 ps for the all-molecule perturbation TI calculations and of 250 ps for the single-molecule perturbation TI calculations was carried out. After this equilibration period, the simulation at each λ point was performed for 100 ps for the all-molecule perturbation TI, for 600 ps for the one-molecule perturbation TI at 300 K, and for 200 ps for the one-molecule perturbation TI at 280 and 320 K.

The one-step perturbation calculations were carried out using an uncharged soft-core reference state as described in Refs. 27 and 28. The van der Waals interactions of the soft-core reference state with the solvent molecules had the form

$$E_{\mathrm{pot,vdW,soft}}(r,\lambda,\alpha) = \frac{C12_{\mathrm{SO}}}{\left(r^6+\lambda^2\alpha\frac{C12_{\mathrm{SO}}}{C6_{\mathrm{SO}}}\right)^2} - \frac{C6_{\mathrm{SO}}}{\left(r^6+\lambda^2\alpha\frac{C12_{\mathrm{SO}}}{C6_{\mathrm{SO}}}\right)}, \tag{27}$$

where α, a parameter that controls the ''softness'' of the interaction, was set to 1.51, and λ, a coupling parameter included here only for implementation reasons, was set to 0.5. The parameters $C6_{\mathrm{SO}}$ and $C12_{\mathrm{SO}}$ of the Lennard-Jones interactions between the soft-core reference and the oxygen atoms of the SPC molecules were computed using geometric combination rules as described in Ref. 34. The reference state was chosen with a mass of 18.0154 amu, and $\sqrt{C6_{SS}} = 0.27322\ (\mathrm{kJ\,mol^{-1}\,nm^6})^{1/2}$ and $\sqrt{C12_{SS}} = 0.05901(\mathrm{kJ\,mol^{-1}\,nm^{12}})^{1/2}$. We note that in Ref. 27 the squares of these parameter values are erroneously attributed to the soft-core oxygen interactions. The height of the soft-core potential used at $r=0$ thus is $9.37\ \mathrm{kJ\,mol^{-1}}$, and not $6.99\ \mathrm{kJ\,mol^{-1}}$ as quoted in Refs. 27 and 28. The initial configuration was generated by replacing one SPC molecule of the equilibrated box described above by the soft-core particle and equilibrating for 10 ps. As described in Ref. 28, the insertion of a SPC (or SPC_{nc}) molecule at the soft-core site was performed with additional translational and rotational sampling. The molecule was first inserted in ten random orientations, and then for each orientation ten random displacements in an interval of $[-0.05\ \mathrm{nm},\ 0.05\ \mathrm{nm}]$ were carried out.

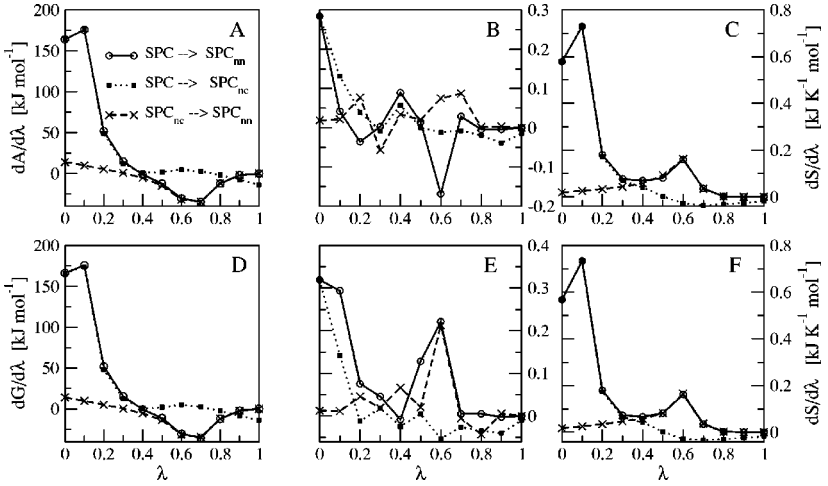

FIG. 2. Thermodynamic integration perturbing a single out of 1000 molecules; SPC: full nonbonded interactions, SPC_{nn}: no nonbonded interactions, SPC_{nc}: no Coulomb interactions. Upper panels: NVT, lower panels: NPT; A and D: $dA^{\mathrm{TI}}/d\lambda$, $dG^{\mathrm{TI}}/d\lambda$ according to Eq. (12); B and E: $dS^{\mathrm{TI}}/d\lambda$ according to Eq. (17); C and F: $dS^{\mathrm{TI},uuv}/d\lambda$ according to Eq. (25). Simulation length per λ point: 600 ps.

