Entropy Calculations on a Reversibly Folding Peptide: Changes in Solute Free Energy Cannot Explain Folding Behavior

Heiko Schäfer,[1] Xavier Daura,[1] Alan E. Mark,[2] and Wilfred F. van Gunsteren[1*]

[1]Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH Zentrum, Zurich, Switzerland
[2]Laboratory of Biophysical Chemistry, University of Groningen, AG Groningen, The Netherlands

ABSTRACT **The configurational entropy of a β-heptapeptide in solution at four different temperatures is calculated. The contributions of the backbone and of the side-chain atoms to the total peptide entropy are analyzed separately and the effective contribution to the entropy arising from correlations between these terms determined. The correlation between the backbone and side-chain atoms amounts to about 17% and is rather insensitive to the temperature. The correlation of motion within the backbone and within side-chains is much larger and decreases with temperature. As the peptide reversibly folds at higher temperatures, its change in entropy and enthalpy upon folding is analyzed. The change in entropy and enthalpy upon folding of the peptide alone cannot account for the observed change in free energy on folding of the peptide in solution. Enthalpic and entropic contributions of the solvent thus also play a key role. Proteins 2001; 43:45–56.** © 2001 Wiley-Liss, Inc.

INTRODUCTION

Entropy is probably the most abused thermodynamic quantity: it is often either ignored or identified as the cause of inexplicable results. In contrast to the free energy that is easily accessible by experiment, determination of the entropy causes grief to the experimentalist and the theoretician. Although entropy can, in principle, be calculated from molecular dynamics simulation trajectories,[1–4] the calculations are hampered by methodological, convergence, and computational problems. Consequently, much less effort has been spent on the development of entropy calculations as compared with free energy calculations. This situation is changing. With the availability of very long simulation trajectories (up to a couple of 100 ns), which cover a high proportion of the accessible configuration space of the molecular system of interest, as well as more effective methods to estimate entropy,[5,6] the role of the entropy in complex molecular systems can now be studied.

The importance of entropy in determining protein folding is well recognized.[7] Hydrophobic clustering is commonly attributed to an increase in entropy within the solvent. The much touted folding funnels have been discussed in terms of an interplay between the configurational entropy of the peptide chain, which would favor the unfolded state and enthalpic interactions (energy gap)

favoring the native or folded conformation.[8] On a more local level, the effects of substitutions involving glycine and proline on stability are often attributed to changes in entropy within the folded or unfolded state. Entropic barriers have also been used to explain the mechanism of protein folding[8] and differences in side-chain entropy have been evoked to explain the differences in the rate of folding of different proteins.[9]

Using experimental calorimetric studies, only the total change in entropy associated with a given process is directly accessible. Despite attempts to correlate changes in nuclear magnetic resonance (NMR) derived order parameters to changes in configurational entropy,[10] estimates of entropic contributions in the literature to date depend primarily on database analysis[11,12] or simple models.[13] The current study is the first in which analysis of the entropy is based on extensive sampling of a realistic model of a protein or peptide under equilibrium conditions. In this work we examine the configurational entropy of a small non-natural peptide in solution[14] that adopts a well-characterized helical structure at room temperature and shows reversible folding/unfolding transitions in 200-ns trajectories. Specifically, we examine the total peptide entropy both above and below the melting temperature, the contributions of the backbone and side-chain atoms to the total peptide entropy, and the magnitude of the effects of correlations between backbone and side-chain atoms. Although clearly there are differences between the folding characteristics of peptides and proteins, the basic physical principles remain essentially the same. The conclusions drawn in regard to the specific peptide system will yield insights into the role played by entropy in folding of proteins in general.

METHOD

Entropy calculations are notoriously difficult, as the entropy depends on the complete phase space of a molecular system and is sensitive to the inclusion of correlations between motions along the many degrees of freedom. To escape from the high dimensionality of phase space, many methods[2–4] employ internal generalized coordinates and

*Correspondence to: Wilfred F. van Gunsteren, Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH Zentrum, CH-8092 Zurich, Switzerland. E-mail: wfvgn@igc.phys.chem.ethz.ch

use direct sampling of degrees of freedom or a harmonic approximation to motion, or a combination of both approaches. The transformation to internal coordinates reduces the dimensionality of the problem but generates additional complications because of mass-metric tensor effects.[2,4]

An approximate expression for the entropy introduced by Schlitter[5] based on a quantum-mechanical harmonic approximation does not require the transformation of simulation trajectories to internal coordinates. The Appendix provides a very brief summary of Schlitter's formula. It was implemented to work with the GROMOS96 simulation package[15,16] and was extensively tested.[6] It could be shown that for condensed-phase molecular systems at physiological temperatures ($\sim$300 K) the formula yields results with errors below 5%.

Schlitter's formula uses the covariance matrix of the atom-positional fluctuation, which is easily computed from the trajectory of a simulation. Because the cartesian coordinates are directly used to calculate the covariance matrix, the overall translational and rotational motion is included, possibly leading to convergence problems. This can be avoided by removing the translation of the center of mass and the rotation around the center of mass of the molecule by performing translational/rotational fits of the trajectory structures of the molecule during calculation of the covariance matrix.[6] So, the entropy of the translational center of mass motion can be excluded from the calculation. The same holds for the overall rotational entropy, with the caveat that overall rotation cannot be rigorously separated from internal motion for a flexible molecule. The entropy contributions of internal degrees of freedom (e.g., torsional angles) could also be obtained using Schlitter's formula. However, this would require a transformation to internal coordinates, with its concomitant metric tensor complications, losing the advantage of Schlitter's formula, i.e., that cartesian coordinates can be used. In that case, it is preferable to use standard formulae to compute the entropy. The use of Schlitter's formula is straightforward when the entropy of subgroups of atoms is of interest. Only the atoms of interest are then included in calculating the covariance matrix.

The off-diagonal elements of the covariance matrix are nonzero because of the correlation between the cartesian coordinates of all atoms. If the entropy is calculated using the complete covariance matrix, these correlations are included. By contrast, if only the diagonal elements of the covariance matrix are used, all correlations between cartesian coordinate components of atoms are ignored, resulting in a larger entropy. In this way, correlation between atoms can be estimated.

A β-HEPTAPEPTIDE IN SOLUTION

The system studied consists of a β-heptapeptide, a non-natural peptide composed of β-amino acid residues. The system has been studied extensively by NMR[17] and molecular dynamics simulations[14,18] and shows helical secondary structure in methanol. Figure 1 shows the structural formula, the helical conformation as deter-
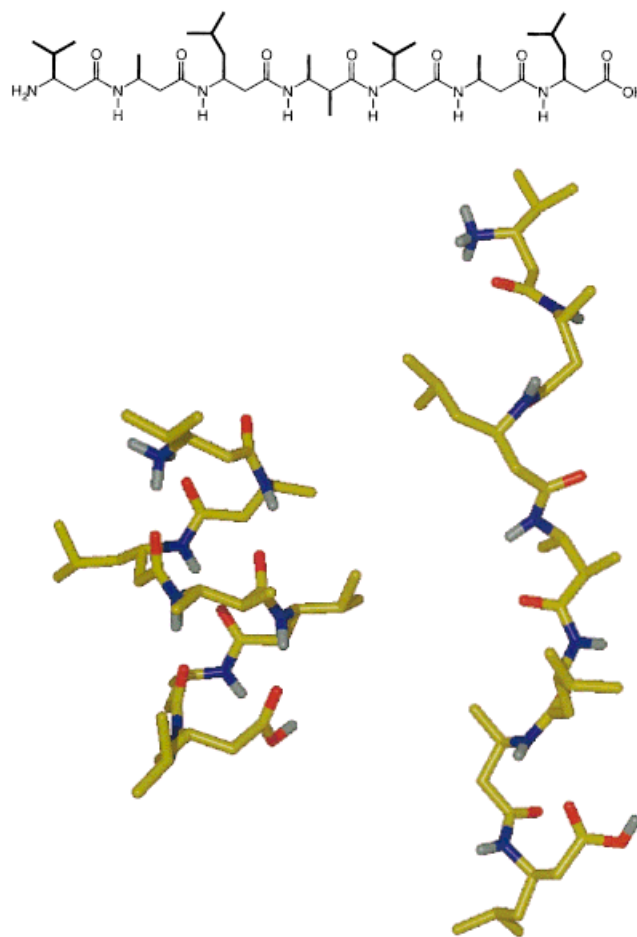

