DOI: 10.1002/chem.200701380

Estimating the Temperature Dependence of Peptide Folding Entropies and Free Enthalpies from Total Energies in Molecular Dynamics Simulations

Ricard Boned,[a] Wilfred F. van Gunsteren,[b] and Xavier Daura*[a, c]

Abstract: The temperature dependence of thermodynamic quantities, such as heat capacity, entropy and free enthalpy, may be obtained by using well-known equations that relate these quantities to the enthalpy of the molecular system of interest at a range of temperatures. In turn, the enthalpy of a molecular system can be estimated from molecular dynamics simulations of an appropriate model. To demonstrate this, we have investigated the temperature dependence of the enthalpy, heat capacity, entropy and free enthalpy of a system that consists of a β-heptapeptide in methanol and have used the statistical mechanics relationships to describe the thermodynamics of the folding/unfolding equilibrium of the peptide. The results illustrate the power of current molecular simulation force fields and techniques in establishing the link between thermodynamic quantities and conformational distributions.

Keywords: GROMOS · molecular dynamics · peptide folding · peptides · thermodynamics

Introduction

One of the main goals of molecular simulation is to establish a link between conformational molecular behaviour, which is often evasive to experimental probes, and thermodynamic properties. The use of appropriate molecular models, force fields and molecular simulation techniques has already become standard for the interpretation, and in some cases prediction or validation, of experimental data.[1]

The state of a molecular system is characterised by thermodynamic quantities, such as, for example, in the NPT ensemble, volume, enthalpy, heat capacity, entropy and, in particular, free enthalpy. The temperature dependence of these quantities may be investigated, as demonstrated herein, through the analysis of molecular dynamics (MD) simulation trajectories at a range of temperatures by using simple thermodynamic equations that relate the enthalpy of a system with its heat capacity, entropy and free enthalpy. This type of analysis can be used, for example, to relate calorimetric data to atomic-resolution molecular motion.

The thermodynamic analysis proposed is illustrated by studying the properties of a β-heptapeptide in methanol (Figure 1).[2,3] β-Peptides are peptides composed of β-amino acids, that is, $H_2N\text{-}C_\beta R\text{-}C_\alpha R'\text{-}COOH$. Small β-peptides of as few as six amino acid residues fold into turns, helices, and sheetlike structures that are analogous to the secondary structures of proteins. In addition, these compounds are resistant to degradation by most common peptidases and proteases. These two properties have made them attractive targets for pharmaceutical developments (see reference [4] for a comprehensive review of β-peptides and their applications). The β-heptapeptide of interest primarily populates an (M)-3_{14}-helical conformation in methanol at 298 K and 1 atm.[2] We performed seven 0.4 μs MD simulations of a model system that consisted of the β-heptapeptide and solvent (methanol) in a periodic cuboidal box under constant pressure (1 atm) and temperature (298, 310, 320, 330, 340, 350 and 360 K, respectively). In these simulations, the peptide undergoes multiple transitions between two generic conformational states, folded ((M)-3_{14}-helix) and unfolded (heterogeneous conformational state grouping all other conformations), in an equilibrium that is temperature dependent. This enables the calculation of the temperature de-

[a] R. Boned, Prof. X. Daura
Institute of Biotechnology and Biomedicine (IBB)
Universitat Autònoma de Barcelona (UAB)
08193 Bellaterra (Spain)
Fax: (+34) 93-581-2011
E-mail: Xavier.Daura@uab.cat

[b] Prof. W. F. van Gunsteren
Laboratory of Physical Chemistry
Swiss Federal Institute of Technology (ETH)
8093 Zürich (Switzerland)

[c] Prof. X. Daura
Catalan Institution for Research and Advanced Studies (ICREA)
08010 Barcelona (Spain)

Supporting information for this article is available on the WWW under http://www.chemeurj.org/ or from the author.

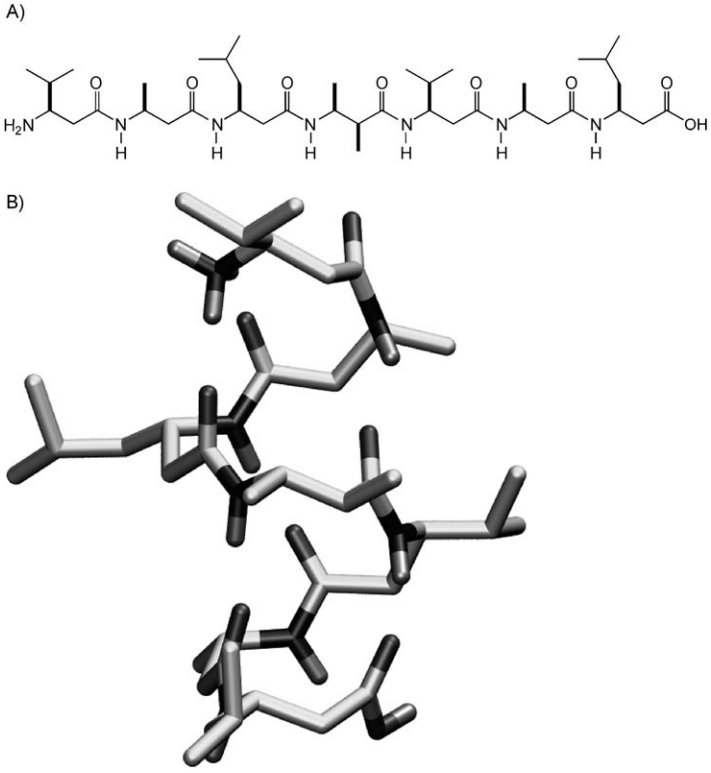

Figure 1. A) Structural formula of the β-heptapeptide studied. In the simulations both end groups were protonated, in accordance with previous studies and experimental data.[2,3,10] B) The NMR spectroscopy model structure in methanol at 298 K and 1 atm.[2]

pendence of changes in volume, enthalpy, heat capacity, entropy, and free enthalpy associated with the folding process.

