

## The “Hot-Solvent/Cold-Solute” Problem Revisited

M. Lingenheil, R. Denschlag, R. Reichold, and P. Tavan\*

*Lehrstuhl für Biomolekulare Optik, Ludwig-Maximilians-Universität München,  
Oettingenstrasse 67, 80538 München, Germany*

Received February 1, 2008

**Abstract:** The temperature steers the equilibrium and nonequilibrium conformational dynamics of macromolecules in solution. Therefore, corresponding molecular dynamics simulations require a strategy for temperature control which should guarantee that the experimental statistical ensemble is also sampled *in silico*. Several algorithms for temperature control have been proposed. All these thermostats interfere with the macromolecule’s “natural” dynamics as given by the Newtonian mechanics. Furthermore, using a single thermostat for an inhomogeneous solute–solvent system can lead to stationary temperature gradients. To avoid this “hot solvent/cold solute” problem, two separate thermostats are frequently applied, one to the solute and one to the solvent. However, such a separate temperature control will perturb the dynamics of the macromolecule much more strongly than a global one and, therefore, can introduce large artifacts into its conformational dynamics. Based on the concept that an explicit solvent environment represents an ideal thermostat concerning the magnitude and time correlation of temperature fluctuations of the solute, we propose a temperature control strategy that, on the one hand, provides a homogeneous temperature distribution throughout the system together with the correct statistical ensemble for the solute molecule while, on the other hand, minimally perturbing its dynamics.

### 1. Introduction

Molecular dynamics (MD) simulations using molecular mechanics (MM) force fields have become a widespread tool to study the equilibrium conformational dynamics of proteins and peptides in solution,<sup>1</sup> including processes of folding and refolding.<sup>2</sup> More recently, also nonequilibrium processes have been simulated in which a protein or peptide is destabilized, for example by applying an external force mimicking the action of an atomic-force microscope,<sup>3–5</sup> by exerting internal mechanical strain,<sup>6,7</sup> by introducing point mutations into the protein sequence,<sup>8,9</sup> or simply by elevating the temperature.<sup>9,10</sup>

The behavior of proteins in solution is steered by the thermodynamic conditions, notably by the temperature. The native state is stable only within a certain temperature range; processes of hot and cold unfolding have been observed.<sup>11</sup> The temperature influences the stability and function of proteins not only directly by changing the relative importance

of the entropy but also indirectly via certain temperature dependent solvent properties such as the dielectric constant<sup>12</sup> or the viscosity.<sup>13</sup> Therefore, if one wants to describe experiments on proteins by MD simulations, the temperature must be properly controlled.

Clearly, an adequate method for temperature control is not the only precondition if one aims at quantitative descriptions of experimental data. In this respect, the quality of the employed force field, the sufficiency of statistical sampling achieved by finite simulation times, and other technical issues are also questions of concern.<sup>14</sup> However, the temperature is of key importance because many experimental observables that can be compared with the information obtained from MD simulations sensitively depend on this parameter. Examples are the temperature factors in X-ray crystallography,<sup>15</sup> the proton exchange and spin relaxation rates in nuclear magnetic resonance spectroscopy [see ref 16 and references therein], and the fluorescence depolarization rates<sup>17</sup> as well as the thermodynamical measures of protein stability.<sup>18</sup>

\* Corresponding author e-mail: tavan@physik.uni-muenchen.de.

The requirement of proper control does not only apply to the temperature, i.e. the average kinetic energy of the system, but also to other ensemble properties (e.g., energy fluctuations) associated with experimental observables. Thus, in a broader sense, the problem of temperature control in MD simulations is also that of generating the correct statistical ensemble (usually canonical or isothermal–isobaric). The accurate generation of a specific statistical ensemble by means of a MD simulation is also relevant for the application of generalized ensemble techniques like replica exchange molecular dynamics<sup>19–21</sup> which has recently become very popular in order to enhance the sampling efficiency. These techniques rely on the assumption that the applied MD method samples the canonical ensemble at the respective temperature.

When simulating macromolecules in solution, the solvent environment, which is essential for the properties of the solute, can either be treated implicitly using continuum approximations or explicitly by including part of the solvent into the simulation system.<sup>14</sup> The following discussions exclusively deal with the latter case and are devoted to the task of controlling the temperature of a solute macromolecule in explicit solvent. This task can comprise additional challenges if nonequilibrium relaxation processes are studied. Here, frequently, energy is released in one part of the system and then dissipated into the rest of the simulation box, e.g. from a solute molecule to the surrounding solvent. Since the kinetics of energy relaxation and heat transport can influence the dynamical properties of the solute,<sup>22</sup> any applied temperature control method should make sure that the natural energy relaxation processes are unimpaired.

Generally, the ideal temperature control scheme for solute–solvent systems would be to simulate the complete simulation system microcanonically, i.e. at constant total energy in the *NVE* ensemble. One can show<sup>23</sup> that, in this situation, an arbitrary subset of degrees of freedom in thermal contact with the rest of the system (e.g., the solute’s kinetic degrees of freedom) will sample the canonical ensemble if the energy fluctuations of the subsystem are insignificant compared with the total energy in the rest of the system. Furthermore, one can show that the subsystem will sample the isothermal–isobaric ensemble if also the subsystem’s volume fluctuations are negligible compared with the volume of the rest of the system. Finally, one expects that all those configurational degrees of freedom of the solute which directly interact with the solvent system will sample the canonical or isothermal–isobaric ensemble, respectively, if additionally the solute–solvent interaction energy is negligible compared with the solvent–internal interaction energy. In MD simulations systems, the latter condition is fulfilled if the solvent atoms by far outnumber those of the solute. Concurrently, by using the *NVE* approach, the solute’s Newtonian dynamics are left completely undisturbed. The *NVE* strategy has been recommended<sup>24</sup> for studies of protein folding kinetics and is occasionally applied<sup>25,26</sup> to eliminate a putative influence of thermostat algorithms on the simulated dynamics.

Unfortunately, the simple *NVE* strategy is not easily applied to extended MD simulations. Numerical inaccuracies

associated with approximation schemes serving to speed up the computations generally lead to a violation of energy conservation. For example, heating may be caused by certain approximate treatments of long-range electrostatic interactions<sup>27,28</sup> or by integrating the equations of motion with multiple-time-step algorithms.<sup>29</sup> Cooling may occur, for instance, if constraining bond lengths or angles with a too loose tolerance<sup>30</sup> or if neighbor lists for the calculation of the van der Waals interactions are not updated frequently enough.<sup>31</sup> The defect of energy conservation could, in principle, be compensated by using an ergostat algorithm which would just scale the velocities of all atoms at every time step by an appropriate factor to keep the total energy exactly constant. However, the rates of algorithmic energy drift can vary among the constituents of an inhomogeneous simulation system leading to unphysical steady state temperature gradients,<sup>32</sup> a problem sometimes referred to as the “hot-solvent/cold-solute” problem.<sup>33</sup> For example, such a gradient can result from an approximate treatment of the electrostatic interactions, which may render a mildly polar solute less affected by algorithmic noise than a strongly polar aqueous solvent.<sup>32,34,35</sup>

Thus, specifically for equilibrium simulations of macromolecules in solution, the applied temperature control has to fulfill an important requirement: The temperature distribution has to be homogeneous throughout the inhomogeneous simulation system. As a strategy guaranteeing such a homogeneous temperature distribution it has been suggested to couple the subsystems independently to separate thermostats.<sup>36</sup> Further below we will check this strategy among others because it is the central aim of this work to determine an optimal strategy for generating a homogeneous temperature distribution in solute–solvent simulation systems.

From a general point of view, the appropriateness of a given temperature control method involves the following three aspects:

- a) Thermodynamics: Does the method generate the expected thermodynamical ensemble in principle (i.e., with simulations of infinite length and in the absence of numerical errors)?
- b) Ergodicity: Does the method generate the expected ensemble within the time typically covered by modern MD simulations?
- c) Dynamics: Is the time dependence and spatial distribution of the thermostatic forces realistic? For a solute in solution, for example, one would prefer to have no such forces at all beyond the thermostating Newtonian interactions with the solvent.

A number of different algorithms has been proposed as realizations of the required thermostats (for a review see ref 36). Each of these algorithms has its specific merits and drawbacks. A critical discussion of these issues is another goal of our study.

For example, the widely used Berendsen thermostat<sup>37</sup> (BT) has the advantage to couple only weakly to the dynamics of the controlled system (see the original paper ref 37 for this issue). On the other hand, it is clear from theoretical considerations that the BT does not create a canonical distribution of microstates,<sup>40</sup> i.e. it introduces artifacts of type

a). Furthermore, the BT violates energy equipartition by redistributing energy from high to low frequency modes, which leads to the so-called “flying-ice-cube effect”.<sup>38,39</sup> It is unclear whether this effect is specific to the Berendsen method and closely related methods or it can occur with any thermostat belonging to the more general class of velocity rescaling algorithms.<sup>38</sup>

The more strongly coupling Nosé-Hoover thermostat<sup>41,42</sup> (NHT) is theoretically expected to generate the canonical distribution of microstates if certain conditions are obeyed thus conforming with the above question a).<sup>43</sup> However, within the time covered by a typical MD simulation, amplitudes of temperature fluctuations were observed which were by 1 order of magnitude larger than those expected for a canonical ensemble.<sup>44</sup> Several studies<sup>42,44–47</sup> have shown that Nosé-Hoover coupled systems do not necessarily acquire ergodicity in a reasonable time [cf. question b) above] if these systems are small, stiff, or at low temperatures. Additionally, by its very construction as a velocity rescaling algorithm, also the NHT could show the flying-ice-cube artifact (although we are not aware of any reports on a corresponding example).

As a reaction to these problems, modifications to both the Berendsen and Nosé-Hoover schemes have been proposed. The most frequently employed variant of the Nosé-Hoover thermostat is the so-called Nosé-Hoover chain,<sup>48</sup> which has been successfully tested by Cheng and Merz<sup>33</sup> as a remedy to the hot-solvent/cold-solute problem. No artifacts or deviations from the canonical ensemble have been reported so far. Only recently, Bussi et al.<sup>49</sup> suggested a modification of the Berendsen scheme in order to reliably generate a canonical distribution for systems that otherwise would sample the microcanonical ensemble. Both, the Nosé-Hoover chain and the modified Berendsen thermostat induce temperature fluctuations of the correct size by artificially scaling the atomic velocities. For systems, however, which anyway sample the canonical ensemble, such a thermostat introduces an unnecessary perturbation of the dynamics, i.e. artifacts of type c). The generic example for such a system is a solute molecule in a sufficiently large explicit solvent system, which, as discussed above, always samples a canonical ensemble although possibly at the wrong temperature because of algorithmic inaccuracies.