TABLE I. Free energies, energies and entropies obtained by thermodynamic integration. Method (m): the interaction of all or a single water molecule is λ dependent; transition between λ-dependent states: SPC all interactions, SPC_{nn} no nonbonded interactions, SPC_{nc} no Coulomb interactions; thermodynamic boundary conditions (tbc): NVT constant volume, NPT constant pressure; T temperature; l simulation length per λ point; $\langle V\rangle$ average volume determined from the simulations at the two endstates. ΔA^{TI} Eq. (11); ΔU^{end}_{pot} difference in potential energy between the two endstates; ΔS^{TI} Eq. (18); ΔS^{end} Eq. (14); $\Delta U^{end,vv}_{pot}$ difference in solvent–solvent potential energy between the two endstates; $\Delta S^{TI,uuv}$ Eq. (25); $\Delta S^{TI,all}$ Eq. (25); The different methods to compute ΔS are indicated by [A], [B], and [C] in the text.

						ΔA^{TI} (ΔG^{TI})	ΔU^{end}_{pot} (ΔH^{end}_{pot})	ΔS^{TI} [A]	ΔS^{end} [B]	$\Delta U^{end,vv}_{pot}$ ($\Delta H^{end,vv}_{pot}$)	$\Delta S^{TI,uuv}$	$\Delta S^{TI,all}$ [C]
m	Transition	tbc	T [K]	l [ps]	$\langle V\rangle$ [nm^3]	[kJ mol^{-1}]		[J K^{-1} mol^{-1}]		[kJ mol^{-1}]	[J K^{-1} mol^{-1}]	
all	$SPC\rightarrow SPC_{nn}$	NVT	300	100	30.73	23.2	41.3	47.0	60.3			
	$SPC\rightarrow SPC_{nc}$		300	100	30.73	28.7	39.4	22.3	35.7			
	$SPC_{nc}\rightarrow SPC_{nn}$		300	100	30.73	−5.3	1.9	22.9	24.0			
single	$SPC\rightarrow SPC_{nn}$	NVT	300	600	30.73	23.1	40.7	10.7	58.7	−41.4	162.5	24.5
			300	200	30.73	23.0	45.3	22.1	74.3	−37.0	157.8	34.5
			280	200	30.73	25.0	29.6	29.7	16.4	−56.4	175.2	−26.2
			320	200	30.73	21.8	48.7	40.3	84.1	−32.1	149.2	48.9
		NPT	300	600	30.80	23.4	59.5	92.4	120.3	−23.6	161.7	83.0
			300	200	30.81	23.3	76.2	62.9	176.3	−8.0	166.5	139.8
			280	200	30.36	24.7	39.7	51.7	53.6	−45.9	171.5	7.6
			320	200	31.34	22.4	59.6	63.7	116.2	−21.3	144.8	78.2
single	$SPC\rightarrow SPC_{nc}$	NVT	300	600	30.73	31.3	52.1	27.8	69.3	−23.1	117.5	40.5
			300	200	30.73	31.5	64.6	69.4	110.3	−10.7	119.0	83.3
		NPT	300	600	30.82	30.9	52	12.9	70.3	−24.2	117.5	36.8
			300	200	30.82	31.1	58.4	5.6	91.0	−18.8	115.4	52.7
			280	200	30.39	31.8	75	24.9	154.3	−3.3	125.6	113.8
			320	200	31.35	30.6	61.7	19.7	97.2	−12.5	106.4	67.3
single	$SPC_{nc}\rightarrow SPC_{nn}$	NVT	300	600	30.73	−7.7	9.7	27.2	58.0	2.7	46.0	55.0
			300	200	30.73	−7.4	−1.7	65.5	19.0	−8.7	45.3	16.3
		NPT	300	600	30.82	−7.8	17	33.9	82.7	10.0	45.9	79.2
			300	200	30.82	−8.0	15.1	25.6	77.0	8.0	45.9	72.6
			280	200	30.37	−6.4	23.8	37.0	107.9	16.8	42.4	102.4
			320	200	31.34	−7.7	−14.4	24.7	−20.9	−21.1	41.0	−24.9