Theoretical background

G/T *differences from temperature integration*: The Gibbs–Helmholtz equation states that at a constant number of particles (N) and pressure (p), if the enthalpy (H) of the system is known, then the temperature dependence of the free enthalpy (G) divided by the temperature (T) is also known [Eq. (1)].[5] Equation (1) can also be rewritten as Equation (2) and the enthalpy is given by Equation (3), in which U is the internal energy and V is the volume:

$$\left(\frac{\partial(G/T)}{\partial T}\right)_p = -\frac{H}{T^2} \tag{1}$$

$$\left(\frac{\partial(G/T)}{\partial(1/T)}\right)_p = H \tag{2}$$

$$H = U+pV \tag{3}$$

When using MD simulations at constant pressure, the enthalpy of the system may be expressed as an ensemble or time average, as shown in Equation (4):

$$H = \langle\mathscr{H}\rangle_{NpT}+p\langle V\rangle_{NpT} = \langle E_V+E_K\rangle_{NpT}+p\langle V\rangle_{NpT} \tag{4}$$

in which $\mathscr{H}$ is the Hamiltonian of the system, E_V is the potential energy and E_K is the kinetic energy. The subindex on the angular brackets indicates the statistical ensemble at which the configurations of the system are generated. Integrating Equation (2) between $1/T_1$ and $1/T_2$ gives Equation (5):

$$\frac{G^{T_2}}{T_2}-\frac{G^{T_1}}{T_1} = \int_{1/T_1}^{1/T_2} H\mathrm{d}(1/T) \tag{5}$$

The enthalpy (H) can be computed from MD simulations at a number of temperatures between T_1 and T_2 and the integral can be then solved numerically.

The Gibbs–Helmholtz equation can be also applied to changes in the state of the system. When the system evolves from state A to state B, the associated change in free enthalpy is defined by Equation (6). The corresponding change in enthalpy is given by Equation (7) and the Gibbs–Helmholtz equation can be then defined by Equation (8):

$$\Delta G_{\mathrm{BA}} = G_{\mathrm{B}}-G_{\mathrm{A}} \tag{6}$$

$$\Delta H_{\mathrm{BA}} = H_{\mathrm{B}}-H_{\mathrm{A}} \tag{7}$$

$$\left(\frac{\partial(\Delta G_{\mathrm{BA}}/T)}{\partial(1/T)}\right)_p = \Delta H_{\mathrm{BA}} \tag{8}$$

Integrating Equation (8) between $1/T_1$ and $1/T_2$ gives Equation (9):

$$\frac{\Delta G_{\mathrm{BA}}^{T_2}}{T_2}-\frac{\Delta G_{\mathrm{BA}}^{T_1}}{T_1} = \int_{1/T_1}^{1/T_2} \Delta H_{\mathrm{BA}}\mathrm{d}(1/T) \tag{9}$$

Equations (5) and (9) can also be derived from the statistical mechanical definition of free enthalpy.[6]

G/T *differences from probabilities*: If A and B are two states in equilibrium, then the relationship between ΔG_{BA} and the equilibrium constant K_{BA} is given by Equation (10) in which R is the gas constant and the equilibrium constant is defined by Equation (11).

$$\frac{\Delta G_{\mathrm{BA}}}{T} = -R\ln K_{\mathrm{BA}} \tag{10}$$

$$K_{\mathrm{BA}} = \frac{P_{\mathrm{B}}}{P_{\mathrm{A}}} \tag{11}$$

P_{A} and P_{B} are the probabilities of finding the system in states A and B, respectively. The probabilities can be estimated from the time spent in each state in equilibrium MD simulations that feature sufficient transitions between the two states. From Equation (10) it follows that Equation (12) can be derived, which can be numerically compared with Equation (9).

$$\frac{\Delta G_{BA}^{T_2}}{T_2}-\frac{\Delta G_{BA}^{T_1}}{T_1}=-R\ln\frac{K_{BA}^{T_2}}{K_{BA}^{T_1}} \quad (12)$$

S *differences from temperature integration*: The entropy (S) of a system at temperature T_2 can be calculated from knowledge of its entropy at temperature T_1 and the heat required to change the temperature from T_1 to T_2 at constant pressure,[5] as shown in Equation (13) in which C_p is the heat capacity at constant pressure [Eq. (14)].

$$S^{T_2}-S^{T_1}=\int_{T_1}^{T_2}\frac{C_p}{T}\mathrm{d}T \quad (13)$$

$$C_p=\left(\frac{\partial H}{\partial T}\right)_p \quad (14)$$

If the heat capacity of the system is constant in the intervals $(T_a,T_b)\in[T_1,T_2]$ and $(T_b,T_c)\in[T_1,T_2]$, then we can derive Equations (15) and (16).

$$C_p^{(T_a,T_b)}\approx\left(\frac{H^{T_b}-H^{T_a}}{T_b-T_a}\right)_p \quad (15)$$

$$C_p^{T_b}\approx\frac{C_p^{(T_a,T_b)}+C_p^{(T_b,T_c)}}{2} \quad (16)$$

By using MD simulations, the heat capacity can be computed with Equation (16) at a number of temperatures between T_1 and T_2, and the integral in Equation (13) can then be solved numerically.

When using Equations (4), (5) and (13), numerical errors due to numerical integration or to the approximation used for the heat capacity, can be evaluated with the relationship derived from the classical definition of free enthalpy, shown in Equation (17):

$$\frac{G^{T_2}}{T_2}-\frac{G^{T_1}}{T_1}=\frac{H^{T_2}}{T_2}-\frac{H^{T_1}}{T_1}-(S^{T_2}-S^{T_1}) \quad (17)$$

Equation (13) can be also used to analyse the effect of a temperature change on the relative entropies of states A and B of a system. The entropy difference between states A and B is defined by Equation (18) and the corresponding heat capacity difference is given by Equation (19).

$$\Delta S_{BA}=S_B-S_A \quad (18)$$

$$\Delta C_{p_{BA}}=C_{p_B}-C_{p_A} \quad (19)$$

The change in the relative entropies of states A and B due to changing the temperature from T_1 to T_2 is then given by Equation (20):

$$\Delta S_{BA}^{T_2}-\Delta S_{BA}^{T_1}=\int_{T_1}^{T_2}\frac{\Delta C_{p_{BA}}}{T}\mathrm{d}T \quad (20)$$

Results and Discussion

System thermodynamics: Basic thermodynamic quantities are given in Table 1. Calculating error bars on these quantities in a realistic way is difficult. Statistical errors on the data set are small and do not take into account the most significant source of error, that is, incomplete sampling of the accessible configurational space or sampling with non-converged weights. In addition to error estimates, we show the evolution of the enthalpy along the last 0.3 μs of simulation (Figure 2). Clearly, in the last 0.2 μs the enthalpy of the system varies by less than 2 kJ mol^{-1} at all temperatures. Table 1 shows that the enthalpy of the system increases with increasing temperature in an almost linear manner (very slightly quadratic, see Figure S3 in the Supporting Information) for the given temperature range.