Fig. 1. Structural formula of the β-heptapeptide, NMR determined helical folded model structure (*left*), and an extended structure from the 350 K simulation (*right*).

mined by NMR at room temperature, and an extended structure taken from the simulation.

The β-heptapeptide in methanol was simulated in a periodic computational box at four different temperatures (298, 340, 350, and 360 K) using the GROMOS96 simulation package.[15,16] The NMR model structure was taken as the initial structure for the simulations at 298, 340, and 350 K. The 360-K simulation was started from a fully extended conformation of the peptide. The simulations at 298 K and 340 K covered 200 ns, those at 350 K and 360 K 50 ns each. Configurations saved at 0.5-ps intervals were used in the calculations. The peptide undergoes reversible folding in the simulations, and the ratio of folded to unfolded structures decreases with increasing temperature from 25:1 at 298 K to 1:3 at 360 K.

The atom-positional root-mean-square deviation (RMSD) from the (folded) NMR model structure was taken as an index to assign the configurations to the folded or unfolded categories. The RMSD of the main chain atoms in residues 2–6 from the NMR model structure is shown for all four temperatures in the lower panels of Figures 2–5. Configurations with an RMSD with respect to the NMR model structure of <0.1 nm are considered to represent the
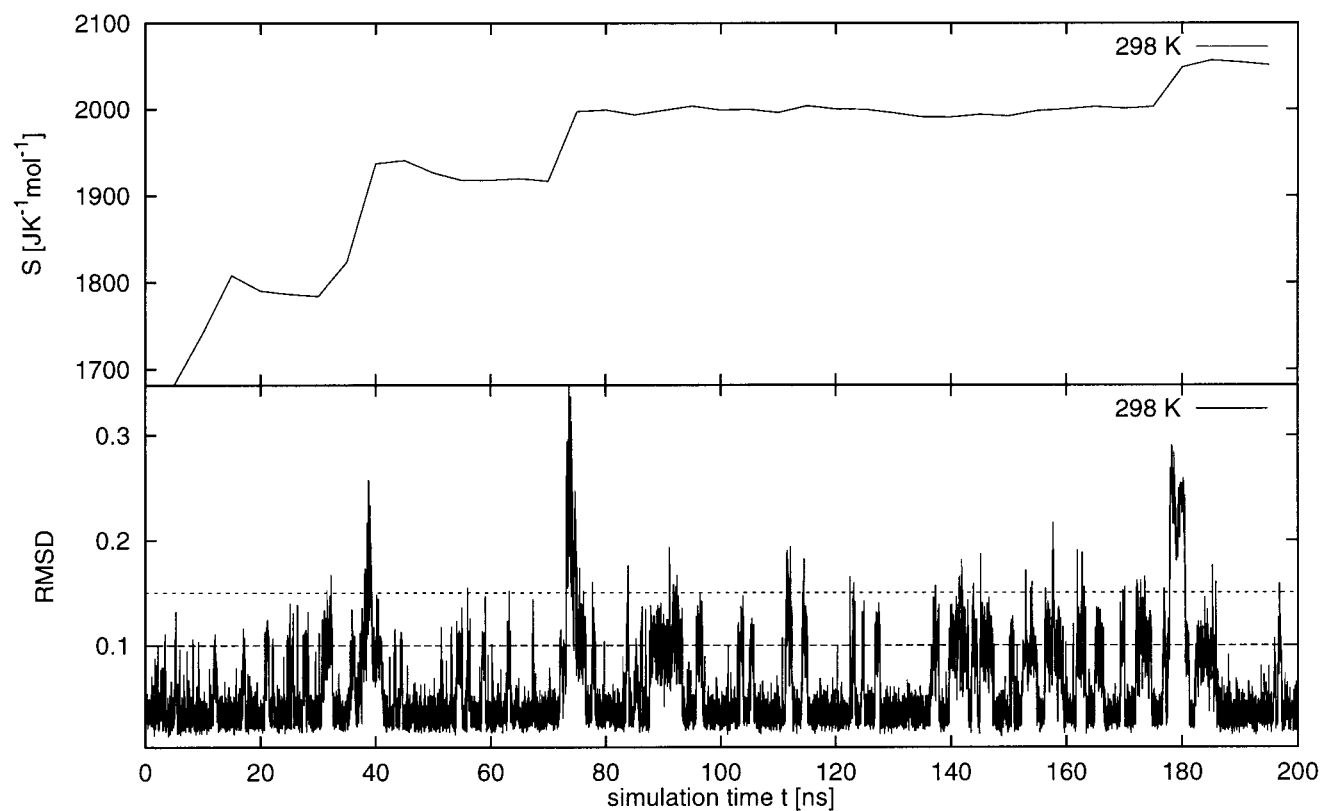
Fig. 2. Absolute entropy S_{all} of the β-heptapeptide (excluding overall translation and rotation) and backbone atom-positional root-mean-square deviation (RMSD) (residues 2–6) from the NMR model structure. Simulation at 298 K.

Fig. 3. Absolute entropy S_{all} of the β-heptapeptide (excluding overall translation and rotation) and backbone atom-positional root-mean-square deviation (RMSD) (residues 2–6) from the NMR model structure. Simulation at 340 K.