IV. RESULTS

The following estimates for entropy differences were computed (whenever applicable):

(A) Thermodynamic integration: ΔS^{TI} as given in Eq. (18).
(B) Entropy as difference of energy/enthalpy and free energy: ΔS^{end} according to Eq. (14) with ΔA^{TI} (or ΔG^{TI}) computed using TI according to Eq. (11) and ΔU^{end}_{pot} (or ΔH^{end}_{pot}) calculated from the endpoints of the TI pathway.
(C) Thermodynamic integration with splitting of the Hamiltonian into solute–solvent versus solvent–solvent terms: $\Delta S^{TI,all}$ as given in Eq. (25), using $\Delta S^{TI,uuv}$ computed using TI including only the solute–solvent terms of the Hamiltonian (the solute–solute terms are zero) and $\Delta U^{end,vv}_{pot}$ (or $\Delta H^{end,vv}_{pot}$) from the endpoints of the TI pathways.
(D) Entropy as temperature derivative of free energy: $\Delta S^{\Delta T}$ through the finite-difference estimate given in Eq. (15) via $\Delta A^{TI}(T\pm\Delta T)$ [or $\Delta G^{TI}(T\pm\Delta T)$], these computed using TI according to Eq. (11).
(E) One-step perturbation: ΔS^{PT} [Eq. (19)].
Due to the nature of the one-step perturbation approach (simulation only at one state, from which the other ones are extrapolated), no one-step perturbation equivalents to [B] and [C] can be given, the corresponding equations ultimately lead to Eq. (19), i.e., the estimates are identical.
(F) One-step perturbation equivalent to [D]: Equation (15) with $\Delta A^{PT}(T\pm\Delta T)$ [or $\Delta G^{PT}(T\pm\Delta T)$] computed using PT according to Eq. (13).

A. Thermodynamic integration

Table I presents the free-energy and entropy differences obtained by thermodynamic integration. The upper part of the table shows the results for perturbing all molecules in the system. The cycle closures of the thermodynamic cycle $SPC\rightarrow SPC_{nc}\rightarrow SPC_{nn}\rightarrow SPC$ ($\Sigma\Delta A^{TI}=0.2$ kJ mol^{-1} and $\Sigma\Delta S^{TI}=-1.8$ J K^{-1} mol^{-1}, where $\Sigma\Delta X$ denotes the sum of the property ΔX in the thermodynamic cycle) illustrate that both the free energy and the entropy are quite well converged. Cycle closures of ΔU^{end} are not meaningful, as they in this case rely on the same endstate simulations and therefore add up to exactly zero. We note that this is not exactly true for all cases discussed below, as sometimes two independent simulations of the endstates are involved. For the ΔA^{TI} values given in this upper part of the table, a statistical error could be estimated which is at most 0.05 kJ mol^{-1}, the statistical error of ΔU^{end}_{pot} is below 0.02 kJ mol^{-1}. Figure 1 shows the (well-converged) derivatives of A and S with re-

spect to λ [Eqs. (12) and (17)], again indicating that one can obtain not only good estimates for ΔA but also reasonable estimates for ΔS with this method (method [A]). No error bars are displayed in the figure, as the ones for $dA/d\lambda$ are too small to be visible and the ones for $dS/d\lambda$ too large due to the sizeable errors of the energy ensemble averages. These errors in the energy ensemble averages seem to cancel partially upon computing entropies, leading to comparatively good results. Comparing the $dA/d\lambda$ and the $dS/d\lambda$ profiles of the three transitions shown, one can attribute the maxima and minima to different types of interactions, namely the first maximum (around $\lambda=0.1$) to switching off electrostatic interactions and the second extremum (around $\lambda=0.6$) to removing the van der Waals interactions. This nicely shows, that the effects of turning off the different types of nonbonded interactions are well separated along this λ path. This makes a more complex procedure, in which first electrostatic and subsequently the remaining nonbonded interactions are switched off, unnecessary. Table I also presents the ΔS^{end} estimates obtained from free-energy differences and energy differences [Eq. (14); method [B]], which differ significantly from the results directly obtained from the TI calculation (ΔS^{TI}; method [A]). Considering the error estimates, ΔS^{end} seems to be more precise than ΔS^{TI}. Comparison of the values of ΔA^{TI} for the transition $\mathrm{SPC}\rightarrow\mathrm{SPC}_{nn}$ with values for the excess free energy of SPC from the literature (expt. 24.0 kJ mol^{-1}; simulated 24–25 kJ mol^{-1} at slightly different simulation conditions[38]) gives excellent agreement.