Concerning temperature control of macromolecules in solution, we want to show how one can (i) generate the appropriate ensemble for the solute molecule in adequate time, (ii) leave invariant the time scales of energy relaxation and of equilibrium fluctuations, and (iii) guarantee a homogeneous temperature distribution in equilibrium simulations with (iv) minimal perturbation of the solute’s Newtonian dynamics.

For this purpose we will scrutinize in section 2 the existing temperature control scenarios for MD simulations of solvent–solute systems by partially recollecting and partially developing associated theoretical concepts. These considerations will lead to the definition of strategies for a minimally invasive control of a solute temperature. In section 3 we will sketch the methods which we employed in a series of quite extended test simulations on peptides in aqueous solution.

As explained in section 4, these simulations were specifically designed to estimate the extent to which the theoretically expected effects of temperature control do actually modify the properties of a solute peptide. Section 5 discusses the results and suggests a practical procedure ensuring a minimally invasive temperature control.

## 2. Theory

**Thermostats.** The most widely used class of thermostat algorithms is based on the rescaling of atomic velocities. The equation of motion for an atom which belongs to a system under the rule of such a thermostat is

$$m_i \ddot{\mathbf{r}}_i(t) = \mathbf{F}_{i,\text{ff}}(t) - m_i \gamma(t) \dot{\mathbf{r}}_i(t) \quad (1)$$

Here, the acceleration  $\ddot{\mathbf{r}}_i(t)$  of atom  $i$  is caused not only by the forces  $\mathbf{F}_{i,\text{ff}}(t)$  derived from an MM force-field but also by a second term  $\mathbf{F}_{i,\text{therm}}(t) \equiv -m_i \gamma(t) \dot{\mathbf{r}}_i(t)$ , which is proportional to the atom’s velocity  $\dot{\mathbf{r}}_i(t)$  and to a generally time dependent thermostat parameter  $\gamma(t)$ .

In the Berendsen scheme,  $\gamma(t)$  is directly given in terms of the instantaneous kinetic temperature<sup>50</sup>  $T(t)$  by

$$\gamma(t) = \frac{1}{2\tau} \left[ 1 - \frac{T_0}{T(t)} \right] \quad (2)$$

with  $\tau$  denoting the coupling time and  $T_0$  the target temperature. For the Nosé-Hoover<sup>41,42</sup> thermostat,  $\gamma(t)$  is coupled on a time scale  $\tau_{\text{NHT}}$  to  $T(t)$  by the differential equation

$$\frac{d\gamma}{dt} = \frac{1}{\tau_{\text{NHT}}^2} \left[ \frac{T(t)}{T_0} - 1 \right] \quad (3)$$

**Perturbation of the Dynamics.** Every thermostat which is described by eq 1 perturbs the Newtonian dynamics generated by the forces  $\mathbf{F}_{i,\text{ff}}(t)$  through the admixture of additional thermostatic forces  $\mathbf{F}_{i,\text{therm}}(t)$ . For a solute–solvent system, these thermostatic forces introduce artifacts of type c) concerning the dynamics (cf. section 1). The resulting perturbation can be measured for a selected atom  $i$  by the quotient

$$\xi_i^2 \equiv \langle \mathbf{F}_{i,\text{therm}}^2 \rangle_D / \langle \mathbf{F}_{i,\text{ff}}^2 \rangle_D \quad (4)$$

where the brackets  $\langle \dots \rangle_D$  denote temporal averages over a simulation of a given duration  $D$ . The perturbation quotients (4) will depend on the system size and on the particular thermostat, i.e. on the form of  $\gamma(t)$ , as well as on the coupling time.

The perturbation quotients  $\xi_i$  are strictly local measures for the influence of a thermostat on a simulated dynamics. However, one may also consider the local perturbation inflicted on a certain group  $G$  of atoms within a simulation system, e.g. on the C $_{\alpha}$ -atoms of a solute peptide embedded in a solvent environment. Then the root mean quotient  $\bar{\xi}_G \equiv \sqrt{\langle \xi_i^2 \rangle_G}$  over the  $\xi_i^2$  belonging to  $G$  can be used to compare how the dynamics of a solute is perturbed in different solute–solvent systems.

Instead of calculating the averages  $\langle \mathbf{F}_{i,\text{therm}}^2 \rangle_D$  required for the evaluation of the  $\xi_i^2$  directly from a simulation, one can also give a simple estimate for these average square forces.

Assuming a sufficiently large simulation system, the velocities of the individual atoms will negligibly contribute to the temperature  $T(t)$ . Hence, the correlation of  $\gamma^2$  and  $\dot{\mathbf{r}}_i^2$  vanishes and one obtains

$$\langle \mathbf{F}_{i,\text{therm}}^2 \rangle_D = \langle m_i^2 \gamma^2 \dot{\mathbf{r}}_i^2 \rangle_D = m_i^2 \langle \gamma^2 \rangle_D \langle \dot{\mathbf{r}}_i^2 \rangle_D \quad (5)$$

Assuming furthermore that the system is in equilibrium, that the atomic velocity distributions are undisturbed by the thermostat, and that the system is free of internal constraints (such as fixed bond lengths), the mean square velocity of atom  $i$  is expected to be  $\langle \dot{\mathbf{r}}_i^2 \rangle_D \approx 3k_B \hat{T}/m_i$ , where  $\hat{T} = \langle T \rangle_D$  is the average temperature determined from the simulation. Equation 5 then becomes

$$\langle \mathbf{F}_{i,\text{therm}}^2 \rangle_D \approx 3m_i k_B \hat{T} \langle \gamma^2 \rangle_D \quad (6)$$

We will check this estimate by sample simulations and show that it already holds for relatively small systems.

Inserting the estimate 6 into eq 4, one can recognize that the perturbation quotients of a given system which is simulated with different thermostatic strategies solely differ with respect to  $\langle \gamma^2 \rangle_D$ . Thus, in this case, comparisons of the mean square scaling activities  $\langle \gamma^2 \rangle_D$  suffice for the evaluation of different thermostatic strategies concerning the size of local perturbations of the dynamics. However, thermostats do not only cause local perturbations of the Newtonian dynamics but may also interfere with ensemble properties like, for example, size and time scales of the temperature fluctuations.

**Temperature Fluctuations.** In an MD simulation, the statistics of the temperature fluctuations provides a probe for artifacts of type a) and b) pertaining the generation of the desired ensemble (section 1). For a system in contact with a heat bath of temperature  $T_b$ , the distribution of microstates is either given by the canonical or by the isothermal–isobaric ensemble. However, with respect to the temperature fluctuations, both ensembles are equal. The associated probability density  $\rho(\mathbf{T})$  for the instantaneous kinetic temperature is a  $\chi^2$ -distribution<sup>51</sup>

$$\rho(\mathbf{T}) = \frac{(N_{\text{DoF}} \mathbf{T} / 2T)^{N_{\text{DoF}}/2}}{\Gamma(N_{\text{DoF}}/2) T} \exp\left[-\frac{N_{\text{DoF}} \mathbf{T}}{2T}\right] \quad (7)$$

where  $T = T_b$  is the expectation value of  $\mathbf{T}$ ,  $N_{\text{DoF}}$  is the number of degrees of freedom (DoF) of the system, and  $\Gamma(\dots)$  denotes the Euler  $\Gamma$ -function. Consequently, the variance  $\sigma_T^2$  of the temperature fluctuations is

$$\sigma_T^2 = \frac{2T^2}{N_{\text{DoF}}} \quad (8)$$

Under the influence of a thermostat, the statistics can deviate from what is expected for a canonical ensemble. This deviation constitutes a measure for the global influence of the thermostat and for how close a simulation is to sampling the canonical ensemble. In the limit  $N_{\text{DoF}} \rightarrow \infty$ , eq 7 becomes a normal distribution.

The size of  $\sigma_T^2$  together with the autocorrelation time<sup>50</sup>  $\tau_c$  of the temperature fluctuations critically influences the accuracy with which the equilibrium temperature  $T$  is determined by a given simulation. The variance  $\sigma_T^2$  of the

time averages  $\hat{T}$  obtained from a set of equilibrium simulations with durations  $D$  can be estimated<sup>50</sup> to be

$$\sigma_{\hat{T}}^2 = 2\sigma_T^2 \frac{\tau_c}{D} \quad (9)$$

In order to judge whether a particular strategy is suited to correctly tune the temperature  $T$ , one has to perform a test simulation which is long enough to determine  $T$  with sufficient accuracy. For a small solute (large  $\sigma_T^2$ ) with a correlation time  $\tau_c$  longer than 10 ps, an accuracy of 1 K may require simulation times of up to 10 ns.

**Power of a Thermostat.** By means of the observables introduced above, one can judge to what extent a thermostat can perturb the dynamical and equilibrium properties of a solute in solute–solvent simulations. Such perturbations can, of course, be avoided by using no thermostat at all. However, as outlined in section 1, this approach is generally not feasible because algorithmic inaccuracies, which are inevitable in large scale simulations using efficient MD codes, represent heat drains or sources that have to be compensated.