Fig. 4.   Absolute entropy S_{all} of the β-heptapeptide (excluding overall translation and rotation) and backbone atom-positional root-mean-square deviation (RMSD) (residues 2–6) from the NMR model structure. Simulation at 350 K.
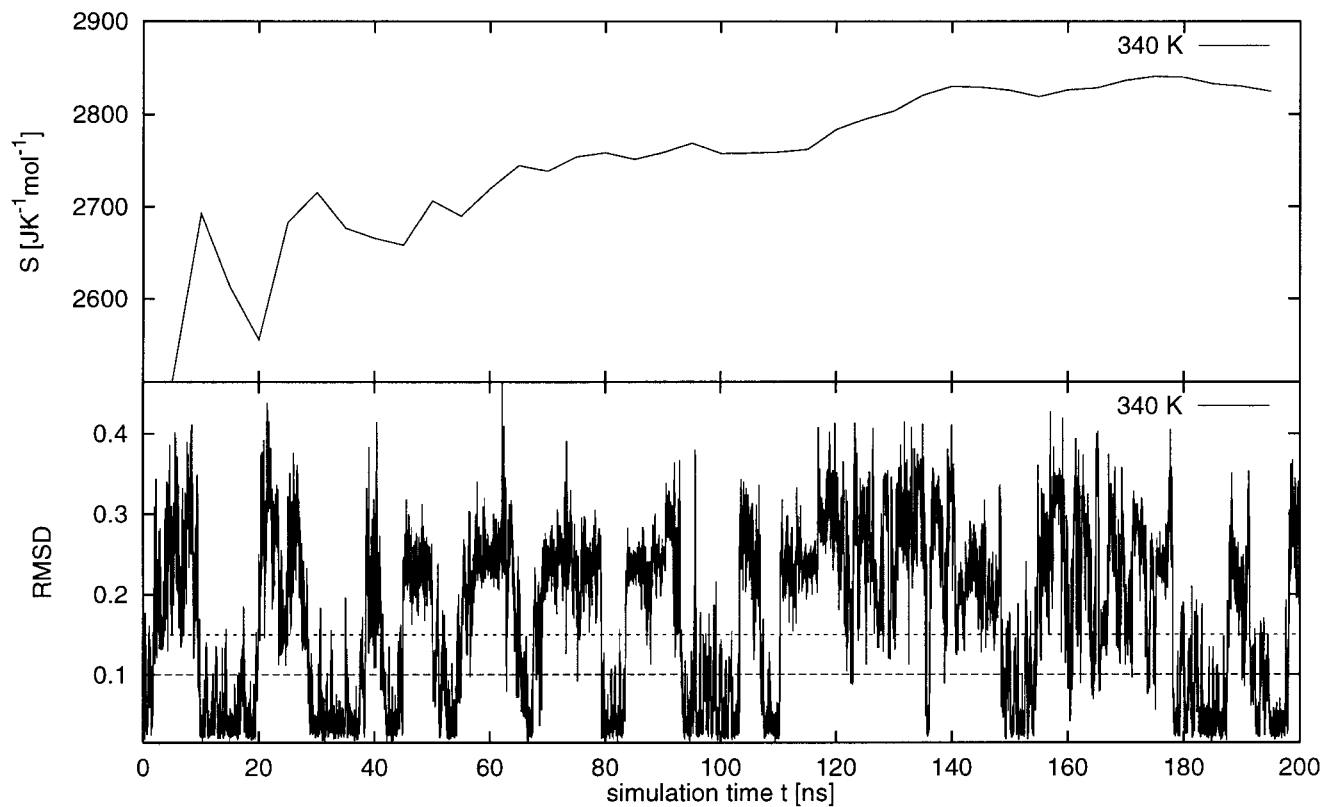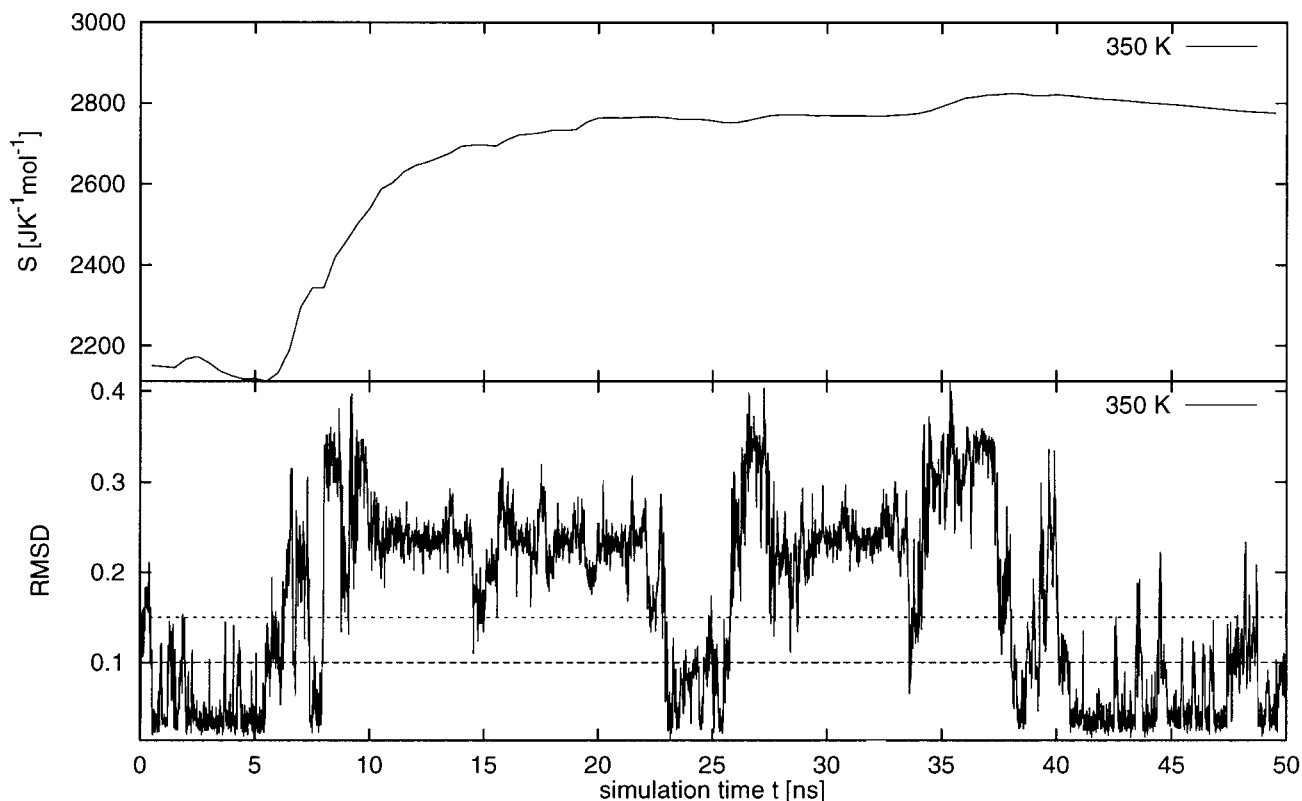
folded state, while configurations with an RMSD of >0.15 nm are considered unfolded.[14]

CONFIGURATIONAL ENTROPY OF THE β-HEPTAPEPTIDE

The configurational entropy of the β-heptapeptide was calculated for all four temperatures, using all configurations (of the β-heptapeptide only, excluding the solvent) in the trajectory irrespective of their fold. The peptide configurations were fitted on top of each other using a least-squares fit of all peptide atoms. This was done in order to remove the overall translation of the center of mass and the overall rotation of the peptide. The overall rotation cannot be removed in an unambiguous way. The effect of fitting based on different subsets of peptide atoms has been analyzed elsewhere.[6]

Figures 2–5 show the peptide entropy in the upper panels and the atom-positional RMSD from the helical model structure in the lower panels, as a function of time for all four temperatures. There is a striking correlation between unfolding events in the RMSD, i.e., large increases in the RMSD, and sharp increases in the entropy. This can best be seen at 298 K, where the peptide stays in the helical fold for nearly the whole simulation. The RMSD shows three major unfolding events mirrored in the entropy by three jumps. These jumps in the entropy occur when the peptide explores new regions in phase space. At higher temperatures, the jumps are less well correlated with the RMSD, especially at the end of the trajectory, when unfolding events may not necessarily open up new regions of phase space. The decrease in entropy that can be observed for some parts of the trajectories is possibly less intuitive. It is best illustrated with the first 20 ns of the simulation at 340 K (Fig. 3). After the initial 10 ns, in which the peptide stays mostly unfolded and every new conformation that is sampled induces a significant increase in entropy, the peptide folds and remains in basically the same conformation for the next 10 ns, reducing its entropy. The effect on the total entropy of this resampling of conformations is still patent toward the end of the trajectory but, as with the sampling of new conformations, the weights become smaller.