The results for perturbing only a single molecule are presented in the lower part of Table I both for constant volume and constant pressure simulations at 280, 300, and 320 K. The convergence behavior differs significantly from the case where all molecules are perturbed. Again, the free-energy differences ΔA^{TI} seem to be reasonably well converged, as can be seen from cycle closures and from the comparison of the results for different simulation lengths per λ value. This is confirmed by Fig. 2, where the derivative of the free energy with respect to λ is presented in the left panels (again, the error bars are too small to be visible). Estimates of the statistical error of ΔA^{TI} in the table range from 0.4 to 0.6 kJ mol^{-1} for the simulations with 600 ps per λ point and from 0.5 to 1.1 kJ mol^{-1} for the simulations with 200 ps per λ point. The $dA/d\lambda$ profiles are qualitatively very similar to those discussed above for Fig. 1 where all molecules were perturbed. The resulting Gibbs free energy of solvation of water (transition $\mathrm{SPC}\rightarrow\mathrm{SPC}_{nn}$) agrees reasonably well with experimental values [26.5 kJ mol^{-1} at 303 K; 27.5 kJ mol^{-1} at 283 K; 25.5 kJ mol^{-1} at 323 K (Ref. 39)], in particular the temperature dependence of ΔG^{TI} seems to be quite decent. As expected, the free-energy estimates do not differ much between the NPT and NVT ensembles. The entropy differences ΔS^{TI} though have not converged after 600 ps of simulation per λ point. This is not only demonstrated by the nonclosure of thermodynamic cycles ($\Sigma\Delta S^{\mathrm{TI}}$ amounts to up to 44 and -46 J K^{-1} mol^{-1} for NVT and NPT at 300 K, respectively) but also by the rather erratic behavior of $dS/d\lambda$ as displayed in the middle panels of Fig. 2. As expected, the estimates of ΔS^{end} are equally or even more erroneous as the ΔS^{TI} estimates, which results from the large error in the potential energy ensemble averages. The statistical error of ΔU^{end} ranges beween 10 and 15 kJ mol^{-1} for the simulations with 600 ps per λ point and is approximately 20 kJ mol^{-1} for the simulations with 200 ps per λ point. It should be noted that the statistical error is not sufficient to describe possible deviations in ΔU^{end} when completely independent simulations are considered. This is probably due to dependence of the total potential energy of the system on the (not entirely converged) pressure. Again it should be noted that one is dealing here with errors in the total potential energy of the system, which is in the present cases in the order of $-40\,000$ kJ mol^{-1}.

TABLE II. $\Delta S^{\Delta T}$ at 300 K determined from ΔA^{TI} and ΔG^{TI} (values can be found in Table I) at 280 K and 320 K via Eq. (15). Thermodynamic boundary conditions (tbc): NVT constant volume, NPT constant pressure.

Transition	tbc	$\Delta S^{\Delta T}$ [J K^{-1} mol^{-1}]
$\mathrm{SPC}\rightarrow\mathrm{SPC}_{nn}$	NVT	80
$\mathrm{SPC}\rightarrow\mathrm{SPC}_{nn}$	NPT	58
$\mathrm{SPC}\rightarrow\mathrm{SPC}_{nc}$	NPT	30
$\mathrm{SPC}_{nc}\rightarrow\mathrm{SPC}_{nn}$	NPT	33