To properly tune this compensation, we consider the work performed by the thermostatic forces  $\mathbf{F}_{i,\text{therm}}(t)$  on the atoms  $i$  for an ensemble of simulation systems with the temperature  $T(t) = \langle \mathbf{T}(t) \rangle_{\text{ens}}$ . The ensemble average power exerted by the thermostat on a given atom  $i$  is

$$\beta_i(t) = \langle \mathbf{F}_{i,\text{therm}}(t) \cdot \dot{\mathbf{r}}_i(t) \rangle_{\text{ens}} \quad (10)$$

Using the definition of  $\mathbf{F}_{i,\text{therm}}$  [see eq 1] and the Berendsen expression 2 for  $\gamma$  leads to

$$\beta_i(t) = \frac{1}{\tau} \langle \varepsilon_{i,\text{kin}}(t) [T_0/T(t) - 1] \rangle_{\text{ens}} \quad (11)$$

with the usual definition for the kinetic energy  $\varepsilon_{i,\text{kin}}(t)$  of atom  $i$ . Employing once more the assumption of a negligible correlation between the velocity [and, thus, the kinetic energy  $\varepsilon_{i,\text{kin}}(t)$ ] of a single atom and the kinetic temperature  $\mathbf{T}(t)$  of the system, one obtains

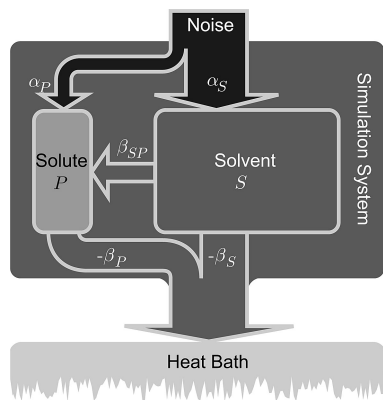
$$\beta_i(t) = \frac{3k_B T_i(t)}{2\tau} [T_0/T(t) - 1] \quad (12)$$

where  $k_B$  is the Boltzmann constant, and  $T_i(t) \equiv 2/3k_B \langle \varepsilon_{i,\text{kin}}(t) \rangle_{\text{ens}}$  is the ensemble average temperature of atom  $i$ . For equilibrated systems the ensemble averages employed in eq 12 can be replaced by temporal averages  $\langle \dots \rangle_D$ . This allows to calculate for every subsystem  $\kappa$  from a simulation the (time) average thermostatic power

$$\hat{\beta}_\kappa = \frac{k_B \hat{T}_\kappa}{2\tau} [T_0/\hat{T} - 1] \quad (13)$$

per degree of freedom using the average temperature  $\hat{T}_\kappa \equiv \langle \mathbf{T}_\kappa \rangle_D$  of the subsystem  $\kappa$ , the corresponding average  $\hat{T} \equiv \langle \mathbf{T} \rangle_D$  of the temperature  $\mathbf{T}(t)$  controlled by the BT, and the thermostat parameters  $T_0$  and  $\tau$ .

Further below we will use eq 13 to determine the thermostatic power exerted by a BT on a solute peptide from sample simulations. These data will be used to check the validity of a heat conduction model which we will now introduce to analyze the hot-solvent/cold-solute problem occasionally hampering MD simulations of inhomogeneous systems.<sup>33</sup>



**Figure 1.** Heat flow model representing specifically the “hot-solvent/cold-solute” case for an inhomogeneous system consisting of two subsystems with different heating rates. The simulation system is coupled to a single thermostat, representing an external heat bath. Bright and dark colors code low and high temperatures, respectively. Heat flows driven by temperature gradients and heat sources are marked by arrows. A detailed discussion is given in the text.

**Heat Flow Model.** In simulations of solute–solvent systems, the algorithmic heat drains or sources may be inhomogeneously distributed and, thus, the temperature may likewise be inhomogeneous. According to requirement (iii) stated at the bottom of the Introduction, such inhomogeneous temperature distributions should be avoided. Figure 1 sketches a heat flow model from which one can derive strategies for the reliable control of the solute temperature. As drawn, the model refers to a particular strategy employing a single thermostat for the whole system. For further reference we denote this strategy by **G**.

The model depicted in Figure 1 consists of two subsystems  $\kappa \in \{P, S\}$  with  $P$  denoting the solute and  $S$  denoting the solvent. The powers  $\alpha_\kappa$  of algorithmic heating per DoF are assumed to be constant and homogeneous within the subsystems. Furthermore, the temperature is assumed to be homogeneous within each subsystem  $\kappa$ , i.e. for the atomic temperatures  $T_i$  we have  $T_i = T_\kappa$  for all  $i \in \kappa$ . According to eq 12, the ensemble average work  $\beta_\kappa(t)$  exerted on atom  $i$  per unit time by the global thermostat then only depends on whether  $i$  is part of  $P$  or  $S$ , respectively. Thus, for the subsystems  $\kappa \in \{P, S\}$  the respective thermostatic powers per DoF are given by

$$\beta_\kappa(t) = \frac{k_B T_\kappa(t)}{2\tau} [T_0/T(t) - 1] \quad (14)$$

If the local temperatures  $T_P$  and  $T_S$  differ, as is assumed in Figure 1, there will be a net heat flow

$$\beta_{SP}(t) = \frac{k_B [T_S(t) - T_P(t)]}{2\tau_{SP}} \quad (15)$$

between  $S$  and  $P$ , which we assume to depend linearly on the temperature difference. Here, the time constant  $\tau_{SP}$  characterizes the thermal coupling of the subsystems.

The heat flowchart shown in Figure 1 immediately suggests stationarity conditions. In the steady state, the net heat flow must individually vanish for each of the two subsystems, i.e.

$$\alpha_P + \beta_{SP} + \beta_P = 0 \quad (16)$$

$$\alpha_S - \beta_{SP} + \beta_S = 0 \quad (17)$$

Now suppose for a moment that the temperature distribution is homogeneous throughout the system, i.e.  $T_P(t) = T_S(t) = T(t)$ . According to eq 14 the thermostatic powers  $\beta_P(t)$  and  $\beta_S(t)$  exerted on the subsystems are then equal, and, by eq 15, the heat flow  $\beta_{SP}(t)$  between  $S$  and  $P$  vanishes. Equations 16 and 17 then immediately require as the stationarity condition that  $\alpha_P = \alpha_S$ , i.e. that the heat sources in the subsystems work at equal powers.

If this is not the case ( $\alpha_P \neq \alpha_S$ ), the temperature cannot be homogeneously distributed in the stationary state, and, by eq 15, there will be a nonvanishing heat exchange  $\beta_{SP} \neq 0$  between the subsystems. As a result, a steady state temperature difference is inevitable whenever, upon applying scenario **G**, a single global thermostat is used to thermostate a system exhibiting inhomogeneities with respect to the rates  $\alpha_\kappa$  of algorithmic heating. This is the origin of the hot-solvent/cold-solute problem as described e.g. in ref 34.

**Separate Thermostats.** To avoid temperature inhomogeneities, it has become a standard in simulations of macromolecules to couple *separate* thermostats to the subsystems.<sup>36,47,52–57</sup>

We will denote this temperature control scenario by **P**. In the following discussion of scenario **P**, we will concentrate on the temperature control of the solute  $P$ , assuming that the temperature of the solvent  $S$  is reliably controlled. Such a reliable control can be achieved by a solvent thermostat combining a coupling time  $\tau_S$  on the subpicosecond time scale (e.g.,  $\tau_S = 0.1$  ps) with a target temperature  $T_{0,S}$  equal to the intended temperature. This choice of thermostat tuning actually is the standard (see e.g. refs 33, 36, 37, 47, 52–56), and, thus, we call it the classical setup.

For a scenario **P**, in which a separate BT is coupled to a (thermally homogeneous) solute  $P$ , the controlled temperature  $T(t)$  is the solute temperature  $T_P(t)$ . Thus, we obtain from eq 14 the simplified expression

$$\beta_P(t) = \frac{k_B}{2\tau_P} [T_0 - T_P(t)] \quad (18)$$

for the power of the thermostat acting on  $P$ . With eqs 18 and 15, the solute’s stationarity condition 16 may be rewritten as

$$\alpha_P + \frac{k_B(T_S - T_P)}{2\tau_{SP}} + \frac{k_B(T_{0,P} - T_P)}{2\tau_P} = 0 \quad (19)$$

where  $T_{0,P}$  denotes the target temperature, and  $\tau_P$  denotes the coupling time of the solute thermostat. The first term characterizes the algorithmic heating within  $P$ , the second term characterizes the heat flow between  $P$  and  $S$ , and the third term characterizes the power  $\beta_P$  of the thermostat separately coupled to  $P$ .

Equation 19 is the quintessence of our linear heat flow model and may be used to predict the effects of three different thermostatic strategies within scenario **P**. In all these strategies,  $S$  is coupled to a classical BT and  $P$  is decoupled from this thermostat. The three strategies are as follows:

**P.1** The solute  $P$  is coupled to a classical thermostat. Here, the use of a correspondingly small coupling time  $\tau_P \approx 0.1$

ps is the standard.<sup>36,52–56,58</sup> For such small  $\tau_P$ , eq 19 is completely dominated by the thermostatic term. The reason for this dominance is that  $\tau_P$  is by at least 1 order of magnitude smaller than the solute–solvent coupling time  $\tau_{SP}$ , which is typically larger than 1 ps (see further below). Neglecting the heat flow contribution, the deviation  $T_P - T_{0,P}$  from the target temperature is given by  $2\alpha_P\tau_P/k_B$ . For moderate algorithmic heating rates  $\alpha_P$ , this deviation is expected to be small because of the short time scale  $\tau_P$ .

**P.2** No separate thermostat is coupled to the solute  $P$ , i.e.  $\tau_P \rightarrow \infty$ , and solely the thermostatted solvent  $S$  acts as a heat bath. We call this strategy “noninvasive” because it does neither alter the Newtonian dynamics nor the energy relaxation properties of  $P$ . The expected temperature difference  $T_P - T_S = 2\alpha_P\tau_{SP}/k_B$  will be small if the local heating  $\alpha_P$  is negligible on the time scale  $\tau_{SP}$  of the thermal coupling between the subsystems.

**P.3** The solute is coupled to a thermostat with a very large coupling time  $\tau_P \gg \tau_{SP}$  to realize a “constant heat flow” (CHF) approach. As suggested by the heat balance eq 19, a homogeneous ( $T_P = T_S$ ) and stationary temperature distribution only requires that the thermostatic power  $\beta_P$  cancels the power  $\alpha_P$  of algorithmic heating, i.e.,  $\alpha_P = k_B(T_P - T_{0,P})/2\tau_P$ . This condition can be satisfied for an arbitrarily large coupling time  $\tau_P$  by a proper choice of the target temperature  $T_{0,P}$ . In the limit  $\tau_P \rightarrow \infty$ , the thermostat variable  $\gamma$  in eqs 1 and 2 becomes a constant  $\gamma_P$ , and the thermostat scheme may actually be described by this single parameter. At large  $\tau_P$ , the thermostat works in a heating/cooling limit as a constant heat source/drain, and this activity solely serves to maintain the energy balance. Because of eq 2, the perturbation of the Newtonian dynamics [cf. eq 1] inflicted by such a CHF thermostat can be made very small. Therefore, we call the CHF approach to the solute’s temperature control, which is applicable to non-negligible local heating rates  $\alpha_P$ , “minimally invasive”.

To set up a CHF simulation as required in strategy **P.3**, the a priori unknown power  $\alpha_P$  of algorithmic heating has to be determined in order to specify the constant thermostat parameter  $\gamma_P$ , or, equivalently, the parameters  $T_{0,P}$  and  $\tau_P$  if a traditional Berendsen thermostat is used in the heating/cooling limit. To this end, the solute temperature  $T_P$  has to be measured in two test simulations with different heating powers  $\beta_P$  of the thermostat. The two heat balance eqs (19) of these tests then constitute a system of linear equations which determines the unknown parameters  $\beta_{SP}$  and  $\alpha_P$ . A detailed description of the setup protocol is given in Appendix A.