Table 1. Absolute thermodynamic quantities.[a]

$\langle T\rangle$ [K]	$\langle p\rangle$ [kJ mol^{-1} nm^{-3}]	$\langle V\rangle$ [nm^3]	$H\pm\delta H$ [kJ mol^{-1}]
297.8	0.058	62.0	−29228 ± 0.6
309.6	0.056	63.0	−28201 ± 0.8
319.6	0.060	63.9	−27327 ± 1.0
329.5	0.058	64.8	−26438 ± 0.8
339.5	0.058	65.7	−25534 ± 0.9
349.5	0.061	66.8	−24617 ± 0.9
359.5	0.062	67.9	−23685 ± 0.8

[a] The averages are given over 4×10^6 time points (1 per 0.1 ps). The enthalpy has been calculated by using Equation (4). Because $\langle p\rangle$ should be the same for all simulations, the average over the seven values (0.059 kJ mol^{-1} nm^{-3}) has been used to calculate the enthalpy. Note that 1 atm = 0.061 kJ mol^{-1} nm^{-3}. The calculation of the errors is described in the Experimental Section.

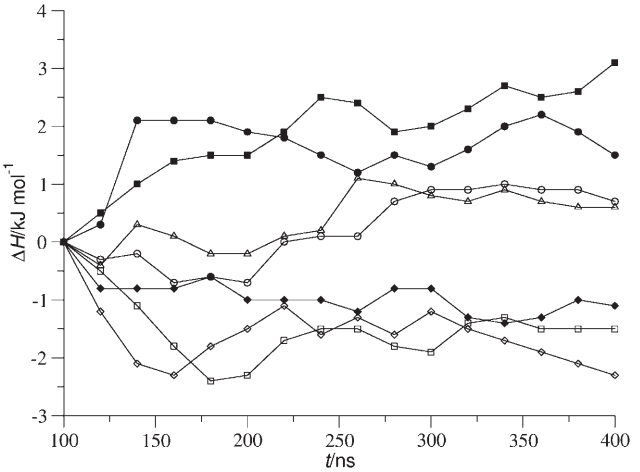

Figure 2. Variation of the enthalpy of the system in the last 0.3 μs of the simulations at 298 (○), 310 (□), 320 (◇), 330 (△), 340 (●), 350 (■) and 360 K (◆). $\Delta H(t')=H(t')-H(t=0.1\ \mu s)$, in which $H(t')$ is calculated by using Equation (4) and the data from the time interval 0 to t'.

By using Equation (5) (integration of the curve in Figure S3 in the Supporting Information from right to left) the relationship between the free enthalpy of the system and the temperature can be investigated, for example, $(G^{360}/360\,\mathrm{K})-(G^{298}/298\,\mathrm{K})=15.4\,\mathrm{kJ\,mol^{-1}\,K^{-1}}$ and $(G^{350}/350\,\mathrm{K})-(G^{310}/310\,\mathrm{K})=9.8\,\mathrm{kJ\,mol^{-1}\,K^{-1}}$. As expected, the free enthalpy of the system per kelvin increases with increasing temperature. Note that the values of these differences do not change when calculated over only the first 0.2 μs.

By using Equation (16), the heat capacity of the system at 310, 320, 330, 340 and 350 K can be estimated. The heat capacity per kelvin is plotted in Figure 3 as a function of temperature. The integration of this curve gives an entropy difference of $S^{350}-S^{310}=10.9\,\mathrm{kJ\,mol^{-1}\,K^{-1}}$. Thus, as common sense would suggest, the entropy of the system increases with increasing temperature. Again the value of this difference does not change when calculated over only the first 0.2 μs.

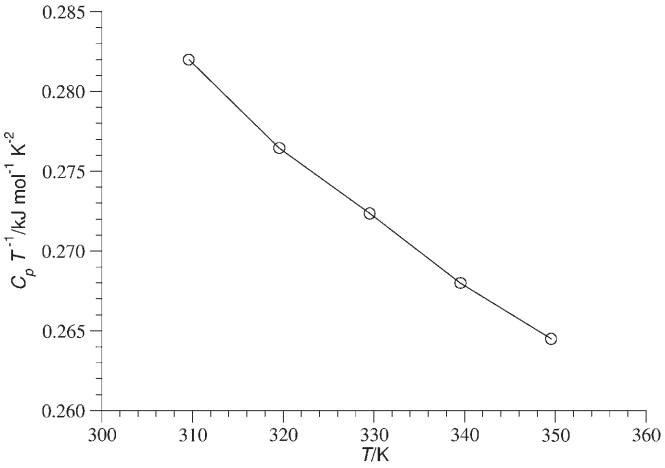

Figure 3. Heat capacity of the system at constant pressure per kelvin as a function of temperature. Error bars are smaller than the circles.

Do the differences obtained for free enthalpy, enthalpy and entropy conform to Equation (17) in the temperature interval 310–350 K? The answer to this numerical check is yes (9.8 = 20.7–10.9), which indicates that the approximation used for the calculation of the heat capacity of the system is valid in this range.