The ordering of the entropies with temperature is generally correct: entropy increases with increasing temperature. The entropy at 350 K is slightly lower than at 340 K, which can be explained by the different lengths of the simulations, 50 ns versus 200 ns, respectively. The trajectory averages at 350 K are generally less well converged. This reversed order of 340 K and 350 K is apparent in most results.

For all four temperatures, the configurational entropy levels off at the end of the trajectory, indicating a relative convergence. It is not necessarily an indication that the whole of phase space has been sampled, but that impor-
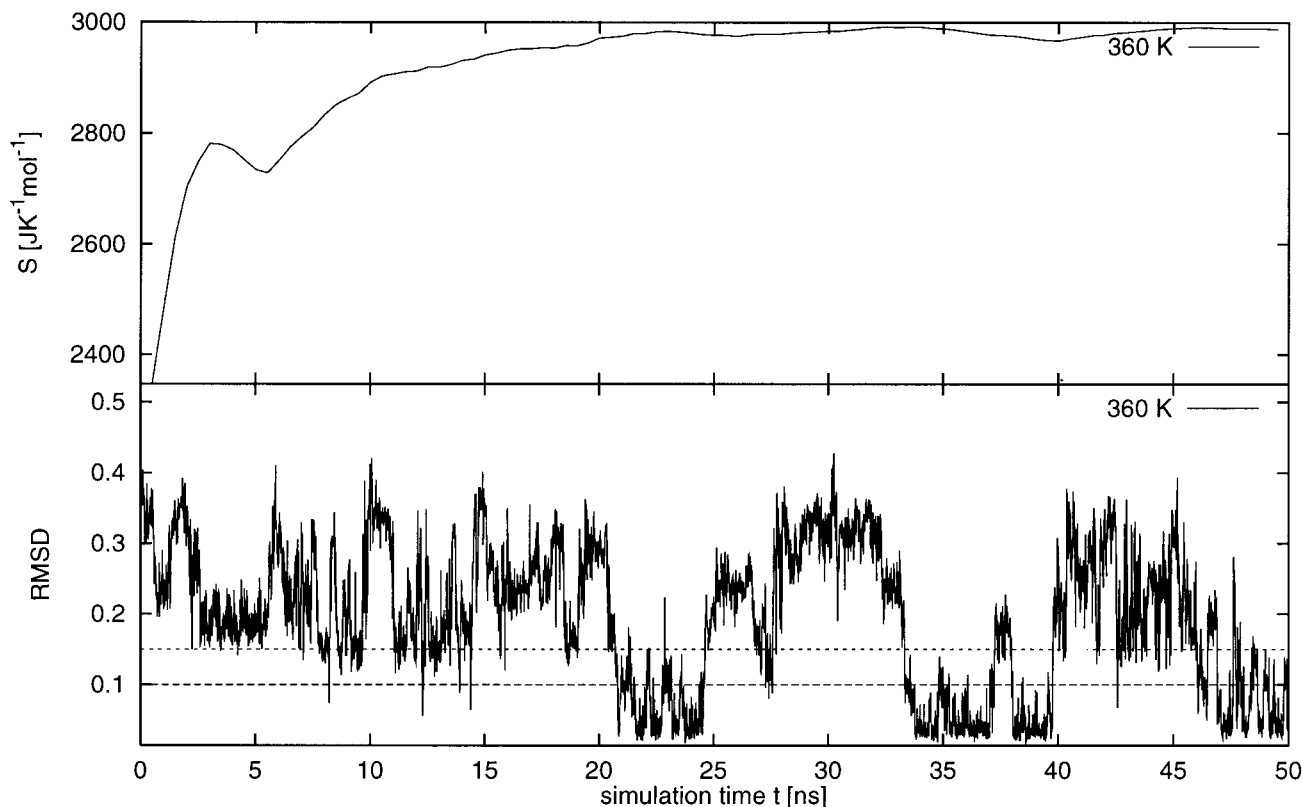
Fig. 5. Absolute entropy S_{all} of the β-heptapeptide (excluding overall translation and rotation) and backbone atom-positional root-mean-square deviation (RMSD) (residues 2–6) from the NMR model structure. Simulation at 360 K.

tant regions of phase space accessible to the peptide have been sufficiently sampled. It has also been shown[6] that the overall rotation of the peptide and the volume of the (periodic) computational box are sampled rather completely.

ENTROPY PER RESIDUE

By using only a subset of the atoms of a molecule in the calculation of the covariance matrix of atom-positional fluctuations, it is possible to calculate the entropy of this particular subset of atoms. It is therefore possible to calculate the entropy of each of the seven residues of the β-heptapeptide. However, when calculating the entropy of a subset of atoms in the system, the correlations with the rest of the system (i.e., all atoms not in the subset) are ignored. If correlations between two subsets of atoms are non-negligible, the sum of the entropies of the two subsets will be larger than the entropy calculated from the combined subsets.[19,20] The entropy per residue was calculated using all peptide configurations in the saved trajectory. Each configuration was fitted onto the first using a least-squares fit for a particular set of atoms. To identify possible artifacts originating in the fitting procedure, three different sets of atoms were used in the least-squares fit: (1) all atoms of the peptide; (2) only the backbone atoms; and (3) four atoms in the (4th) central residue (the carbon bound to the nitrogen and its three covalently

bound neighbors). Because the residues possess different numbers of atoms, the entropy per residue was normalised using the number of atoms.

The entropies S_x^{res} of residue x for all four temperatures and the three different translational/rotational fits are shown in Figure 6. At 298 K, the central residue shows a lower entropy than the ends of the peptide chain. Using only four atoms in the central residue in the least-squares fit increases, not unexpectedly, the difference in entropy between the centre and ends of the peptide chain. The differences in residual entropy are smaller at higher temperatures, as the mobility of the central residues increases because of more frequent unfolding events. The use of only four atoms in the central residue for fitting results in an artificially low entropy for the central residue. In this case, even at the highest temperature, there is a pronounced entropy difference between the center and the ends of the peptide chain.