In the following we will examine the temperature control strategies **G** and **P.1–P.3** introduced above by test simulations. Based on these results, we formulate guidelines for a temperature control satisfying the four conditions summarized at the bottom of the Introduction.

### 3. Methods

**MD Simulation Techniques.** The software packages EGO-MMII<sup>27</sup> and GROMACS<sup>59</sup> were used in several series of MD simulations. Besides EGO we also applied GROMACS because it provides an NHT in addition to a BT,

because it is computationally efficient for very small systems, and because it can provide data for a crosscheck of results. In EGO the electrostatic interactions are treated combining structure-adapted multipole expansions<sup>60,61</sup> with a moving-boundary reaction-field approach<sup>62</sup> and a multiple-time-step integration.<sup>29,63</sup> In the GROMACS simulations we used the PME method<sup>28</sup> with a 10 Å cutoff for the real space contribution, with a grid spacing of 0.5 Å, and with a sixth order interpolation of the charges to the grid. For both EGO and GROMACS simulations, the van der Waals interactions were truncated at 10 Å. If not stated otherwise, the simulations were carried out with explicit solvent using periodic boundary conditions and with a BT ( $\tau_S = 0.1$  ps,  $T_{0,S} = 300$  K) rapidly coupled to the solvent to guarantee that the solvent was closely kept at the target value. Bond lengths were constrained using the M-SHAKE algorithm<sup>30</sup> with relative tolerances of  $10^{-4}$  when using GROMACS, which is the recommended default value, and  $10^{-6}$  when using EGO, which is hard-coded in the source code in this case.

We applied different simulation protocols to vary the heating properties within the simulation systems. Here, the first parameter was the software used for simulation, which we denote by E for EGO or G for GROMACS. Since the M-SHAKE algorithm is known<sup>30</sup> to have a cooling effect, we varied the number of constraints by either constraining no bonds at all (N), only bonds involving a hydrogen atom (H), or all bonds (A). The last parameter which presumably influences the heating in the system is the length  $\Delta t$  of the basic integration time step, which we simply denote by its value in femtoseconds. Thus, a standard EGO simulation (constraints on bonds involving hydrogen atoms and  $\Delta t = 1$  fs) would be denoted by E/H/1.

**Model Systems.** The first model system was a polyaniline octapeptide (8ALA) with charged termini described by the GROMOS96 force field<sup>64</sup> and embedded in a cubic box of 20 Å edge-length containing 236 simple point charge (SPC) water molecules.<sup>65</sup> The number of DoF for the peptide then is 153/143/103 for N-/H-/A-constraining, respectively. The starting conformation was always fully extended. The system was equilibrated for 300 ps during which solute and solvent were coupled to separate BTs ( $\tau_P = \tau_S = 0.1$  ps,  $T_{0,S} = T_{0,P} = 300$  K). The second model system was an alanine dipeptide (ALDI) described by the CHARMM22 force field<sup>66</sup> in a cubic box of 21.3 Å edge length containing 324 water molecules modeled by the transferable three-point intermolecular potential (TIP3P).<sup>66,67</sup> Here, the number of peptide DoF is 66/54/45 for N-/H-/A-constraining, respectively. The system was prepared as described for 8ALA, except that the equilibration time was only 100 ps.

**MD Simulations.** A first series of seven MD simulations of 8ALA in SPC water served to study the various situations encountered in the temperature control of inhomogeneous systems. Table 1 associates acronyms to these simulations and lists the employed parameters. In particular, in the last simulation G/A/2\_P.3, the CHF approach was applied to the peptide. Using the data from the preceding simulation G/A/2\_P.2 ( $\tau_P = \infty$ ,  $\hat{T}_P = 293.4$  K), in which only  $S$  was coupled to a classical BT and the data from an independent 10 ns test simulation with an additional CHF thermostat coupled

**Table 1.** Simulation Names and Associated Parameters in Series #1<sup>a</sup>

name	software	protocol			thermostat parameters			
		C	$\Delta t/\text{fs}$	$D/\text{ns}$	$\tau_{\text{sys}}/\text{ps}$	$\tau_S/\text{ps}$	$\tau_P/\text{ps}$	$T_{0,P}/\text{K}$
E/H/1_G	EGO	H	1	20	0.1	—	—	—
E/H/2_G	EGO	H	2	20	0.1	—	—	—
E/H/2_P.2	EGO	H	2	20	—	0.1	—	—
G/H/2_P.2	GROMACS	H	2	20	—	0.1	—	—
G/A/2_P.2	GROMACS	A	2	20	—	0.1	—	—
G/A/2_P.1	GROMACS	A	2	20	—	0.1	0.1	300
G/A/2_P.3	GROMACS	A	2	20	—	0.1	500	2340

<sup>a</sup> The simulation names code the varied parameters and temperature control scenarios. C specifies the type of bond length constraints,  $\Delta t$  the size of the basic integration time step, and  $D$  the duration of the simulation. The parameters  $\tau$  specify the coupling times of the BTs coupled to the whole system (sys), to the solvent (S), or to the solute (P).  $T_{0,P}$  is the target temperature of a thermostat coupled to the solute. The solute peptide was 8ALA in SPC water. See the text for further information.

**Table 2.** Simulation Parameters in Series #2<sup>a</sup>

name	software	protocol			thermostat parameters			
		C	$\Delta t/\text{fs}$	$D/\text{ns}$	$\tau_{\text{sys}}/\text{ps}$	$\tau_S/\text{ps}$	$\tau_P/\text{ps}$	$T_{0,P}/\text{K}$
CHF.0	GROMACS	A	2	200 × 2	—	0.1	—	—
CHF.1	GROMACS	A	2	200 × 2	—	0.1	500	2340
CHF.2	GROMACS	A	2	200 × 2	—	0.1	500	4800
CHF.3	GROMACS	A	2	200 × 2	—	0.1	500	7700
CHF.4	GROMACS	A	2	200 × 2	—	0.1	500	11100
CLS.1	GROMACS	A	2	200 × 2	—	0.1	0.1	300
CLS.2	GROMACS	A	2	200 × 2	—	0.1	0.1	340

<sup>a</sup> The model peptide was 8ALA in SPC water at  $\hat{T}_S = 300$  K. Except for the simulation set CHF.0, in which only  $T_S$  was controlled, separate BTs were applied to S and P. See the caption to Table 1 for further information.

to the peptide ( $\tau_P = 500$  ps,  $T_{0,P} = 4800$  K,  $\hat{T}_P = 307.9$  K), the unknown parameters in eq 19 were determined as described in the section 2. We found the values  $\alpha_P = -2.04k_B$  K/ps for the algorithmic heating rate and  $\tau_{SP} = 1.61$  ps for the solute–solvent coupling time, which actually is in the picosecond time range as claimed further above. To realize a CHF thermostat maintaining the peptide at  $T_P \approx 300.0$  K, these values were inserted into eq 19 yielding a “target temperature”  $T_{0,P} = 2340$  K. If the assumptions underlying our heat-flow model are correct, this choice should compensate through  $\beta_P = -\alpha_P$  the algorithmic energy drift in the G/A/2\_P.3 simulation.

The setup of a second series of simulations was chosen such that the effects of the local temperature and of a thermostat on the dynamics of 8ALA can be studied. We performed seven sets of 200 simulations each. Every single simulation had a duration of 2 ns, amounting to 400 ns per set and a total of 2.8  $\mu\text{s}$  of simulation time. The simulation parameters are summarized in Table 2. All simulations were performed with the G/A/2 protocol. In the first set (CHF.0), no thermostat was coupled to the peptide, while in the following four sets (CHF.1 to CHF.4) a BT targeting at increasingly large temperatures  $T_{0,P}$  was coupled in an extremely slow fashion to the peptide. In the last two sets (CLS.1 and CLS.2), a separate classical BT was coupled to the peptide using either the same ( $T_{0,P} = 300$  K) or a slightly higher ( $T_{0,P} = 340$  K) target temperature as compared to  $T_{0,S}$ . The 200 initial conditions were obtained by taking snapshots every 20 ps from a 2 ns preparatory simulation at 300 K, with the peptide’s  $C_\alpha$  atoms harmonically coupled to their initial coordinates of an extended conformation.

To compare the different thermostatic strategies discussed in section 2, we determined the corresponding thermostatic

**Table 3.** Simulation Parameters in Series #3<sup>a</sup>

peptide	software	protocol			thermostat parameters		
		C	$\Delta t/\text{fs}$	$D/\text{ns}$	$\tau_{\text{sys}}/\text{ps}$	$\tau_P/\text{ps}$	$T_{0,P}/\text{K}$
8ALA	GROMACS	N	1	0.25	NHT	0.064	300
8ALA	GROMACS	N	1	0.25	NHT	0.256	300
8ALA	GROMACS	N	1	0.25	NHT	1.024	300
8ALA	GROMACS	N	1	0.25	NHT	4.096	300
8ALA	GROMACS	N	1	0.25	NHT	16.384	300
8ALA/ALDI	EGO	N	1	0.25	BT	0.001	300
8ALA/ALDI	EGO	N	1	0.25	BT	0.004	300
8ALA/ALDI	EGO	N	1	0.25	BT	0.016	300
8ALA/ALDI	EGO	N	1	0.25	BT	0.064	300
8ALA/ALDI	EGO	N	1	0.25	BT	0.256	300
8ALA/ALDI	EGO	N	1	0.25	BT	1.024	300
8ALA/ALDI	EGO	N	1	0.25	BT	4.096	300

<sup>a</sup> For nomenclature see the caption to Table 1. In all simulations the solvent was coupled with  $\tau_S = 0.1$  ps to a Nosé-Hoover (NHT) or a Berendsen (BT) thermostat, respectively.

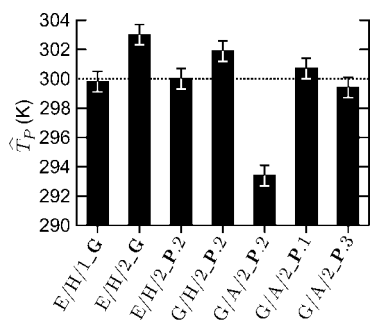
forces (eqs 5 and 6) and perturbation ratios (eq 4) in a third series of relatively short 250 ps simulations. Simulations were performed for 8ALA with varying coupling strengths and BTs and NHTs, respectively. Additionally, we determined the thermostatic forces and the perturbation ratio also for ALDI and Berendsen coupling again varying the coupling strength. The simulation parameters of the third series are given in Table 3. As these simulations served to compare thermostatic and force-field forces, no bond lengths were constrained thus eliminating constraint forces.