Folding thermodynamics: As mentioned in the Theoretical background section, the same type of formulae can be used to study peptide-folding thermodynamics. The atom-positional root-mean-square differences (RMSD) between peptide structures taken at 0.5 ps intervals from the simulation and the NMR spectroscopy model structure are shown in Figure 4. Clearly, multiple events of unfolding and refolding exist at each temperature (see the Experimental Section for the criterion used to distinguish between the folded (F) and unfolded (U) peptide structures). This is an essential condition for the present analysis. After all (saved) configurations of the system were classified as F (the peptide in the (*M*)-3_{14}-helical conformation) or U, the differences in Table 2 could be calculated. As an indication of numerical precision (i.e., not accuracy), in addition to estimated errors given in Table 2, the evolutions in the last 0.3 μs of the simulation of $\Delta\langle V\rangle_{FU}$ and ΔH_{FU} are shown in Figure 5. Intriguingly, the volume of the system seems to be very slightly, but also systematically, larger for the folded state than for the unfolded state (Table 2). Figure 5A shows that the sign of this difference is indeed significant (fluctuations in the last 0.2 μs are lower than 0.005 nm^3). This suggests that the solvent-excluded volume is very slightly larger in the folded state, which corresponds to a more compact average structure of the peptide. A clear example comes from the comparison of an extended and a helical confomer. Whereas in an extended conformation the peptide is fully accessible to the solvent, the helical conformation has a cylindrical solvent-inaccessible core. A similar observation has been made for α-peptides adopting an α-helical fold in aqueous solution.[7] A remarkable conclusion from Table 2 is that, within this temperature range, enthalpy is more favourable for folding at the higher temperatures. This property, which might appear surprising at first sight, has also been observed experimen-

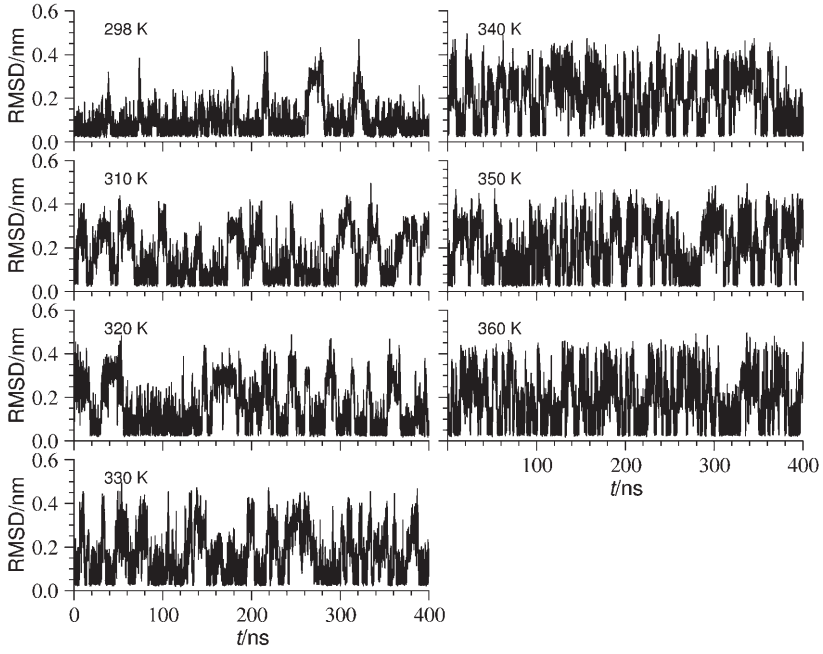

Figure 4. Atom-positional RMSD between peptide structures taken at 0.5 ps intervals from the simulation and the helical NMR spectroscopy model structure. Translational and rotational least-squares fitting and RMSD calculation were based on all peptide groups (O-C-N-H atoms) and the C_β and C_α atoms of residues 2 to 5.

Table 2. Folding thermodynamic quantities.[a]

$\langle T\rangle$ [K]	$\langle V\rangle_F$ [nm^3]	$\Delta\langle V\rangle_{FU}$ [nm^3]	H_F [kJ mol^{-1}]	$\Delta H_{FU}\pm\delta(\Delta H_{FU})$ [kJ mol^{-1}]	$P_F\pm\delta P_F$
297.8	61.98	0.02	−29230.0	−8.1±1.1	0.79±0.07
309.6	62.98	0.02	−28205.4	−9.3±1.1	0.48±0.16
319.6	63.86	0.03	−27333.6	−13.0±1.3	0.52±0.16
329.5	64.79	0.03	−26444.3	−11.7±1.1	0.49±0.12
339.5	65.76	0.03	−25542.5	−13.6±1.2	0.35±0.14
349.5	66.79	0.03	−24627.7	−15.9±1.2	0.35±0.15
359.5	67.89	0.03	−23693.7	−13.9±1.2	0.39±0.11

[a] Volumes (V_X), enthalpies (H_X) and probabilities (P_X) have been calculated over a set of 8×10^5 configurations of the system (1 per 0.5 ps), previously classified as unfolded (U subscript) or folded (F subscript). Enthalpies have been calculated by using Equation (4), with an average pressure of 0.059 kJ mol^{-1} nm^{-3} (see Table 1). Probabilities have been estimated by conformation counting. The calculation of errors is described in the Experimental Section. Only the upper errors (sum of errors) are shown for the folding enthalpies.