When only one molecule is perturbed, it is possible to determine the solute–solvent entropy difference $\Delta S^{\mathrm{TI},uuv}$ [Eq. (25)]. As $\Delta S^{\mathrm{TI},uuv}$ relies only on solute–solute and solute–solvent potential energy terms, it converges much better than the total entropy difference ΔS^{TI}, where the necessary ensemble averages include the total potential energy. This is reflected in the right panels of Fig. 2, where the derivatives of $\Delta S^{\mathrm{TI},uuv}$ with respect to λ are displayed. Comparing the profiles of $dS^{\mathrm{TI},uuv}/d\lambda$ with those obtained for $dS/d\lambda$ when perturbing all molecules (Fig. 1), one notices a qualitative agreement of maxima and minima for the different transitions. A similar agreement can only be guessed for the middle panels of Fig. 2, where all interactions are considered. However, if one wants to use the (well-converged) estimates of the solute–solvent entropy difference $\Delta S^{\mathrm{TI},uuv}$ to compute the difference of the total entropy $\Delta S^{\mathrm{TI,all}}$ [Eq. (25); method [C]], an accurate estimate of the solvent–solvent potential energy difference $\Delta U_{\mathrm{pot}}^{\mathrm{end},vv}$ between states *a* and *b* is required. The corresponding columns in Table I show, that this is exactly the problem, the values of $\Delta U_{\mathrm{pot}}^{\mathrm{end},vv}$ are by far too ill-converged to result in a good estimate for $\Delta S^{\mathrm{TI,all}}$.

Thermodynamic integrations were not only carried out at 300 K but also at 280 and 320 K. Thus ΔS at 300 K can also be determined using the free-energy difference estimates at $T\pm\Delta T$ [Eq. (15); method [D]]. The simulations at each λ point were performed only for 200 ps, as this seemed to be sufficient to obtain a reasonable estimate for the free-energy differences at 300 K. The results are presented in Table II. The resulting entropy of solvation at constant pressure (transition $\mathrm{SPC}\rightarrow\mathrm{SPC}_{nn}$: 58 J K^{-1} mol^{-1}) agrees very well with the experimental value[39] (51 J K^{-1} mol^{-1}). Note that cycle closure is not necessarily a good measure to assess the quality of the resulting entropy differences. The fact that the cycle for $\Delta S^{\Delta T}$ closes reasonably well (since $\Delta S^{\Delta T}$ values rely on free-energy differences) does not say anything about the accuracy of $\Delta S^{\Delta T}$, since it is not related to the tempera-

TABLE III. One-step perturbation (length of each simulation: 1 ns). Thermodynamic boundary conditions (tbc): NVT constant volume, NPT constant pressure. ΔA^{PT} Eq. (13); ΔS^{PT} Eq. (19); $\Delta S^{\mathrm{PT},uuv}$ Eq. (26).

Transition	tbc	T [K]	$\langle V\rangle$ [nm^3]	ΔA^{PT} (or ΔG^{PT}) [kJ mol^{-1}]	ΔS^{PT} [J K^{-1} mol^{-1}]	$\Delta S^{\mathrm{PT},uuv}$ [J K^{-1} mol^{-1}]
SPC→SPC$_{nn}$	NVT	300	30.73	19.4	217.5	106.7
		280	30.73	16.4	−209.8	116.4
		320	30.73	17.9	−112.6	99.7
	NPT	300	30.85	20.7	45.1	101.7
		280	30.41	18.6	−219.0	96.0
		320	31.39	15.9	96.6	96.7
SPC→SPC$_{nc}$	NVT	300	30.73	28.1	153.9	54.6
		280	30.73	23.7	−463.8	65.6
		320	30.73	28.6	−185.8	46.0
	NPT	300	30.85	29.2	47.7	50.1
		280	30.41	25.1	−190.3	47.0
		320	31.39	25.1	27.8	47.6
SPC$_{nc}$→SPC$_{nn}$	NVT	300	30.73	−8.7	63.6	52.1
		280	30.73	−7.3	254.0	50.8
		320	30.73	−10.6	73.2	53.7
	NPT	300	30.85	−8.5	−2.6	51.6
		280	30.41	−6.5	−28.7	49.0
		320	31.39	−9.2	68.8	49.1

ture dependence of the free energies involved. Another weak point of this approach is, that, as opposed to the TI and PT formulas which are in principle exact, this finite-difference approach assumes a constant difference in heat capacity Δc_v (or Δc_p for NPT) over the temperature range of interest.