Finally, a fourth series of slightly more extended simulations (500 ps) was designed to examine how the solute’s variance of temperature fluctuations (cf. the corresponding paragraph in section 2) is affected by the coupling times of a BT. We studied 8ALA and ALDI in water and in vacuum

**Table 4.** Simulation Parameters in Series #4<sup>a</sup>

system		protocol				thermostat parameters	
peptide	environment	software	C	$\Delta t/\text{fs}$	$D/\text{ns}$	$\tau_P/\text{ps}$	$T_{0,P}/\text{K}$
8ALA	water/vac	EGO	A	2	0.5	0.001	300
8ALA	water/vac	EGO	A	2	0.5	0.004	300
8ALA	water/vac	EGO	A	2	0.5	0.016	300
8ALA	water/vac	EGO	A	2	0.5	0.064	300
8ALA	water/vac	EGO	A	2	0.5	0.256	300
8ALA	water/vac	EGO	A	2	0.5	1.024	300
8ALA	water/vac	EGO	A	2	0.5	4.096	300
ALDI	water/vac	EGO	H	2	0.5	0.001	300
ALDI	water/vac	EGO	H	2	0.5	0.004	300
ALDI	water/vac	EGO	H	2	0.5	0.016	300
ALDI	water/vac	EGO	H	2	0.5	0.064	300
ALDI	water/vac	EGO	H	2	0.5	0.256	300
ALDI	water/vac	EGO	H	2	0.5	1.024	300
ALDI	water/vac	EGO	H	2	0.5	4.096	300

<sup>a</sup> For nomenclature see the caption to Table 1. BTs were used for solvent and solute. The solvent was coupled with a coupling time of 0.1 ps.



**Figure 2.** Average peptide temperature  $\hat{T}_P$  observed in the first series of simulations on 8ALA in SPC water. The associated acronyms and parameters characterizing the members of the series are given in Table 1.

by E/H/2 simulations using the same set of coupling times as in series #3. Table 4 summarizes the simulations of the last series.

## 4. Results and Discussion

**Temperature Control Scenarios.** As outlined above, the series of equilibrium simulations on the model peptide 8ALA in explicit water as characterized by Table 1 served to exemplify the problems connected with the temperature control of inhomogeneous systems. Figure 2 shows the average peptide temperatures obtained in these sample simulations. Using eq 9, the remaining uncertainty of these average temperatures was estimated to  $\sigma_{\hat{T}_P} < 0.7$  K. The solvent temperatures were 300.0 K where not mentioned explicitly.

In simulation E/H/1\_G, we used the standard simulation protocol for EGO (see section 3 for details), which includes a classical BT coupled to the whole simulation system and, thus, represents an example for scenario G outlined in section 2. Neither the resulting temperatures of the peptide (cf. Figure 2) nor of the solvent showed any statistically significant deviations from the 300 K target value suggesting that in E/H/1\_G the algorithmic noise was weak.

Figure 2 indicates that this behavior was lost in simulation E/H/2\_G, in which the basic integration time step  $\Delta t$  was doubled to 2 fs. For our sample system, this doubling of  $\Delta t$

led to a 3.0 K increase of the peptide temperature, indicating that the modified simulation setup has caused certain algorithmic inaccuracies. When using EGO, the choice of a larger  $\Delta t$  is expected to reduce the accuracy of the integration algorithm because the employed highly efficient multiple-time-step algorithm does not exactly guarantee energy conservation and because the corresponding violation increases with the size of  $\Delta t$  (see refs 63 and 29 for a discussion). According to Figure 2, the combination of a global Berendsen thermostat with a reduced accuracy of integration in simulation E/H/2\_G apparently led to a moderately elevated temperature for the peptide and to a slightly (0.3 K) cooler temperature for the larger solvent system. Nevertheless, the temperature of the total system was accurately kept at 300.0 K by the thermostat. Apart from changed signs (hot solute in cold solvent), this result is an example for the classical problem reported in the literature,<sup>32,34,68</sup> which can arise in scenario G from indiscriminately coupling a thermostat to all parts of an inhomogeneous system and which is described by the heat flow model sketched in Figure 1.

However, as demonstrated by the average peptide temperature displayed in Figure 2 for simulation E/H/2\_P.2, this temperature control problem was eliminated by simply decoupling the peptide from the thermostat, i.e. by realizing scenario P.2. This observation suggests that in the E/H/2 simulations the solvent experiences a considerable cooling, whereas the level of algorithmic noise within the peptide is very low. According to our experience, such a decoupling of the solute is a proper solution for most temperature control problems which can occur in simulations of inhomogeneous systems using either EGO or GROMACS.

The fact that the application of scenario P.2 cannot always remove such problems is demonstrated by the results of simulation G/H/2\_P.2, which was carried out with GROMACS using the same settings as in the EGO simulation E/H/2\_P.2. According to Figure 2, in the G/H/2\_P.2 simulation the peptide was by about 2 K too hot, indicating that the rate  $\beta_{SP}$  of heat transport from the peptide  $P$  into the solvent  $S$  was too slow to compensate the algorithmic heating  $\alpha_P > 0$  of the solute occurring in this case.



It may be expected that introducing additional M-SHAKE constraints into the peptide system leads to a local cooling,<sup>30</sup> which might compensate the observed algorithmic heating of  $P$ . This is the reason why we carried out simulation G/A/2\_P.2, which differs from G/H/2\_P.2 only in the number of constraints (50 vs 10) within the peptide. In fact, Figure 2 displays for simulation G/A/2\_P.2 a peptide temperature which is by 6.6 K cooler than that of the solvent, implying that the original heating has been overcompensated by the local cooling. A deviation of this size is unacceptable in simulations serving to probe the equilibrium properties of the solute. Thus, the simulation setup G/A/2 is a typical case in which one of the two remaining temperature control strategies **P.1** and **P.3** described in the section 2 should be applied.

Hence, in simulation G/A/2\_P.1 we utilized a separate classical BT for temperature control of the peptide, while in simulation G/A/2\_P.3 we applied a CHF thermostat. Figure 2 shows that in both cases there is no significant deviation of the observed peptide temperatures from the solvent temperature. Both methods are capable of correctly thermostating the solute. For the CHF thermostat we conclude that the choice of parameters (cf. section 3) was correct and that the underlying heat flow model describes the situation in this case. This success has motivated us to further scrutinize the validity of this model.

**Validity of the Heat Flow Model.** The second quite extended series of simulations (see Table 2 for the parameters) can serve to assess the validity of eq 19, which expresses the contents of the model. With eq 18 the model 19 can be equivalently reformulated as

$$T_P = T_S + \frac{2\tau_{SP}}{k_B}(\beta_P + \alpha_P) \quad (20)$$

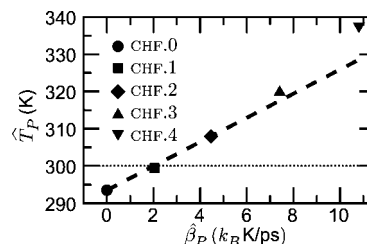
showing that the solute temperature  $T_P$  should depend linearly on the heating power  $\beta_P$  of the solute thermostat. To specify the unknown parameters  $\alpha_P$  and  $\tau_{SP}$  in eq 20, one needs measurements of  $T_S$  and  $T_P$  from two simulations employing different heating powers  $\beta_P$ .

Estimates  $\hat{\beta}_P$  for the heating powers  $\beta_P$  can be determined from simulations by evaluating eq 13 specifically for the case of a solute thermostat, i.e. for  $\kappa = P$ ,  $T_0 = T_{0,P}$ ,  $\tau = \tau_P$ , and  $\hat{T} = \hat{T}_P$ . One obtains

$$\hat{\beta}_P = \frac{k_B}{2\tau_P}[T_{0,P} - \hat{T}_P] \quad (21)$$

which is, up to the use of different averages, identical to eq 18. Thus, at a constant coupling time  $\tau_P$ , the heating power  $\beta_P$  is steered by the choice of the target temperature  $T_{0,P}$  and measured through the average peptide temperature  $\hat{T}_P$ . Therefore, the linear relationship 20 between  $T_P$  and  $\beta_P$  can be checked by comparing with data points  $(\hat{T}_P, \hat{\beta}_P)$  obtained from simulations employing different target temperatures  $T_{0,P}$ .

An inspection of the first five simulation sets in series #2 listed in Table 2 shows that this set qualifies both for the evaluation of the unknown parameters in eq 20 and for the check of this linear equation. In all simulation sets of series #2, the simulation protocol was G/A/2 just like in the

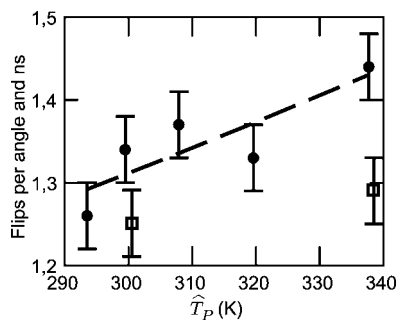


**Figure 3.** The average temperatures  $\hat{T}_P$  of the peptide ALA8 (in SPC water at  $\hat{T}_S = 300$  K) resulting from constant local heating with different powers  $\beta_P$  in the simulation sets CHF.0 to CHF.4 of series #2 (cf. Table 2). The prediction of linear heat flow model eq 20 is drawn as a dashed line, and the solvent temperature  $\hat{T}_S$  is indicated as a dotted line (see the text for explanation).

simulation G/A/2\_P.2 of the first series. However, the temperature control scenario **P.2** (no separate thermostat for the peptide) was employed only in simulation CHF.0. In the remaining CHF simulations a BT was coupled to  $P$  using an extremely slow coupling time  $\tau_P = 500$  ps combined with a large and increasing target temperature (cf. Table 2). According to eq 21 this choice leads to a heating power  $\beta_P$  of this thermostat, which increases from simulation CHF.0 ( $\beta_P = 0$ ) to simulation CHF.4. Figure 3 shows the observed stationary peptide temperatures  $\hat{T}_P$  as a function of the observed heating power  $\hat{\beta}_P$ . In the case of the simulation set CHF.0 (black dot) the result of simulation G/A/2\_P.2 (cf. Figure 2) is closely recovered because the same temperature control setting **P.2** was applied, i.e.  $\hat{T}_P$  was by 6.5 K smaller than the solvent temperature of  $\hat{T}_S = 300$  K. With nonzero and successively growing  $\hat{\beta}_P$  the peptide temperature  $\hat{T}_P$  is seen to increase.