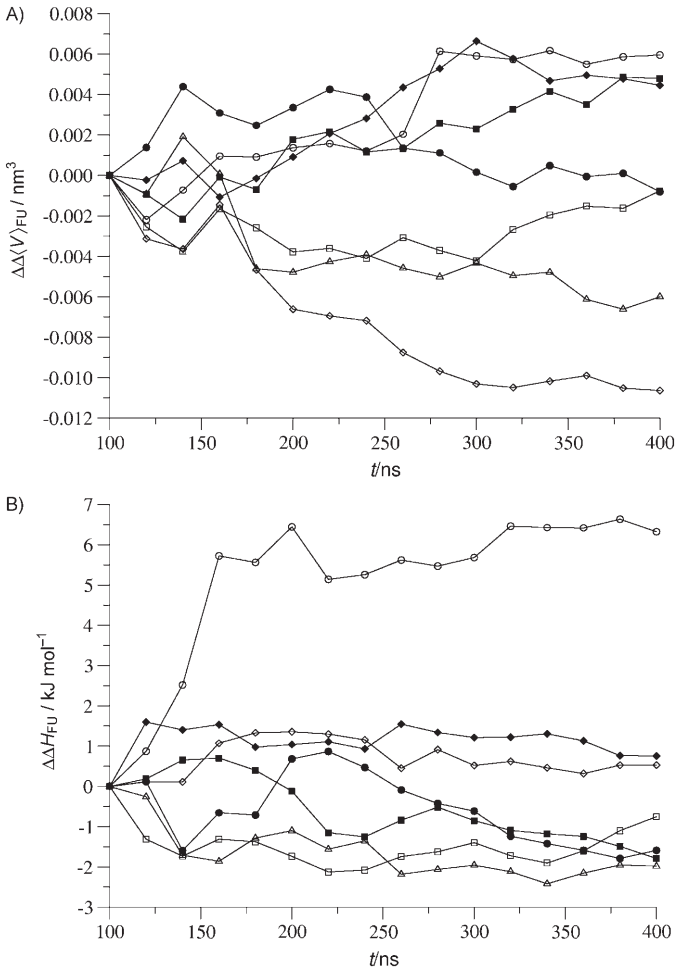

Figure 5. Variation of A) the folding volume ($\Delta\langle V\rangle_{FU}=\langle V\rangle_F-\langle V\rangle_U$) and B) the folding enthalpy ($\Delta H_{FU}=H_F-H_U$) of the system in the last 0.3 μs of the simulations at 298 (○), 310 (□), 320 (◇), 330 (△), 340 (●), 350 (■) and 360 K (◆). $\Delta\Delta\langle V\rangle_{FU}(t')=\Delta\langle V\rangle_{FU}(t')-\Delta\langle V\rangle_{FU}(t=0.1\ \mu s)$, $\Delta\Delta H_{FU}(t')=\Delta H_{FU}(t')-\Delta H_{FU}(t=0.1\ \mu s)$, in which $\Delta\langle V\rangle_{FU}(t')$ and $\Delta H_{FU}(t')$ are calculated by using the data for the time interval 0 to t'.

tally in relation to protein folding in aqueous solution.[8] It means that the average gain in interaction energy upon folding (the kinetic energy must be on average the same for the folded and unfolded states) increases as the temperature increases. This implies that the loss of interaction energy with increasing temperature is faster for the unfolded state than for the folded state. The progression is, however, markedly non-monotonic (see Figure S4 in the Supporting Information). This may be attributed to lack of convergence of population weights in 0.4 μs, but the possibility of non-monotonically decreasing converged probabilities at particular temperature intervals cannot be completely ruled out. Such behaviour has been observed for other systems,[9] and could be due to the inability of the force field to reproduce system properties at temperatures far from the parameterisation temperature, or alternatively, could have a physical explanation. Note that, in general, in the last 0.2 μs of the simulation, the folding enthalpy of the system varies by a maximum of 1.5 kJ mol^{-1} (2 kJ mol^{-1} at 340 K, Figure 5B). As expected, at 298 K the exploration of the unfolded state occurs at a much lower rate due to the dominance of the folded conformation, thus making the convergence of this quantity slow. In terms of folding enthalpies per kelvin, $(\Delta H_{FU}^{350}/350\ K)-(\Delta H_{FU}^{310}/310\ K)=-15\times10^{-3}$ kJ mol^{-1} K^{-1}. Although this number is small, its sign may be considered significant because the value is given per kelvin and the enthalpy, entropy and free enthalpy changes upon folding are small for this system (see below).

As expected, the probability of the folded state (P_F, Table 2) decreases with increasing temperature, although again with a non-monotonic progression likely to be a result of a lack of convergence. Figure 6 shows ΔG_{FU}, which was derived from P_F by using Equation (10), as a function of temperature and its evolution in the last 0.3 μs of the simulation. The two panels indicate that differences between the calculated ΔG_{FU} values are not significant within the ranges 310 to 330 and 340 to 360 K. The slow increase of the folding free enthalpy, or slow decrease of the population of folded structures, with increasing temperature agrees with experimental data of a qualitative nature (CD spectra and chemical shifts of the amino groups) over the same temperature range.[10]

By using Equation (9) (integration of the curve in Figure S4 in the Supporting Information from right to left) we obtain values of $(\Delta G_{FU}^{360}/360\ K)-(\Delta G_{FU}^{298}/298\ K)=7\times10^{-3}$ kJ mol^{-1} K^{-1} and $(\Delta G_{FU}^{350}/350\ K)-(\Delta G_{FU}^{310}/310\ K)=5\times10^{-3}$ kJ mol^{-1} K^{-1} (7×10^{-3} and 4×10^{-3} kJ mol^{-1} K^{-1}, respectively, when calculated over only the first 0.2 μs). Thus, the free enthalpy of folding per kelvin increases with increasing temperature. A difference of 5×10^{-3} kJ mol^{-1} K^{-1} between 310 and 350 K means that the entropy must compensate the counter affect of the enthalpy (-15×10^{-3} kJ mol^{-1} K^{-1}, see above) with -20×10^{-3} kJ mol^{-1} K^{-1}. By using Equation (12), the free enthalpy differences become $(\Delta G_{FU}^{360}/360\ K)-(\Delta G_{FU}^{298}/298\ K)=15\times10^{-3}$ kJ mol^{-1} K^{-1} and $(\Delta G_{FU}^{350}/350\ K)-(\Delta G_{FU}^{310}/310\ K)=4\times10^{-3}$ kJ mol^{-1} K^{-1} (18×10^{-3} and 3×10^{-3} kJ mol^{-1} K^{-1}, respectively, when calculated over only the first 0.2 μs). The relative mismatch between the two sets