B. One-step perturbation

Table III presents the results of the one-step perturbation calculations. As observed in previous studies,[28] the one-step perturbation method with additional translational and rotational sampling when inserting a (dipolar) molecule at the soft-core site yields reasonably good estimates for free-energy differences ΔA^{PT}. However, comparison with the values obtained by TI, ΔA^{TI} (Table I), shows that the method is of no use to predict a reasonable temperature dependence of the free-energy. This implies that the one-step perturbation finite-difference method [F] [Eq. (15)] cannot be used to obtain accurate $\Delta S^{\Delta T}$ values. In Fig. 3, the convergence of ΔA^{PT} at 300 K is presented. It shows that the best convergence is achieved for the transition SPC$_{nc}$→SPC$_{nn}$, where no charges are involved. This is reasonable, since the uncharged soft-core particle is a better reference for uncharged states *a* and *b* than for a state involving partial charges. This can apparently not be completely compensated by the additional translational and rotational sampling. Table III also shows differences of total entropies ΔS^{PT} [Eq. (19); method [E]] and of solute–solvent entropies $\Delta S^{\mathrm{PT},uuv}$ [Eq. (26)]. The total entropy differences are completely unrealistic results with a huge scattering, whereas the estimates of $\Delta S^{\mathrm{PT},uuv}$ seem to be better converged, at least if one compares results for the same transition at different temperatures and for different ensembles. Yet, these $\Delta S^{\mathrm{PT},uuv}$ estimates do not match the values $\Delta S^{\mathrm{TI},uuv}$ obtained previously by thermodynamic integration (Table I). This is probably due to the fact, that $\Delta S^{\mathrm{PT},uuv}$ in Eq. (26) contains essentially two contributions, one from the solute–solvent free-energy difference and the second from the solvent–solvent energy/enthalpy difference, which is obtained by extrapolating the solvent–solvent energy at state *b* from the simulation at state *a*, which probably yields a very poor estimate for this energy. The convergence behavior of ΔS^{PT} and $\Delta S^{\mathrm{PT},uuv}$ for the simulations at 300 K is monitored in Fig. 4, showing that indeed the overall entropy difference is completely erratic, whereas the solute–solvent entropies do converge, at least towards some value.

V. CONCLUSIONS

We have compared the accuracy of various formulas and procedures to compute entropy differences, in particular the excess entropy and the entropy of solvation using liquid water as a test system. The methodological problem of calculating entropy is that any method requires an accurate esti-

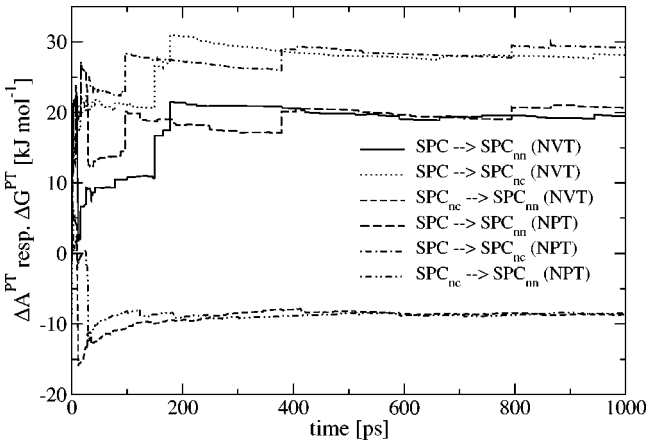

FIG. 3. Convergence of ΔA^{PT} [Eq. (13)] or ΔG^{PT} for one-step perturbation calculations at 300 K. SPC: full nonbonded interactions, SPC$_{nn}$: no nonbonded interactions, SPC$_{nc}$: no Coulomb interactions.