The dashed line in Figure 3 expresses the linear relation 20 between  $\beta_P$  and  $T_P$ . The required parameters were determined as  $\alpha_P = -2.02k_B$  K/ps and  $\tau_{SP} = 1.60$  ps from the simulation sets CHF.0 and CHF.2. Therefore, the dashed line linearly interpolates between the data points  $(\hat{\beta}_P, \hat{T}_P)$  of these two simulation sets. The above values of the parameters  $\alpha_P$  and  $\tau_{SP}$  closely agree with those calculated earlier (see Methods) for setting up the CHF thermostat used in simulation G/A/2\_P.3. This result is expected because in both cases the parameters  $\alpha_P$  and  $\tau_{SP}$  were computed from simulations employing the same parameters.

In simulation set CHF.1, the peptide temperature was nearly identical to  $\hat{T}_S$  with  $\hat{T}_P = 299.5$  K (black square in Figure 3) because here the thermostat parameters were chosen equal to those of the simulation G/A/2\_P.3 (series #1), which realizes the **P.3** strategy. The temperature  $T_P$  predicted for CHF.1 by the dashed line deviates by only 0.5 K from the observed average. This deviation is probably significant because the temperature averages shown in the figure are extremely well converged ( $\sigma_{T_P} < 0.1$  K) due to the extended statistics. If a similar interpolation would be constructed using the data from the simulation sets CHF.3 or CHF.4 instead of CHF.2, the error in the prediction for CHF.1 would increase to 1.1 K or 2.2 K, respectively, with increasing violation of the approximate linear relation 20 between  $\beta_P$  and  $T_P$ . In the case of 8ALA in explicit water, the assumption of a linear thermal coupling between solvent



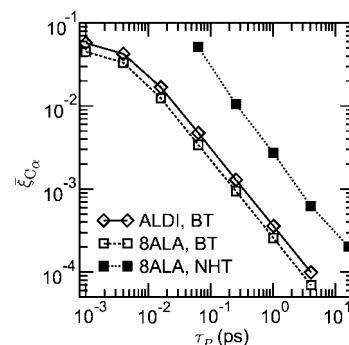
**Figure 4.** Temperature dependence of the peptide backbone dynamics of 8ALA. The graph shows the average number of transitions per angle and nanosecond of the  $\psi$ -dihedral angles between the  $\alpha$ -type region  $[-60^\circ, -30^\circ]$  and the  $\beta$ -type region  $[95^\circ, 145^\circ]$  for the five CHF simulation sets (filled circles) and the two CLS sets (empty squares) over the observed average peptide temperature  $\hat{T}_P$ . The error bars give the range of plus/minus one standard deviation. Additionally, an Arrhenius<sup>69</sup> model (dashed line) fitted to the CHF data is plotted. The simulation parameters are summarized in Table 2.

and solute (eq 15), thus, obviously breaks down if  $\hat{T}_P$  deviates by more than about 10 K from  $\hat{T}_S$ , which is probably also true for related simulation systems. In test simulations serving to set up a CHF thermostat through eq 19, the deviation  $|\hat{T}_P - \hat{T}_S|$  should, thus, be smaller than about 10 K if one wants to guarantee an accurate tuning of  $T_P$  in applications of strategy P.3.

**Backbone Dynamics.** As we have seen further above, the use of an inappropriate strategy for temperature control can lead to peptide temperatures considerably deviating from that of the solvent. It seems likely that such a deviation can entail an altered conformational dynamics of the peptide. To check this expectation, we analyzed the second simulation series also in this respect. Due to the extremely slow thermostat coupling employed in CHF.0 to CHF.4, here, the dynamics should be exclusively affected by differences in the peptide temperatures.

Figure 4 shows how the kinetics of conformational transitions in 8ALA is modified by  $T_P$  in CHF.0–4 (black dots). This kinetics is measured by local flip rates of backbone torsional angles (see the figure caption). As expected, the flip rates increase with the temperature. A simple Arrhenius model<sup>69</sup> fitted to the CHF data is drawn as a dashed line. This model yields an energy barrier of  $434k_B$  K for the backbone flips. This value is well in the range of typical barrier heights reported for biomolecules in the literature.<sup>70</sup>

Having estimated the influence of the temperature on the conformational dynamics of our sample peptide 8ALA in SPC water, it seems appropriate to check whether a separate classical BT (as frequently applied in strategy P.1) changes the dynamics. Here, particularly a slowing down seems possible because a rapidly coupled thermostat can interfere with long-lasting energy fluctuations within the peptide, which are caused by random in- and outflow of energy from the solvent. For the purpose of such a check, we carried out the simulation sets CLS.1–2 listed in Table 2, in which a classical BT separately coupled to  $P$  enforced temperatures  $\hat{T}_P$  of about 300 K and 340 K, respectively.



**Figure 5.** Root mean perturbation quotients  $\bar{\xi}_{C_\alpha}$  at the  $C_\alpha$  atoms of 8ALA and ALDI evaluated from simulation series #3 for the NHTs and BTs, respectively, for different coupling times  $\tau_P$  of the peptide thermostats.

Figure 4 compares the flip rates observed when using a classical Berendsen thermostat (open squares) with the data for the CHF thermostat (filled circles) and demonstrates that our expectation is actually met. Thus, if one wants to sample the equilibrium fluctuations of a peptide in solution by MD as rapidly as possible, or if one wants to gain access to the kinetics of nonequilibrium relaxation processes, the separate coupling of a classical BT to a small peptide seems counterproductive.

We interpret the above result by the following physical picture: A rapidly coupled BT likewise dampens fluctuations to higher and lower energies, thus leading to the correct average temperature. However, barrier crossings are enabled by rare accidental accumulations of a critical amount of energy in the respective collective coordinates. Particularly by dampening the higher energy fluctuations of the peptide, a classical BT makes such accumulations and, thus, barrier crossings less likely. Note that we have additionally checked the performance of a NHT in the same setting. We found no reduction of flip rates (data not shown) as could be expected for a thermostat maintaining the canonical energy fluctuations.

**Local Perturbations of the Dynamics.** The flips of backbone dihedral angles are collective movements and, therefore, are not directly related to the perturbation which a thermostat inflicts on the dynamics of individual atoms. To check the latter, we collected from simulation series #3 (cf. Table 3) all those forces acting on the  $C_\alpha$  atoms of 8ALA which are required for the evaluation of the perturbation quotients (4). We carried out this data collection for BTs and NHTs with coupling times  $\tau_P$  covering 4 orders of magnitude. In the case of the smaller ALDI model, we concentrated on the Berendsen approach.

Figure 5 shows the resulting perturbation ratios (4) evaluated using the approximate expression 6. As demonstrated by the squares marking the 8ALA results, the perturbations  $\bar{\xi}_{C_\alpha}$  are small for both thermostats and decrease over a wide range linearly with the inverse of  $\tau_P$ . For the classical BT ( $\tau_P = 0.1$  ps) the  $\bar{\xi}_{C_\alpha}$  are only about 0.5%. Furthermore, the smaller ALDI model exhibits slightly larger  $\bar{\xi}_{C_\alpha}$  (open diamonds) than 8ALA (open squares). However, this size-induced difference is much smaller than that between the NHTs and BTs. At a given  $\tau_P$ , Nosé-Hoover coupling inflicts perturbations which are by 1 order of

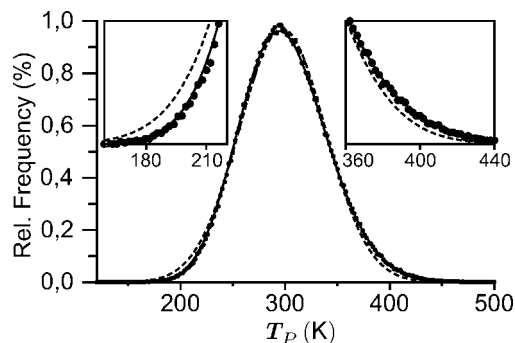
magnitude larger than in the Berendsen case (cf. Figure 5). For a Berendsen coupling of maximal strength ( $\tau_p = 0.001$  ps) the perturbation is comparable to that of a NHT with  $\tau_p$  as large as 0.064 ps. Furthermore, for Nosé-Hoover coupling  $\tau_p$  cannot be chosen larger than about 0.256 ps where  $\xi_{Ca}$  is about 1% and, thus, not particularly small. In the given case of 8ALA, one otherwise observes long-lasting and artificial temperature oscillations, i.e. the so-called Toda daemon<sup>44</sup> (data not shown).

One can compare the perturbations shown in Figure 5 to those which are inflicted by a CHF thermostat as employed in strategy P.3. In simulation G/A/2\_P.3, the peptide 8ALA was kept at 300 K with a perturbation ratio of  $\xi_{Ca} \approx 10^{-4}$ . As can be seen from Figure 5, this ratio corresponds to a Berendsen coupling time larger than 1 ps in the classical thermostat setup. However, a classical BT with  $T_{0,p} = T_S$  and  $\tau_p \geq 1$  ps cannot properly control the temperature because then  $\tau_p$  is in the range of solvent–solute coupling time ( $\tau_{SP} = 1.6$  ps), i.e. is too slow (cf. section 2). On the other hand, a more strongly coupled thermostat with  $\tau_p = 0.1$  ps does the job, but then the perturbation is more than ten times stronger than for a CHF thermostat.

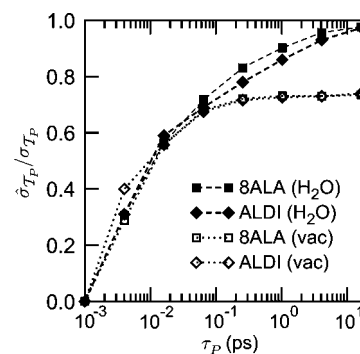
The above analysis was based on data for perturbation ratios derived through the approximate expression 6 and, therefore, depends on the validity of this equation. The first assumption made in the derivation (cf. section 2) of eq 6 was that the atomic velocities  $\dot{\mathbf{r}}_i(t)$  and the thermostat variable  $\gamma(t)$  are uncorrelated. We have checked this assumption for simulation series #3 by evaluating eq 5 with and without taking the correlation into account; the relative difference was less than  $10^{-2}$  for both 8ALA and ALDI (data not shown).