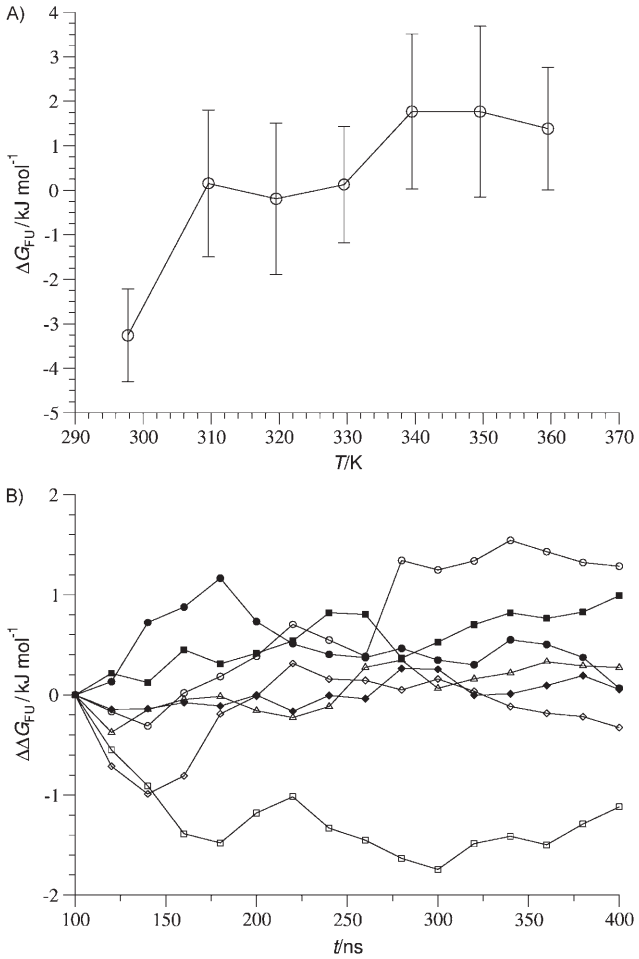

Figure 6. A) Folding free enthalpy ($\Delta G_{FU}=G_F-G_U$) as a function of temperature and B) its variation in the last 0.3 μs of the simulations at 298 (○), 310 (□), 320 (◇), 330 (△), 340 (●), 350 (■) and 360 K (◆). ΔG_{FU} calculated from P_F and P_U ($P_U=1-P_F$, Table 2) by using Equation (10). $\Delta\Delta G_{FU}(t')=\Delta G_{FU}(t')-\Delta G_{FU}(t=0.1\ \mu s)$, in which $\Delta G_{FU}(t')$ is calculated by using the data for the time interval 0 to t'. Error bars are given according to Equation (28) and δP_F from Table 2.

of values obtained from Equations (9) and (12) reflects the differences in the convergence properties of population probabilities and associated enthalpies,[11] and also the lack of an energy component in the criterion used to define the similarity cutoff for clustering.

The change in the heat capacity of the system upon folding (per kelvin) is plotted in Figure 7 as a function of temperature. The convergence problems already seen for the enthalpy become obvious here (Figure S4 in the Supporting Information). The integration of this curve gives the entropy difference of $\Delta S^{350}_{FU}-\Delta S^{310}_{FU}=-15\times10^{-3}\ kJ\,mol^{-1}\,K^{-1}$ ($-9\times10^{-3}\ kJ\,mol^{-1}\,K^{-1}$ when calculated over only the first 0.2 μs), which is close to the expected value of $-20\times10^{-3}\ kJ\,mol^{-1}\,K^{-1}$. This result indicates that, contrary to the situation observed for the enthalpy, entropy is more favourable to folding at lower temperatures. In other words, the entropy loss upon folding is larger the higher the temperature (the entropy of the unfolded state grows faster with temperature than that of the folded state).

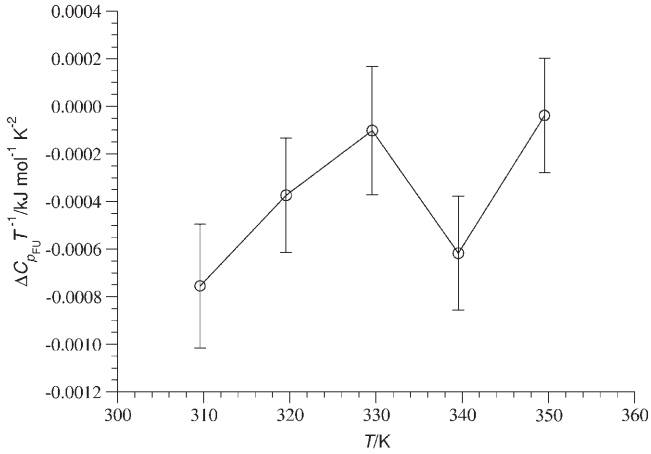

Figure 7. Folding heat capacity of the system at constant pressure ($\Delta C_{p_{FU}}=C_{p_F}-C_{p_U}$) per kelvin as a function of temperature. Error bars are given according to Equation (30) and the upper $\delta(\Delta H_{FU})$ values from Table 2.

Finally, we note that differences between thermodynamic quantities in the folded and unfolded states as defined herein cannot be directly compared with those measured experimentally. This is because herein the distinction between the F and U states is based on a conformational criterion (RMSD from helix) applied to the simulation to partition the ensemble into F and U conformers at one thermodynamic state point, whereas in experiments folding is induced by changing the thermodynamic state point (temperature, co-solvent) and measuring average properties over the whole conformational ensemble.

Contribution of interaction terms to folding: The contribution of different interaction (potential) energy terms to folding is given in Table 3. Internal (peptide) bonding interactions, internal non-bonding interactions and solvent–solvent interactions favour folding at all temperatures, whereas peptide–solvent non-bonding interactions strongly disfavour folding. Although this is not unexpected, it is remarkable that the contribution of solvent–solvent interactions to folding is similar in magnitude to the contribution of the internal non-bonding interactions of the peptide. Due to exact

Table 3. Contribution of interaction terms to folding.[a]

$\langle T\rangle$ [K]	$\Delta\langle E_{v,bp}\rangle_{FU}$ [kJ mol^{-1}]	$\Delta\langle E_{v,nbpp}\rangle_{FU}$ [kJ mol^{-1}]	$\Delta\langle E_{v,nbps}\rangle_{FU}$ [kJ mol^{-1}]	$\Delta\langle E_{v,nbss}\rangle_{FU}$ [kJ mol^{-1}]
297.8	−6.1	−58.4	97.8	−41.4
309.6	−7.4	−53.7	89.8	−37.6
319.6	−8.1	−68.4	113.3	−50.0
329.5	−7.4	−75.7	123.2	−51.6
339.5	−8.5	−73.8	117.4	−49.0
349.5	−8.8	−73.2	115.4	−49.0
359.5	−9.3	−77.2	121.3	−48.6

[a] Average interaction potential energies (E_v) have been calculated over a set of 8×10^5 configurations of the system (1 per 0.5 ps), previously classified as U or F. $E_{v,bp}$ = internal (peptide) bonding interactions, $E_{v,nbpp}$ = non-bonding internal interactions, $E_{v,nbps}$ = non-bonding peptide–solvent interactions and $E_{v,nbss}$ = non-bonding solvent–solvent interactions.