ENTROPY OF THE BACKBONE AND SIDE-CHAINS

The atoms of the peptide were split into two groups: one consisting of all (46) backbone atoms, the other consisting of all (18) side-chain atoms. For both subsets, the entropy was calculated using all configurations in each saved trajectory. The peptide configurations were fitted on top of each other using a least-squares fit for all atoms. Entropy
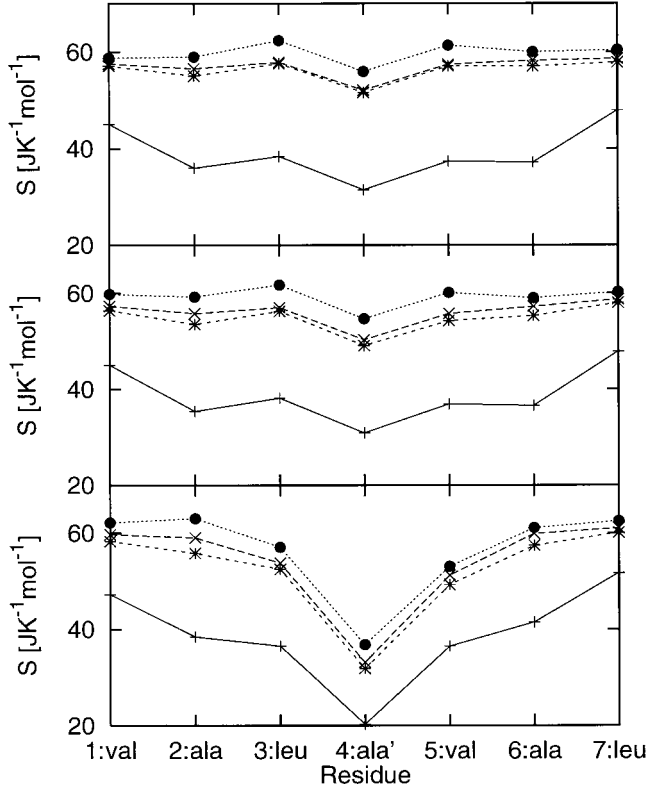
Fig. 6.   Entropy per residue per atom for 298 K (+), 340 K (∗), 350 K (×), and 360 K (●). The entropy has been normalized using the number of atoms in each residue. Different sets of atoms are used in the translational/rotational least-squares fit aimed at removal of overall translational and rotational contributions to the entropy. *Top,* all atoms. *Middle,* all backbone atoms. *Bottom,* four atoms of the central residue.

TABLE I. Configurational Entropy of the Entire β-Heptapeptide (S_{all}), Its Backbone (S_{bb}), and Side-Chains (S_{sc}) After a Structural Least-Squares Fit Using All Atoms[†]

T	S_{all} (J K^{-1} mol^{-1})	S_{all}^{-} (J K^{-1} mol^{-1})
298 K	2051 (32)	4381 (69)
340 K	2825 (44)	5774 (90)
350 K	2776 (43)	5864 (92)
360 K	2987 (47)	5881 (92)
T	S_{bb} (J K^{-1} mol^{-1})	S_{bb}^{-} (J K^{-1} mol^{-1})
298 K	1359 (30)	2912 (63)
340 K	1923 (42)	3934 (86)
350 K	1882 (41)	3993 (87)
360 K	2020 (44)	3977 (87)
T	S_{sc} (J K^{-1} mol^{-1})	S_{sc}^{-} (J K^{-1} mol^{-1})
298 K	1041 (58)	1469 (82)
340 K	1370 (76)	1840 (102)
350 K	1358 (75)	1872 (104)
360 K	1437 (80)	1904 (106)

[†]Entropy was calculated using the complete covariance matrix (no superscript) and using only the diagonal elements (superscript −). The entropy per atom is given in parentheses.

TABLE II. Distribution of Entropy Between Backbone and Side-Chains, and Correlation Between Backbone and Side-Chains[†]

	T (K)			
	298	340	350	360
Decrease in entropy due to correlation $\Delta S_{sc/bb}^{corr} = S_{sc} + S_{bb} - S_{all}$ (J K^{-1} mol^{-1})	349 (5)	469 (7)	464 (7)	470 (7)
Percentage of the correlation $\Delta S_{sc/bb}^{corr}/S_{all}$ (%)	17 (17)	17 (17)	17 (17)	16 (16)
Percentage of the entropy in the backbone $\dfrac{S_{bb}}{S_{bb} + S_{sc}}\dfrac{S_{all} - \Delta S_{sc/bb}^{corr}}{S_{all}}$ (%)	47 (28)	49 (30)	48 (29)	49 (30)
Percentage of the entropy in the side-chains $\dfrac{S_{sc}}{S_{bb} + S_{sc}}\dfrac{S_{all} - \Delta S_{sc/bb}^{corr}}{S_{all}}$ (%)	36 (55)	35 (54)	35 (54)	35 (54)

[†]$\Delta S_{sc/bb}^{corr}$ is the decrease in entropy because of correlation between the side-chains and the backbone. The percentage of the correlation on the configurational entropy S_{all} and the percentage of entropy contributed by the side-chains and the backbone are given. Values calculated on a per-atom basis are given in parentheses.

was calculated using the full covariance matrix, thus including all correlations within the particular (sub)set of atoms (backbone or side-chains), and neglecting all correlations by using only the diagonal elements of the covariance matrix. Table I shows the results of the calculation of the entropy of the entire peptide, S_{all}; of the backbone atoms, S_{bb}; and of the side-chain atoms, S_{sc}. Superscript $^-$ indicates an entropy calculation in which all correlations were neglected. Because the backbone and side-chain subgroups of atoms have different numbers of atoms, the entropy values were normalized through division by the number of atoms. The normalized values are given in parentheses in Table I.

Comparing the total peptide entropy S_{all} with the value S_{all}^- obtained by neglecting all correlations, the importance of motional correlations becomes clear: inclusion of correlations reduces the entropy by about a factor of 2 at the four temperatures.

Comparing the entropy of the side-chains, the backbone, and the entire peptide, the backbone contributes more entropy than do the side-chains. However, per atom, the side-chains contain more entropy than the backbone. This larger share of the entropy in the side-chains can be attributed to the side-chains greater flexibility. Unfortunately, artifacts from the least-squares fit cannot be

neglected: if the backbone moves, the side-chains will also be moved, and this motion will not be removed by the least-squares fit. Nevertheless, at 298 K, the least-squares fit has been shown to be good,[6] and the entropy per atom in the side-chains at 298 K is roughly 50% higher than in the backbone.