The second assumption was that the individual atomic velocities  $\dot{\mathbf{r}}_i(t)$  are drawn from an undisturbed Maxwell distribution and can be checked by comparing results of the exact expression 5 with results of the approximate expression 6. We evaluated these expressions for the trajectories of series #3 and determined the root-mean-square deviations. In the worst case of a BT at the maximum coupling strength ( $\tau_p = 0.001$  ps), we found root-mean-square deviations amounting to 8.3% of the mean thermostatic force for 8ALA and to 14% for ALDI. In view of the moderate statistics provided by the 250 ps simulations employed in series #3, the estimate 6 is fairly reliable. Thus, eq 6 is adequate if one wants to estimate thermostatic forces.

**Temperature Fluctuations.** In our suggestion of the minimally invasive CHF thermostat characterizing strategy P.3 we were guided by the notion that a properly thermostatted explicit solvent system is a canonical heat bath for an uncontrolled solute. To check this assumption, we compare in Figure 6 the canonical  $\chi^2$ -distribution (eq 7) for the instantaneous peptide temperature  $T_p(t)$  with results from simulation G/A/2\_P.3. For the 103 degrees of freedom of 8ALA, the  $\chi^2$ -distribution (solid line) resembles a Gaussian (dashed line), which is expected for very large systems. Remarkably, the MD results (circles) closely reproduce the slight asymmetry of the  $\chi^2$ -distribution. This agreement strongly indicates that the peptide has sampled the canonical ensemble in the simulation G/A/2\_P.3. We have verified this



**Figure 6.** Distribution of the instantaneous temperature  $T_p(t)$  of 8ALA (in SPC water at 300 K) during the 20 ns MD simulation G/A/2\_P.3 (dots). The dashed line is a Gaussian fit to the data. The canonical distribution (eq 7) is drawn as a solid line.



**Figure 7.** Ratio  $\hat{\sigma}_{T_p}/\sigma_{T_p}$  of measured and canonical temperature fluctuations for various coupling times  $\tau_p$  of a Berendsen solute thermostat. The model peptides are 8ALA (squares) and ALDI (diamonds). Simulations were performed in explicit water (H<sub>2</sub>O, filled symbols) and vacuum (vac, empty symbols) for both peptides. Simulation parameters are given in Table 4.

result for a series of further CHF simulations. It did not change for larger solvent systems and was independent of the coupling time for the solvent thermostat provided that the solvent temperature remained well-tuned (data not shown).

To estimate how a classical BT separately coupled to a peptide (strategy P.1) affects its global statistical properties, we determined the temperature fluctuations of the peptides 8ALA and ALDI, respectively, as measured by the standard deviation  $\hat{\sigma}_{T_p}$  in a fourth series of simulations (for details see Table 4). Figure 7 shows the ratio of  $\hat{\sigma}_{T_p}$  and  $\sigma_{T_p}$ , which is the value theoretically expected for a canonical ensemble and is given by eq 8. For peptides in explicit solvent the figure shows that  $\hat{\sigma}_{T_p}/\sigma_{T_p}$  is always smaller than one and approaches that limit for large  $\tau_p$ . Thus, in the classical setting ( $\tau_p = 10^{-2}$  ps) a BT strongly suppresses the canonical temperature fluctuations. These fluctuations successively become restored with increasing  $\tau_p$ . The full range of canonical fluctuations is reached at coupling times  $\tau_p > 10$  ps, i.e. at values exceeding the solvent-peptide heat coupling time  $\tau_{SP}$  by a factor of 10. As a result, the separate BT is effectively disconnected from the peptide, the solute–solvent heat exchange term  $\beta_{SP}$  dominates the heat balance eq 19, and strategy P.1 reduces to the noninvasive strategy P.2.

Figure 7 not only reveals the general suppression of temperature fluctuations within a peptide by a classical BT but also demonstrates through a comparison with vacuum simulation data that these fluctuations are caused (i) by a fast exchange of kinetic and potential energy within a peptide and (ii) by a slower energy exchange with the solvent.

In vacuum simulations, the exchange of kinetic and potential energy within the peptide is the only cause of temperature fluctuations. As shown by the data, a rapidly coupled Berendsen thermostat ( $\tau_P < 0.1$  ps) suppresses these microcanonical fluctuations in the same way as it suppresses the canonical temperature fluctuations of a solvated peptide. However, at slower coupling times  $\tau_P$  the thermostat is seen to no longer affect the microcanonical fluctuations. The clear saturation of  $\hat{\sigma}_{T_P}/\sigma_{T_P}$  at  $\tau_P > 0.1$  ps demonstrates that the microcanonical fluctuations occur on time scales below 0.1 ps. In contrast, additional fluctuations of a solvated peptide are still suppressed by the thermostat with even slower coupling. Thus, as claimed above, they occur on longer time scales.

In order to retain the correct statistics for the solute, it is important to choose the coupling time  $\tau_P$  for the thermostat longer than the typical time scale of the canonical fluctuations, which, in our case, is in the range of 10 ps, as can be seen from Figure 7. However, this time may even be longer for more weakly coupling solvents or larger solutes.

## 5. Conclusions

Every thermostat changes the dynamics of the controlled system to a larger or lesser extent. Measured on a microscopic scale, these changes are by about 1 order of magnitude smaller for BTs than for NHTs (cf. the data on the perturbation quotients displayed in Figure 5). On the other hand, NHTs, in contrast to BTs, guarantee the canonical ensemble. For instance, as shown by the results on the temperature fluctuations (Figure 7), BTs suppress all those canonical energy fluctuations which are slower than the time scale  $\tau$  at which the BT is coupled to the system.

Whether such changes can modify the specific observables to be extracted from a simulation and to be compared with experimental data is *a priori* unclear in many cases. Even if one suspects that a given thermostat could possibly introduce an artifact into the computation of a certain observable, one may have to spend an enormous computational effort for a statistically clear proof. In fact, to prove a suspected dampening of peptide flip rates by a standard BT, we had to spend about 400 ns of simulation time on each of the data points to get the statistical certainty shown in Figure 4.

Especially if the popular strategy P.1 is applied to a solute–solvent system, the specific drawbacks of the various thermostat algorithms may directly affect the properties of the solute. The P.1 strategy with a BT is expected to cause artifacts of type a), i.e. artifacts resulting from an incorrect thermodynamical ensemble. In fact, as we have shown for a sample peptide, the dampening of the canonical energy fluctuations due to the BT can lead to reduced peptide flip rates. Furthermore, one expects that the combination of P.1 with the NHT will render the solute vulnerable to artifacts of type b), i.e. lacking ergodicity. Using the P.1 strategy with

other thermostats which suffer neither from type a) nor type b) drawbacks (e.g., the Nosé–Hoover chain) still perturbs the dynamics much more strongly than necessary, i.e. such a strategy is prone to introduce artifacts of type c) (dynamics).

Given the need for some sort of temperature control in large scale MD simulations of complex systems, the optimal strategies to avoid artifacts of types a), b), and c) are P.2 or P.3, respectively. Here, the minimally invasive strategy P.3, which employs a constant heat flow to compensate the algorithmic heat production in the solute, has to be applied only if the noninvasive strategy P.2 turns out to be ineffective in a sufficiently extended test simulation. Strategy P.3 reduces the perturbation of the solute's dynamics to a minimum while keeping it nevertheless properly tempered. The precise protocol to set up a P.3 scheme is given in the Appendix.

The preservation of the canonical ensemble within the solute through strategies P.2 and P.3 (despite the use of a standard BT for the solvent which strongly perturbs the temperature fluctuations in this part of the system) is the most important result of this paper and proves our hypothesis that an explicitly simulated solvent of the correct temperature  $T_S$  represents the optimal thermostat for a solute. Admittedly, our quantitative analysis of the applicability of strategies P.2 and P.3 is restricted to relatively small peptides because an extended statistics is required for reliable results. Already for the small peptides with their short temperature autocorrelation times of 15 ps, it takes more than 10 ns to determine the average temperature with an accuracy of 1 K. For larger systems, the temperature autocorrelation times increase and so do the simulation times required for accurate temperature measurements. Too short simulations can easily lead to the false impression that the solute temperature sizably differs from the solvent temperature. To our experience, the non-invasive strategy P.2 can suffice for quite large solvent–solute systems. For instance, reinspecting a simulation<sup>8</sup> of the C-terminal domain of the human prion protein (residues 125–228), which employed a global thermostat coupling (strategy G), we found that the protein temperature deviated by more than 10 K from that of the solvent. Subsequent simulations of a slightly larger fragment (residues 114–228), which employed strategy P.2 but otherwise the same simulation setup, showed no significant temperature difference. In the few cases in which one observes a seemingly intolerable temperature difference between solute and solvent, one can still use the solvent as the heat bath by applying the minimally invasive strategy P.3 to keep the solute well tempered.

It should be noted that our heat flow model and the associated setup protocol for the constant heat flow strategy P.3 are restricted to two subsystems with homogeneous local algorithmic heating rates. For simulations of more complex systems such as protein–DNA assemblies in solution, for which one expects more than two different heating rates, a constant heat flow strategy can be analogously designed. However, it will become increasingly difficult to determine the local heating rates of the various subsystems which have to be compensated.

**Acknowledgment.** This work was supported by the Bavarian joint research project ForPrion (LMU02), by the VolkswagenStiftung (I/79 884), and by Deutsche Forschungsgemeinschaft (SFB533/C1, SFB749/C4).

## Appendix: Setting up Strategies P.2 and P.3

Here, we give a detailed description of the steps needed in order to set up a simulation system containing a macromolecule  $P$  in thermal equilibrium with an explicit solvent environment  $S$  according to the strategies P.2 and P.3, respectively, using the standard Berendsen algorithm. After preparation (e.g., removal of close solvent–solute contacts by energy minimization), the following steps are necessary:

a) Heating phase: The subsystems are heated using two separate classical BTs (e.g.,  $\tau_S = \tau_P = 0.1$  ps) to the temperature  $T_{\text{sim}}$  desired in the production simulation. Depending on the initial deviations of the solute temperature  $T_P$  and solvent temperature  $T_S$ , it may take a simulation time of up to  $30\tau_{SP}$  for the respective subsystems to safely attune to  $T_{\text{sim}}$ .

b) Relaxation phase I: The solute is decoupled from its thermostat ( $\tau_P = \infty$ ) and relaxes to its new steady state temperature  $T_{P,1}$ . The time constant for the relaxation to the steady state is the solvent–solute coupling time  $\tau_{SP}$ . Since  $\tau_{SP}$  is still unknown, an upper limit estimate (e.g.,  $\tau_{SP} \approx 20$  ps) should be used to determine the relaxation time  $t_{\text{relax}} \approx 10\tau_{SP}$ .

c) Test simulation I: Here, the solute remains decoupled from its thermostat and the simulation serves to determine its average temperature  $T_{P,1}$ . If the deviation from equilibrium measured by  $|T_{P,1} - T_{\text{sim}}|$  is less than an acceptable tolerance  $\Delta T_P$ , then the noninvasive strategy P.2 is applicable, and one may directly continue the simulation for data production f).