statistical mechanical solvent–solvent enthalpy–entropy compensation, that is, $\Delta H_{ss} = T\Delta S_{ss}$,[12] the enthalpy change due to solvent rearrangement upon folding is not a driving force for this process.[13]

Conclusion

This study illustrates the combined use of MD simulation trajectories and basic thermodynamic formulae for the investigation of the temperature dependence of fundamental properties of molecular systems in equilibrium. More specifically, we have investigated the temperature dependence of the enthalpy, heat capacity, entropy and free enthalpy of a system that consists of a β-heptapeptide in methanol and have used analogous relationships to describe the folding/unfolding equilibrium of the peptide. The results lead us to the following conclusions:

1) The enthalpy, entropy and free enthalpy per kelvin of the system increase with increasing temperature, which indicates a progressive loss of interaction energy and order.
2) The volume of the system in the folded state is very slightly, but systematically, larger than that in the unfolded state, which suggests a slightly more compact (larger solvent-excluded volume) average peptide structure in the folded state.
3) The enthalpy of the system is more favourable to folding at the higher temperatures, which indicates that the loss of interaction energy in the system with increasing temperature is faster for the unfolded state than for the folded state.
4) The entropy of the system is more favourable to folding at lower temperatures, which indicates that the loss of order in the system with increasing temperature is faster for the unfolded state than for the folded state.
5) The folding entropy dominates the temperature dependence of the folding free enthalpy. Thus, the change in free enthalpy per kelvin upon folding increases with increasing temperature.
6) Internal (peptide) bonding interactions, internal non-bonding interactions and solvent–solvent interactions contribute favourably to the enthalpy of folding at all temperatures, whereas peptide–solvent non-bonding interactions strongly disfavour folding.
7) The contribution of solvent–solvent interactions to the enthalpy of folding is similar in magnitude to the contribution of the internal non-bonding interactions of the peptide.

Although none of these conclusions are completely unexpected, they illustrate the power of current molecular simulation force fields and techniques in establishing the link between thermodynamic quantities and conformational states and the energy–entropy compensation effects that dominate polypeptide folding processes.

Experimental Section

Simulation setup: All simulations were performed by using the GROMOS96 simulation software[14] and the GROMOS 43A1 force field (including the methanol model).[14,15] β-Amino acid topologies were generated based on their α-amino acid equivalents.[3,16] A simulation at 298 K was started for a cuboidal periodic system that contained the β-heptapeptide in its (*M*)-3_{14}-helical conformation (Figure 1) and 962 methanol molecules at liquid density after energy minimisation. The setup of this simulation has been described in detail elsewhere (simulation MDT1 in reference [16]). The simulation was performed by using periodic boundary conditions, constant temperature and pressure (1 atm) and a twin-range cutoff of 0.8/1.4 nm for non-bonded interactions. Simulations at 310, 320, 330, 340, 350 and 360 K were branched from the simulation at 298 K after 0.2 ns and the bath temperature was reset to the target temperature in one single step. Each of the seven simulations spanned 0.4 μs.

Trajectory analysis: Trajectory coordinates and energies were stored every 0.5 and 0.1 ps, respectively. Translational and rotational least-squares fitting of structures and atom-positional RMSD were based on all peptide groups (O-C-N-H atoms) and the C_β and C_α atoms of residues 2 to 5. The distinction between F and U structures was based on an atom-positional RMSD criterion: an RMSD matrix (containing the RMSD between every pair of structures) was calculated for 8×10^4 peptide structures taken at regular intervals (5 ps) from the simulation at 298 K. The distribution of RMSD values was plotted (Figure S1 in the Supporting Information) and the first minimum in the distribution (0.1 nm) was taken as similarity cutoff. This cutoff was then used to perform a conformational clustering of all 8×10^4 structures (see reference [17] for a description of the clustering algorithm). The central member of the most-populated cluster at 298 K ((*M*)-3_{14}-helical, Figure S2 in the Supporting Information) was taken as the reference folded structure (for all temperatures). Subsequently, 8×10^5 system configurations (1 per 0.5 ps) from each temperature were classified as F or U, depending on the atom-positional RMSD between the structure of the peptide in that configuration and the reference folded structure ($\leq$0.1 nm or $>$0.1 nm, respectively).

Error analysis: The errors in absolute enthalpy values (δH, Table 1) were calculated by using a block averaging method, with the correlation time estimated from the statistical inefficiency (s) given by Equation (21),[18] in which σ indicates the standard deviation of a distribution, τ_{all} is the number of time steps or system configurations used for the analysis and from Equation (4) $\langle E\rangle_{all}$ is given by Equation (22).

$$\delta H = \sigma(\langle E\rangle_{all}) = \left(\frac{s}{\tau_{all}}\right)^{1/2}\sigma(E) \tag{21}$$

$$\langle E\rangle_{all} = \langle \mathscr{H}\rangle + p\langle V\rangle \tag{22}$$

If the simulation trajectory is divided into n_b blocks, each of which contains τ_b time steps, the statistical inefficiency is defined by Equation (23), in which $\sigma^2(\langle E\rangle_b)$ is given by Equation (24):

$$s = \lim_{\tau_b\to\infty}\frac{\tau_b\sigma^2(\langle E\rangle_b)}{\sigma^2(E)} \tag{23}$$

$$\sigma^2(\langle E\rangle_b) = \frac{1}{n_b}\sum_{b=1}^{n_b}(\langle E\rangle_b - \langle E\rangle_{all})^2 \tag{24}$$

The statistical inefficiency is an estimate of the size at which the data blocks become statistically uncorrelated (assuming Gaussian statistics). Thus, for $\tau_b = s$ Equation (25) can be derived from Equations (21) and (23).

$$\sigma(\langle E\rangle_{all}) = \left(\frac{1}{n_b}\right)^{1/2}\sigma(\langle E\rangle_b) \tag{25}$$

For the calculation of errors in folding enthalpy values ($\delta(\Delta H_{FU})$, Table 2) the following procedure was used: At each temperature, the s

value obtained in the calculation of the error in H (Table 1) was used to define the data block size ($\tau_b = s$, corresponding to a time length t_b). The E values (i.e., $\mathscr{H}+pV$) within each data block were distributed into F and U data blocks according to the peptide conformation. This generated $n_b{}^F$ ($n_b{}^F \leq n_b$, $\tau_b{}^F \leq \tau_b$ and variable through blocks, $t_b{}^F = t_b$) and $n_b{}^U$ ($n_b{}^U \leq n_b$, $\tau_b{}^U \leq \tau_b$ and variable through blocks, $t_b{}^U = t_b$) blocks. These two sets of data blocks were then used to estimate the errors in H_F and H_U by using Equation (25). Finally, a lower bound [Eq. (30), i.e., independent variables] and an upper bound (sum of errors, that is, dependent variables) were calculated for the error in $\Delta H_{FU} = H_F - H_U$.