Table II shows the distribution of the entropy between the backbone and the side-chains, and the correlation between backbone and side-chains. Again, it is clear that the side-chains contribute relatively more to the entropy per atom than does the backbone. The distribution is only slightly dependent on the temperature: the ratio between entropy in the side-chains to entropy in the backbone is
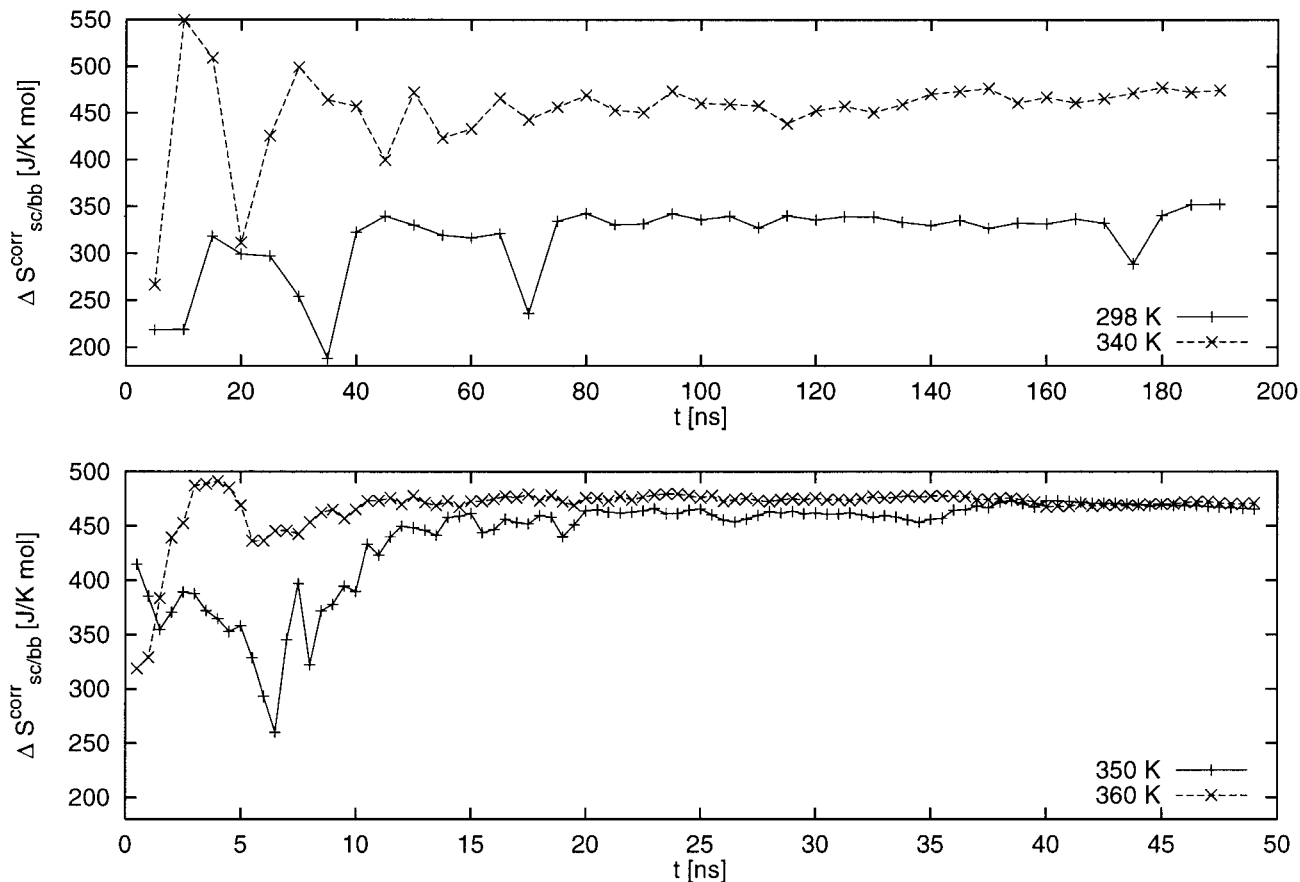
Fig. 7. Decrease in the entropy due to the correlation between the side-chains and the backbone. Entropies of the side-chains, backbone, and the entire molecule were all calculated after a translational and rotational least-squares fit on all atoms. The curves show the difference between the sum of the backbone and side-chain entropies and the entropy of the entire peptide using all atoms (and including all correlations between the atoms) [eq. (1)].

slightly higher at 298 K than at the higher temperatures. As unfolding events become more frequent, the backbone motion is increased relative to that of the side-chains; therefore, the backbone contributes more to the entropy.

Correlation Between Backbone and Side-Chains

The motion of the side-chains and the backbone is, naturally, correlated. In calculating the configurational entropy S_{all} of the peptide, the correlations between all atoms of the peptide are taken into account. In the calculations of S_{bb} and S_{sc}, only the correlations inside the particular subgroup of atoms (backbone, side-chains) are taken into account. Therefore, the correlation between the side chain and the backbone atoms can be calculated as

$$\Delta S_{sc/bb}^{corr} = S_{sc} + S_{bb} - S_{all} \tag{1}$$

Table II shows the values for $\Delta S_{sc/bb}^{corr}$ for all four temperatures. A jump in the correlation between 298 K and 340 K mirrors a sudden increase in the configurational entropy S_{all} as unfolding sets in. The absolute increase in correlation is misleading because the percentage of the correlation $\Delta S_{sc/bb}^{corr}$ on the configurational entropy (Table II, second line) shows no significant change with increasing temperature. The correlation between backbone and side-

chain motion is rather insensitive to the temperature. This is not so surprising for a small peptide with a helical secondary structure, which positions the side-chains on the peptide surface.

$\Delta S_{sc/bb}^{corr}$ is shown as a function of simulation time in Figure 7. The correlation between the backbone and the side-chains shows large fluctuations in the beginning but quickly reaches a constant level. Especially at 298 K, the curve shows distinct dips at 35, 70, and 175 ns, time points at which major unfolding events take place (Fig. 2). When the peptide explores new conformations, the correlation between backbone and side-chains decreases. The correlation builds up again very quickly when the peptide stays in the new conformation.

Cooperative Motion Within the Backbone and Within the Side-Chains

Correlation does not occur only between the side-chains and the backbone of the peptide but within the side-chains and within the backbone as well. The entropy of the subsets (side-chain atoms, backbone atoms) was not only calculated using the full covariance matrix (S_{bb}, S_{sc}), but also using only the diagonal elements of the covariance matrix (S_{bb}^-, S_{sc}^-). The difference between these two values

TABLE III. Effect of Correlation Within Subsets of Atoms[†]

	T (K)			
	298	340	350	360
Correlation within backbone $\Delta S_{\text{bb}}^{\text{corr}} = S_{\text{bb}}^{-} - S_{\text{bb}} \,(\text{J K}^{-1}\,\text{mol}^{-1})$	1553 (34)	2010 (44)	2111 (46)	1957 (43)
Correlation within side chains $\Delta S_{\text{sc}}^{\text{corr}} = S_{\text{sc}}^{-} - S_{\text{sc}} \,(\text{J K}^{-1}\,\text{mol}^{-1})$	428 (24)	470 (26)	513 (29)	467 (26)
Difference in correlation $\Delta\Delta S_{\text{bb/sc}}^{\text{corr}} = \Delta S_{\text{bb}}^{\text{corr}} - \Delta S_{\text{sc}}^{\text{corr}} \,(\text{J K}^{-1}\,\text{mol}^{-1})$	1125 (10)	1541 (18)	1597 (17)	1490 (17)
Percentage of correlation within the backbone $\Delta S_{\text{bb}}^{\text{corr}}/S_{\text{bb}} \,(\%)$	114	105	112	97
Percentage of correlation within the backbone (w.r.t. S_{bb}^{-}) $\Delta S_{\text{bb}}^{\text{corr}}/S_{\text{bb}}^{-} \,(\%)$	53	51	53	49
Percentage of correlation within side-chains $\Delta S_{\text{sc}}^{\text{corr}}/S_{\text{sc}} \,(\%)$	41	34	38	32
Percentage of correlation within side-chains (w.r.t. S_{sc}^{-}) $\Delta S_{\text{sc}}^{\text{corr}}/S_{\text{sc}}^{-} \,(\%)$	29	26	27	25

[†]$\Delta S_{\text{sc}}^{\text{corr}}$ is the decrease in the entropy because of correlation within and between all side-chains, $\Delta S_{\text{bb}}^{\text{corr}}$ is the entropy decrease because of correlation inside the backbone. Values in parentheses are normalized through division by the number of atoms. $\Delta\Delta S_{\text{sc/bb}}^{\text{corr}}$ is the difference between the correlation within the backbone and side-chains. The correlation is also given in percentage with respect to the configurational entropy $S_{xx\,=\,\text{sc,bb}}$ and with respect to the uncorrelated configurational entropy (superscript $-$).