The necessary simulation time  $t_1$  for the test depends on the tolerable uncertainty  $\sigma_{T_{P,1}}^2$  of the measured solute temperature  $\hat{T}_{P,1}$ , which forms an upper bound for the uncertainty  $\sigma_{T_P}^2$  in the prediction of the production run temperature  $T_P$ . If  $\Delta T_P$  is the accuracy required for the prediction, we should make sure that  $\sigma_{T_{P,1}}^2 \leq \Delta T_P^2$ . By eq 9 the simulation time then is  $t_1 = 2\tau_c \sigma_{T_P}^2 / \Delta T_P^2$ , where  $\tau_c$  is the temperature autocorrelation time of the solute, and  $\sigma_{T_P}$  is the standard deviation of its temperature fluctuations, which were observed during the test run. One typically obtains simulation times of several nanoseconds.

d) Relaxation phase II: The solute is coupled to a separate thermostat with a coupling time  $\tau_P \geq 500$  ps intended for the P.3 production run. Using an estimate for  $\tau_{SP}$  (e.g., 1 ps), a reasonable choice for the target temperature is given by  $T_{0,P,2} = -\tau_P / \tau_{SP} \cdot |T_{P,1} - T_{\text{sim}}|$  (leading to 2-fold overcompensation if  $\tau_{SP}$  was exact). The duration of this relaxation phase is the same as in step b).

e) Test simulation II: The average temperature  $T_{P,2}$  is determined. The simulation time  $t_2$  should be equal to  $t_1$  in step c).

f) Production simulation: If strategy P.2 turned out to be applicable in step c), the settings in this simulation are chosen identically (in fact, one may regard the test run as the initial part of the production simulation). Otherwise, the target

temperature  $T_{0,P}$  for a P.3 simulation is determined from the two test simulations by

$$T_{0,P} = T_{\text{sim}} + \frac{T_{0,P,2} - T_{P,2}}{T_{P,2} - T_{P,1}} (T_{\text{sim}} - T_{P,1}) \quad (22)$$

## References

- (1) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646–652.
- (2) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garca, A. E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 168–174.
- (3) Grubmüller, H.; Heymann, B.; Tavan, P. *Science* **1996**, *271*, 997–999.
- (4) Rief, M.; Oesterhelt, F.; Heymann, B.; Gaub, H. E. *Science* **1997**, *275*, 1295–1297.
- (5) Brockwell, D. J.; Paci, E.; Zinober, R. C.; Beddard, G. S.; Olmsted, P. D.; Smith, D. A.; Perham, R. N.; Radford, S. E. *Nat. Struct. Biol.* **2003**, *10*, 731–737.
- (6) Kucera, K.; Lambry, J. C.; Martin, J. L.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5805–5807.
- (7) Spörlein, S.; Carstens, H.; Satzger, H.; Renner, C.; Behrendt, R.; Moroder, L.; Tavan, P.; Zinth, W.; Wachtveitl, J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7998–8002.
- (8) Hirschberger, T.; Stork, M.; Schropp, B.; Winklhofer, K. F.; Tatzelt, J.; Tavan, P. *Biophys. J.* **2006**, *90*, 3908–3918.
- (9) Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M. V.; Alonso, D. O. V.; Daggett, V.; Fersht, A. R. *Nature* **2003**, *421*, 863–867.
- (10) Shea, J.-E.; Brooks, C. L. I. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499–535.
- (11) Privalov, P. L. Physical basis of the stability of the folded conformations of Proteins. In *Protein Folding*; Creighton, T. E., Ed.; W. H. Freeman: San Francisco, 1992.
- (12) Munishkina, L. A.; Phelan, C.; Uversky, V. N.; Fink, A. L. *Biochemistry* **2003**, *42*, 2720–2730.
- (13) Vitkup, D.; Ringe, D.; Petsko, G. A.; Karplus, M. *Nat. Struct. Biol.* **2000**, *7*, 34–38.
- (14) Tavan, P.; Carstens, H.; Mathias, G. Molecular Dynamics Simulations of Proteins and Peptides: Problems, Achievements, and Perspectives. In *Protein Folding Handbook. Part I*; Buchner, J., Kiefhaber, T., Eds.; Wiley-VCH: 2005.
- (15) Frauenfelder, H.; Petsko, G. A.; Tsernoglou, D. *Nature* **1979**, *280*, 558–563.
- (16) Ishima, R.; Torchia, D. A. *Nat. Struct. Biol.* **2000**, *7*, 740–743.
- (17) Ichiye, T.; Karplus, M. *Biochemistry* **1983**, *22*, 2884–2893.
- (18) Lazaridis, T.; Karplus, M. *Biophys. Chem.* **2003**, *100*, 367–395.
- (19) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (20) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.
- (21) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749–13754.
- (22) Carstens, H. Konformationsdynamik lichtsichtbarer Peptide: Molekulardynamiksimulationen und datengetriebene Modell-

- bildung, Dissertation, Fakultät für Physik, Ludwig-Maximilians-Universität München, 2004.
- (23) Koper, G. J. M.; Reiss, H. *J. Phys. Chem.* **1996**, *100*, 422–432.
- (24) Swope, W. C.; Pitera, J. W.; Suits, F. *J. Phys. Chem. B* **2004**, *108*, 6571–6581.
- (25) Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. J. C.; Zhestkov, Y.; Zhou, R. *J. Phys. Chem. B* **2004**, *108*, 6582–6594.
- (26) Nutt, D. R.; Smith, J. C. *J. Chem. Theory Comput.* **2007**, *3*, 1550–1560.
- (27) Mathias, G.; Egwolf, B.; Nonella, M.; Tavan, P. *J. Chem. Phys.* **2003**, *118*, 10847–10860.
- (28) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (29) Grubmüller, H.; Tavan, P. *J. Comput. Chem.* **1998**, *19*, 1534–1552.
- (30) Kraeutler, V.; van Gunsteren, W. F.; Hünenberger, P. H. *J. Comput. Chem.* **2001**, *22*, 501–508.
- (31) Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155–179.
- (32) Oda, K.; Miyagawa, H.; Kitamura, K. *Mol. Simul.* **1996**, *16*, 167–177.
- (33) Cheng, A.; Merz, K. M. *J. Phys. Chem.* **1996**, *100*, 1927–1937.
- (34) Guenot, J.; Kollman, P. A. *Protein Sci.* **1992**, *1*, 1185–1205.
- (35) Arnold, G. E.; Ornstein, R. L. *Proteins* **1994**, *18*, 19–33.
- (36) Hünenberger, P. H. *Adv. Polym. Sci.* **2005**, *173*, 105–149.
- (37) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (38) Harvey, S. C.; Tan, R. K.-Z.; Cheatham, T. E. *J. Comput. Chem.* **1998**, *19*, 726–740.
- (39) Chui, S.-W.; Clark, M.; Subramaniam, S.; Jakobsson, E. *J. Comput. Chem.* **2000**, *21*, 121–131.
- (40) Morishita, T. *J. Chem. Phys.* **2000**, *113*, 2976–2982.
- (41) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (42) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- (43) Tuckerman, M. E.; Liu, Y.; Ciccotti, G.; Martyna, G. J. *J. Chem. Phys.* **2001**, *115*, 1678–1702.
- (44) Holian, B. L.; Voter, A. F.; Ravelo, R. *Phys. Rev. E* **1995**, *52*, 2338–2347.
- (45) Posch, H. A.; Hoover, W. G.; Vesely, F. J. *Phys. Rev. A* **1986**, *33*, 4253–4265.
- (46) D'Alessandro, M.; Tenenbaum, A.; Amadei, A. *J. Phys. Chem. B* **2002**, *106*, 5050–5057.
- (47) Tobias, D. J.; Martyna, G. J.; Klein, M. L. *J. Phys. Chem.* **1993**, *97*, 12959–12966.
- (48) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635–2643.
- (49) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101(1–7)
- (50) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Oxford University Press: Oxford, 1987.
- (51) Stange, K. *Angewandte Statistik*; Springer: Heidelberg, 1970.
- (52) Weber, W.; Hünenberger, P. H.; McCammon, J. A. *J. Phys. Chem. B* **2000**, *104*, 3668–3675.
- (53) Daura, X.; Gademann, K.; Schäfer, H.; Jaun, B.; Seebach, D.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **2001**, *123*, 2393–2404.
- (54) Villareal, M. A.; Montich, G. G. *J. Biomol. Struct. Dyn.* **2005**, *23*, 135–142.
- (55) van den Bosch, M.; Swart, M.; Snijders, J. G.; Berendsen, H. J.; Mark, A. E.; Oostenbrink, C.; van Gunsteren, W. F.; Canters, G. W. *ChemBioChem* **2005**, *6*, 738–746.
- (56) Monticelli, L.; Simoes, C.; Belvisi, L.; Colombo, G. *J. Phys.: Condens. Matter* **2006**, *18*, S329–S345.
- (57) Fox, T.; Kollman, P. A. *Proteins* **1996**, *25*, 315–334.
- (58) Kong, Y.; Ma, J.; Karplus, M.; Lipscomb, W. N. *J. Mol. Biol.* **2006**, *356*, 237–247.
- (59) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.
- (60) Niedermeier, C.; Tavan, P. *J. Chem. Phys.* **1994**, *101*, 734–748.
- (61) Niedermeier, C.; Tavan, P. *Mol. Simul.* **1996**, *17*, 57–66.
- (62) Mathias, G.; Tavan, P. *J. Chem. Phys.* **2004**, *120*, 4393–4403.
- (63) Eichinger, M.; Grubmüller, H.; Heller, H.; Tavan, P. *J. Comput. Chem.* **1997**, *18*, 1729–1749.
- (64) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.
- (65) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Interaction Models for Water in Relation to Protein Hydration. In Intermolecular Forces*; Pullman, B., Ed.; D. Reidel Publishing Company: Dordrecht, 1981.
- (66) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (67) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (68) Guenot, J.; Kollman, P. A. *J. Comput. Chem.* **1993**, *14*, 295–311.
- (69) Gardiner, C. W. *Handbook of Stochastic Methods*, 2nd ed.; Springer: Berlin, 1985.
- (70) Zuckerman, D. M.; Lyman, E. *J. Chem. Theory Comput.* **2006**, *2*, 1200–1202.