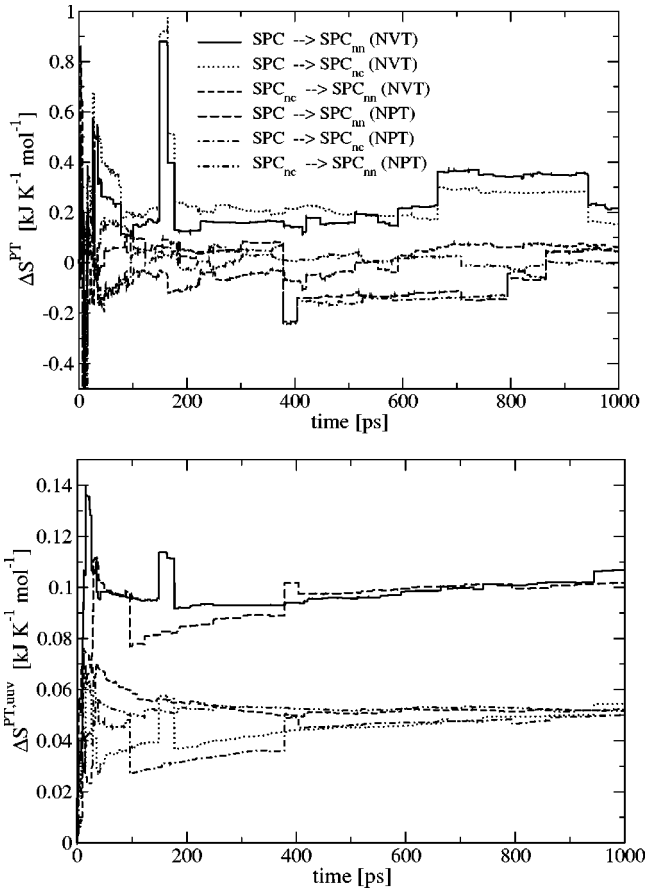

FIG. 4. Convergence of ΔS^{PT} [Eq. (19), upper panel] and $\Delta S^{\mathrm{PT},uuv}$ [Eq. (26), lower panel] for one-step perturbation calculations at 300 K. SPC: full nonbonded interactions, SPC_{nn}: no nonbonded interactions, SPC_{nc}: no Coulomb interactions.

mate of an ensemble average that includes the complete Hamiltonian of the system. Such an ensemble average shows enormous fluctuations and therefore takes a very long simulation time to converge. Two different types of perturbations were considered: perturbing all molecules in the system (excess entropy) and perturbing a single molecule (entropy of solvation). For the first case good, well-converged estimates for entropy differences could be obtained, while the second case was by far more problematic. In this case, comparing the entropy difference obtained by thermodynamic integration, ΔS^{TI}, and the one obtained from the free-energy difference (from thermodynamic integration) combined with the energy/enthalpy difference of the endpoint simulations, ΔS^{end}, the following observations were made: (i) with the applied simulation lengths of 600 ps per λ value (or per endstate), neither method yields an accurate result for ΔS, (ii) ΔS^{TI} requires a value for the ensemble average $\langle \partial \mathcal{H}/\partial \lambda \rangle_\lambda \langle \mathcal{H} \rangle_\lambda - \langle \partial \mathcal{H}/\partial \lambda \mathcal{H} \rangle_\lambda$ which shows smaller fluctuations than the ensemble average $\langle \mathcal{H} \rangle_\lambda$ needed for ΔS^{end}, as the latter ensemble average is an extensive quantity, thus growing with system size, (iii) on the other hand, for ΔS^{TI} a comparatively long simulation has to be carried out at each λ point, whereas for ΔS^{end} the simulations at each λ point can be much shorter, as they are only needed to determine ΔA^{TI}, the gained simulation time can in principle be invested in performing longer simulations of the two endstates. The most promising approach is to determine the temperature dependence of the corresponding free-energy difference by thermodynamic integration. Using thermodynamic integration, one can also get a decent estimate for the difference in the solute–solvent entropies. For processes in which the solvent–solvent energy/enthalpy term is nearly constant, the solute–solvent energy/enthalpy and entropy terms will be representative for for the total energy and entropy terms, respectively. For example,[40,41] the relative contributions of enthalpy and entropy to the free enthalpy of solvation of a series of small solutes in different binary and ternary aqueous solutions could be understood using the solute–solvent energy and entropy terms investigated here. The one-step perturbation approach is a poor approximation to determine entropy differences, even when using it to estimate an entropy difference through determination of the temperature dependence of the corresponding free-energy difference. None of the techniques considered seems suitable to give a perspective for the calculation of the entropy of ligand-protein binding or entropy of polypeptide folding.

ACKNOWLEDGMENT

The authors wish to thank Nico van der Vegt for many stimulating discussions.