TABLE IV. Peptide Entropy of Folded Configurations (Atom-Positional RMSD for Residues 2–6 From NMR Model Structure < 0.1 nm) S_{fold} and of Unfolded Configurations (RMSD > 0.15 nm) S_{unfold}[†]

	T (K)			
	298	340	350	360
Entropy of folded configurations $S_{\text{fold}} \,(\text{J K}^{-1}\,\text{mol}^{-1})$	1810	1976	1995	2048
Entropy of unfolded configurations $S_{\text{unfold}} \,(\text{J K}^{-1}\,\text{mol}^{-1})$	2337	2854	2763	2921
Enthalpy of folded configurations $H_{\text{fold}} \,(\text{kJ mol}^{-1})$	−500	−473	−467	−463
Enthalpy of unfolded configurations $H_{\text{unfold}} \,(\text{kJ mol}^{-1})$	−429	−386	−411	−371
Difference in entropy $\Delta S_{\text{fu}} = S_{\text{fold}} - S_{\text{unfold}} \,(\text{J K}^{-1}\,\text{mol}^{-1})$	−527	−878	−768	−873
Difference in enthalpy $\Delta H_{\text{fu}} = H_{\text{fold}} - H_{\text{unfold}} \,(\text{kJ mol}^{-1})$	−71	−87	−56	−92
Free energy of folding $\Delta G_{\text{fu}} = \Delta H_{\text{fu}} - T\Delta S_{\text{fu}} \,(\text{kJ mol}^{-1})$	86	212	213	222
Free energy from cluster analysis $\Delta G_{\text{fu}}^{\text{cl}} \,(\text{kJ mol}^{-1})$	−7.4	1.6	1.3	3.3
Free energy from counting $\Delta G_{\text{fu}}^{\text{ratio}} \,(\text{kJ mol}^{-1})$	−8.1	1.9	1.7	3.5

[†]Together with the peptide enthalpy of the folded/unfolded configurations $H_{\text{fold}}/H_{\text{unfold}}$ the values S_{fold} and S_{unfold} are used to estimate a peptide free energy of folding ΔG_{fu}. $\Delta G_{\text{fu}}^{\text{cl}}$ is the free energy of folding calculated from the cluster analysis of trajectory conformations, while $\Delta G_{\text{fu}}^{\text{ratio}}$ is calculated by counting folded versus unfolded conformations. The latter two values include peptide–solvent enthalpies and correlations and solvent enthalpies and correlations, whereas ΔG_{fu} only contains peptide contributions to enthalpy and entropy.

TABLE V. Contributions From Backbone and Side-Chains to the Change in Configurational Entropy upon Folding[†]

	T (K)			
	298	340	350	360
Entropy of the backbone in folded configurations $S_{\text{fold}}^{\text{bb}}$ (J K^{-1} mol^{-1})	1126 (24)	1239 (27)	1247 (27)	1286 (28)
Entropy of the backbone in unfolded configurations $S_{\text{unfold}}^{\text{bb}}$ (J K^{-1} mol^{-1})	1520 (33)	1897 (41)	1831 (40)	1942 (42)
Entropy in the side-chains in folded configurations $S_{\text{fold}}^{\text{sc}}$ (J K^{-1} mol^{-1})	931 (52)	999 (55)	1015 (56)	1033 (57)
Entropy in the side-chains in unfolded configurations $S_{\text{unfold}}^{\text{sc}}$ (J K^{-1} mol^{-1})	1179 (66)	1385 (77)	1352 (75)	1402 (78)
Entropy change upon folding in backbone $\Delta S_{\text{fu}}^{\text{bb}}$ (J K^{-1} mol^{-1})	−394 (−9)	−657 (−14)	−584 (−13)	−657 (−14)
Entropy change upon folding in side-chains $\Delta S_{\text{fu}}^{\text{sc}}$ (J K^{-1} mol^{-1})	−247 (−14)	−386 (−21)	−337 (−19)	−369 (−20)

[†]Values per atom are given in parentheses.

$$\Delta S_{\text{xx}}^{\text{corr}} = S_{\text{xx}}^{-} - S_{\text{xx}} \tag{2}$$

is the decrease in entropy because of correlation between the atoms of set xx [backbone (bb) or side-chains (sc)].

The entropy of the backbone and of the side-chains, including (S_{xx}) or excluding (S_{xx}^{-}) correlations, is shown in Table I. The change in entropy because of correlations within the backbone ($\Delta S_{\text{bb}}^{\text{corr}}$) or within the side-chains ($\Delta S_{\text{sc}}^{\text{corr}}$) are shown in Table III. Values normalized through division by the number of atoms are given in parentheses. The decrease in entropy due to correlation in the backbone $\Delta S_{\text{bb}}^{\text{corr}}$ is consistently larger than in the side-chains $\Delta S_{\text{sc}}^{\text{corr}}$. This is indicative of a more correlated motion of the backbone even at higher temperatures where unfolded configurations dominate. This is hardly surprising, as the backbone is, by definition, connected by bonds, whereas the side-chains are not directly connected to each other. Table III also gives the relative decrease in entropy with respect to the configurational entropies S_{bb} and S_{sc}. With increasing temperature the importance of correlations decreases within the backbone, as well as within the side-chains. Yet there is a greater decrease in the backbone, most likely because of the onset of unfolding.

FOLDING

In the previous sections, the configurational entropy of the peptide was calculated using all configurations, irrespective of their conformation. As stated before, the β-heptapeptide is showing reversible folding, and an RMSD criterion was used to categorize conformations as folded or unfolded.[14] The same RMSD criterion has been used here, i.e., an atom-positional RMSD with respect to the NMR model structure of <0.1 nm counts as folded, >0.15 nm as unfolded. We note that the absolute numbers computed for the change in entropy, enthalpy, and free energy (from the previous two quantities and from counting of configurations) upon folding are dependent on the precise RMSD criteria used for the definition of the folded and unfolded states, but the conclusions drawn on the relative weights of entropy and enthalpy and on the comparison with the free energies from counting of configurations are not sensitive to these criteria.

Entropy Change Upon Folding

Using only folded configurations in the entropy calculation yields the configurational entropy of the folded state, the unfolded configurations yield the entropy of the unfolded state. Table IV gives the entropy of the folded and unfolded configurations (S_{fold} and S_{unfold}, respectively) for all four temperatures. The change in entropy upon folding $\Delta S_{\text{fu}} = S_{\text{fold}} - S_{\text{unfold}}$ is also given. The loss in configurational entropy upon folding of the peptide is substantial and is higher at the high temperatures. As already mentioned, the computed entropies exclude contributions from translation and rotation. This means, for example, that the contribution to ΔS_{fu} from the change in rotational entropy upon folding arising from the (possibly) different moments of inertia of the folded and unfolded conformations has been neglected.