The errors in probability values (δP_F, Table 2) were derived by block averaging the number of folded configurations in each simulation. Thus, the standard deviation (σ) of the number of folded configurations was calculated for blocks of 5, 10, 20, 25, 40, 50, 80 and 100 ns, and extrapolated to 400 ns from the (initial) linear part of the curve that describes the dependence of σ on block size.

Propagation of errors was treated within the following general approach. If $Z = f(A)$, the relationship between δZ and δA is given by Equation (26):

$$\delta Z = \frac{dZ}{dA}\delta A \tag{26}$$

Likewise, if $Z = f(A,B)$ and A and B are independent, then δZ is given by Equation (27):

$$\delta Z = \left[\left(\frac{\partial Z}{\partial A}\delta A\right)^2 + \left(\frac{\partial Z}{\partial B}\delta B\right)^2\right]^{1/2} \tag{27}$$

The expressions given in Equations (28), (29) and (30) were then derived for the errors in ΔG_{FU} (Figure 6A), ΔH_{FU} (Figure S4 in the Supporting Information) and $\Delta C_{p_{FU}} T^{-1}$ (Figure 7):

$$\delta(\Delta G_{FU}) = RT\frac{1}{P_F(1-P_F)}\delta P_F \tag{28}$$

$$\delta(\Delta H_{FU}) = [(\delta H_F)^2 + (\delta H_U)^2]^{1/2} \tag{29}$$

$$\delta(\Delta C_{p_{FU}} T_b{}^{-1}) = [(\alpha\delta(\Delta H_{FU}^{T_a}))^2 + ((\alpha-\gamma)\delta(\Delta H_{FU}^{T_b}))^2 + (\gamma\delta(\Delta H_{FU}^{T_c}))^2]^{1/2} T_b{}^{-1} \tag{30}$$

in which T_a, T_b, and T_c are defined in Equations (15), (16), and α and γ are given by Equations (31) and (32).

$$\alpha = \frac{1}{2(T_b - T_a)} \tag{31}$$

$$\gamma = \frac{1}{2(T_c - T_b)} \tag{32}$$

Acknowledgements

The authors thank the Swiss National Science Foundation and its National Competence Center for Research (NCCR) in Structural Biology, and the Spanish MEC/FEDER, ref. BIO2003–02848, for financial support. The Port d'Informació Científica (PIC) and the Òliba project of UAB are thanked for providing storage and computational resources.

[1] a) J. Norberg, L. Nilsson, *Q. Rev. Biophys.* **2003**, *36*, 257–306; b) S. A. Adcock, J. A. McCammon, *Chem. Rev.* **2006**, *106*, 1589–1615; c) W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt, H. B. Yu, *Angew. Chem.* **2006**, *118*, 4168–4198; *Angew. Chem. Int. Ed.* **2006**, *45*, 4064–4092.
[2] D. Seebach, P. E. Ciceri, M. Overhand, B. Jaun, D. Rigo, L. Oberer, U. Hommel, R. Amstutz, H. Widmer, *Helv. Chim. Acta* **1996**, *79*, 2043–2066.
[3] X. Daura, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, *J. Mol. Biol.* **1998**, *280*, 925–932.
[4] D. Seebach, A. K. Beck, D. J. Bierbaum, *Chem. Biodiversity* **2004**, *1*, 1111–1239.
[5] P. W. Atkins, *Physical Chemistry*, 6th edition, Oxford University Press, Oxford, **1998**, pp. 45–140.
[6] W. F. van Gunsteren, X. Daura, A. E. Mark, *Helv. Chim. Acta* **2002**, *85*, 3113–3129.
[7] a) T. Imai, Y. Harano, A. Kovalenko, F. Hirata, *Biopolymers* **2001**, *59*, 512–519; b) D. Paschek, S. Gnanakaran, A. E. Garcia, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6765–6770.
[8] G. I. Makhatadze, K. S. Kim, C. Woodward, P. L. Privalov, *Protein Sci.* **1993**, *2*, 2028–2036.
[9] J. Pitera, private communication.
[10] K. Gademann, B. Jaun, D. Seebach, R. Perozzo, L. Scapozza, G. Folkers, *Helv. Chim. Acta* **1999**, *82*, 1–11.
[11] L. J. Smith, X. Daura, W. F. van Gunsteren, *Proteins: Struct. Funct. Genet.* **2002**, *48*, 487–496.
[12] a) H.-A. Yu, M. Karplus, *J. Chem. Phys.* **1988**, *89*, 2366–2379; b) H. Qian, J. J. Hopfield, *J. Chem. Phys.* **1996**, *105*, 9292–9298.
[13] N. F. A. van der Vegt, M. E. Lee, D. Trzesniak, W. F. van Gunsteren, *J. Phys. Chem. B* **2006**, *110*, 12852–12855.
[14] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, vdf Hochschulverlag AG an der ETH Zürich and BIOMOS b. v., Zürich, Groningen, **1996**.
[15] X. Daura, A. E. Mark, W. F. van Gunsteren, *J. Comput. Chem.* **1998**, *19*, 535–547.
[16] X. Daura, W. F. van Gunsteren, D. Rigo, B. Jaun, D. Seebach, *Chem. Eur. J.* **1997**, *3*, 1410–1417.
[17] X. Daura, W. F. van Gunsteren, A. E. Mark, *Proteins: Struct. Funct. Genet.* **1999**, *34*, 269–280.
[18] R. Friedberg, J. E. Cameron, *J. Chem. Phys.* **1970**, *52*, 6049–6058.

Received: September 3, 2007
Published online: April 9, 2008