[1] D. L. Beveridge and F. M. DiCapua, Annu. Rev. Biophys. Biophys. Chem. **18**, 431 (1989).
[2] T. P. Straatsma and J. A. McCammon, Annu. Rev. Phys. Chem. **43**, 407 (1992).
[3] P. M. King, in *Computer Simulation of Biomolecular Systems, Theoretical, and Experimental Applications*, edited by W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson (Escom Science, Leiden, The Netherlands, 1993), Vol. 2, pp. 267–314.
[4] W. F. van Gunsteren, T. C. Beutler, F. Fraternali, P. M. King, A. E. Mark, and P. E. Smith, in *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*, edited by W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson (Escom Science, Leiden, The Netherlands, 1993), Vol. 2, pp. 267–314.
[5] P. Kollman, Chem. Rev. (Washington, D.C.) **93**, 2395 (1993).
[6] A. E. Mark, in *Encyclopedia of Computational Chemistry*, edited by P. von Rague Schleyer (Wiley, New York, 1998), Vol. 2, pp. 1070–1083.
[7] C. Chipot and D. A. Pearlman, Mol. Simul. **28**, 1 (2002).
[8] M. Karplus and J. N. Kushick, Macromolecules **14**, 325 (1981).
[9] A. Di Nola, H. J. C. Berendsen, and O. Edholm, Macromolecules **17**, 2044 (1984).
[10] O. Edholm and H. J. C. Berendsen, Mol. Phys. **51**, 1011 (1984).
[11] O. L. Rojas, R. M. Levy, and A. Szabo, J. Chem. Phys. **85**, 1037 (1986).
[12] J. Schlitter, Chem. Phys. Lett. **215**, 617 (1993).
[13] J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).
[14] R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).
[15] H. Schäfer, L. J. Smith, A. E. Mark, and W. F. van Gunsteren, Proteins **46**, 215 (2002).
[16] H. Schäfer, X. Daura, A. E. Mark, and W. F. van Gunsteren, Proteins **43**, 45 (2001).
[17] C. L. Brooks III, J. Phys. Chem. **90**, 6680 (1986).
[18] S. H. Fleischmann and C. L. Brooks III, J. Chem. Phys. **87**, 3029 (1987).
[19] B. Guillot, Y. Guissani, and S. Bratos, J. Chem. Phys. **95**, 3643 (1991).
[20] D. E. Smith and A. D. J. Haymet, J. Chem. Phys. **98**, 6445 (1992).
[21] N. Lu, D. A. Kofke, and T. B. Woolf, J. Phys. Chem. **107**, 5598 (2003).
[22] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981), pp. 331–342.
[23] M. Mezei and D. L. Beveridge, Ann. N.Y. Acad. Sci. **482**, 1 (1986).
[24] P. E. Smith and W. F. van Gunsteren, J. Chem. Phys. **100**, 577 (1994).
[25] G. Hummer and A. Szabo, J. Chem. Phys. **105**, 2004 (1996).

[26] H. Liu, A. E. Mark, and W. F. van Gunsteren, J. Phys. Chem. **100**, 9485 (1996).
[27] H. Schäfer, W. F. van Gunsteren, and A. E. Mark, J. Comput. Chem. **20**, 1604 (1999).
[28] J. W. Pitera and W. F. van Gunsteren, J. Phys. Chem. B **105**, 11264 (2001).
[29] H.-A. Yu and M. Karplus, J. Chem. Phys. **89**, 2366 (1988).
[30] H. Qian and J. J. Hopfield, J. Chem. Phys. **105**, 9292 (1996).
[31] M. Watanabe and W. P. Reinhard, Phys. Rev. Lett. **65**, 3301 (1990).
[32] L. W. Tsao, S. Y. Sheu, and C. Y. Mou, J. Chem. Phys. **101**, 2302 (1994).
[33] S. D. Hong and D. J. Jang, Chem. Lett. **9**, 946 (2002).
[34] 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).
[35] 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 AG an der ETH Zürich, Zürich, 1996).
[36] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[37] I. G. Tironi, R. Sperb, P. E. Smith, and W. F. van Gunsteren, J. Chem. Phys. **102**, 5451 (1995).
[38] J. Hermans, A. Pathiaseril, and A. Anderson, J. Am. Chem. Soc. **110**, 5982 (1988).
[39] A. Ben-Naim and Y. Marcus, J. Chem. Phys. **81**, 2016 (1984).
[40] N. F. A. van der Vegt and W. F. van Gunsteren, J. Phys. Chem. B (to be published).
[41] N. F. A. van der Vegt, D. Trzesniak, B. Kasumaj, and W. F. van Gunsteren, Chem. Phys. Chem. (to be published).