The entropy calculation can again be performed on the two subsets of atoms constituting the backbone and the side-chains. The contribution of the backbone and side-chains to the change in configurational entropy upon folding is given in Table V. In absolute terms the change in
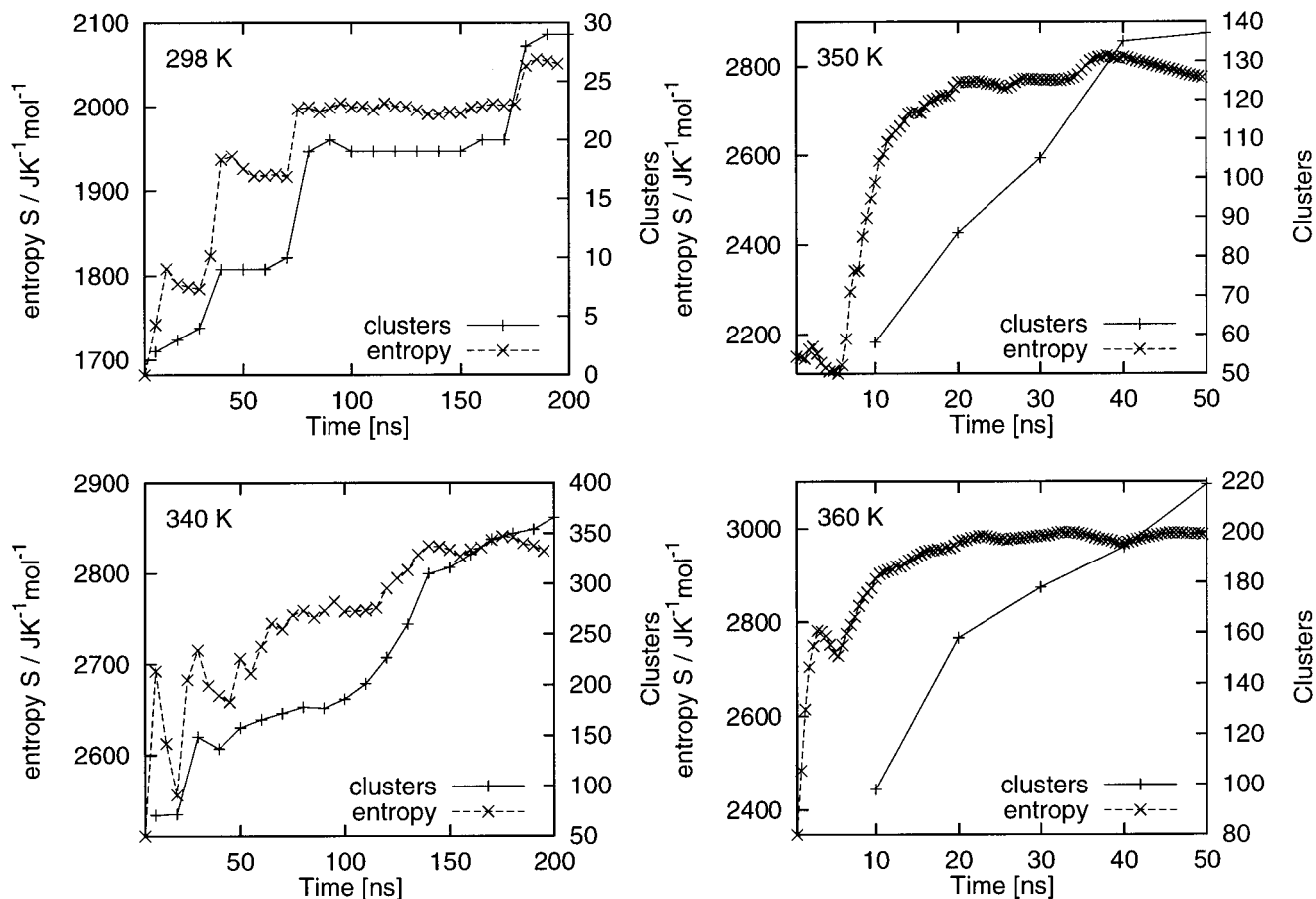
Fig. 8. Configurational entropy of the peptide S_{all} at 298 K, 340 K, 350 K, and 360 K compared with the number of conformational clusters sampled up to time t. Left axes show the entropy, right axes the number of clusters. A cluster of conformations is defined by an atom-positional root-mean-square deviation (RMSD) (residues 2–6, backbone atoms) of <0.1 nm between the central conformation and all other conformations of the cluster.

entropy in the backbone ΔS_{fu}^{bb} is clearly dominating. Normalized through division by the number of atoms (numbers within parentheses) the side-chains show a greater loss in entropy upon folding. Even though the peptide is very short, the folded helical conformation seems to restrict the mobility of the side-chains significantly.

Free Energy of Folding

The thermodynamic quantity that ultimately determines the folding of the β-heptapeptide is the free energy of folding ΔG_{fu}. Daura et al.[14,18] estimated ΔG_{fu} by calculating the ratio of folded to unfolded configurations by simple counting. Their results (ΔG_{fu}^{ratio}) are summarised in Table IV. Another route to the free energy uses the thermodynamic relation

$$\Delta G = \Delta H - T\Delta S \qquad (3)$$

where ΔH is the change in enthalpy. The enthalpy change upon folding ΔH_{fu} was obtained by averaging only intramolecular interactions of the peptide separately for the folded and the unfolded configurations. Because our entropy calculations were restricted to the configurational entropies of the peptide, and solvent entropy as well as peptide–

solvent correlations were neglected, the enthalpies were calculated accordingly. The enthalpies of the folded and unfolded configurations H_{fold} and H_{unfold}, respectively, are shown in Table IV. In general, the gain in intramolecular interaction energy upon folding increases with increasing temperature.

The free energies of folding ΔG_{fu} according to eq. (3) are also given in Table IV. The values range from 86 kJ mol^{-1} at 298 K up to 222 kJ mol^{-1} at 360 K. The large positive values for the free energy of folding would indicate that the folded state is highly improbable. Therefore, the β-heptapeptide would be unfolded at all four temperatures. This is obviously not true and implies that the role of the solvent is vital in the folding process.[21] ΔG_{fu} captures only the solute (peptide) part of the free energy; the solvent contribution is completely ignored. Thus, it represents the gain in free energy of the peptide upon folding, not of the peptide and its solvent environment. The peptide loses entropy upon folding (ΔS_{fu} is negative), which is insufficiently offset by a lowering of the enthalpy upon folding. This results in the positive free energy change. The solvent must therefore compensate the entropy loss of the peptide upon folding. Direct interactions between the solvent and

peptide cannot explain the difference. At 340 K, the potential energies of the interaction between the solvent and the peptide are virtually identical in the folded and unfolded configurations. Therefore, it is apparent that to compensate for the entropy loss of the peptide upon folding, there must be either an increase in entropy in the solvent, a loss in peptide–solvent correlations, or an increase in solvent–solvent interactions.

EXPLORATION OF PHASE SPACE

Ultimately the entropy of a system depends on the entire phase space. The stepwise increase in the configurational entropy was correlated to unfolding events (see Configurational Entropy of the β-Heptapeptide, above). The entropy shows some convergence as unfolding events become more frequent, yet a reliable confirmation of convergence cannot be given. Daura et al.[18] clustered the trajectories according to a RMSD criterion (atom-positional RMSD for the backbone atoms of residues 2–6 < 0.1 nm). By plotting the number of clusters explored up to time t, an attempt was made to obtain a rough idea about the amount of phase space sampled. The number of clusters containing one or more members at time t is compared with the configurational entropy at time t in Figure 8.

While at 298 K the number of clusters follows the stepwise increase in entropy very closely, at higher temperatures the number of clusters is consistently increasing, while the entropy levels off. The number of sampled clusters contains no information about their relative weight and the entropy sometimes decreases while the number of clusters increases. The probability of finding a new cluster with high relative weight decreases with increasing simulation time, and so it decreases the amount of configurational information contained in these new clusters.

CONCLUSIONS

The calculation of configurational entropies of small peptides is now feasible with the availability of trajectories of dozens of nanoseconds. Using Schlitter's formula,[5] problems concerning the mass-metric tensor can be avoided. The use of a least-squares fit to remove overall rotation of the peptide introduces some ambiguity that needs to be considered.[6]

The configurational entropy of the peptide generally shows the expected behavior as a function of temperature. The stepwise increase in the entropy as a function of time could be correlated with unfolding events of the peptide. Analyzing the entropy per residue shows an entropy difference between the centre and the ends of the peptide chain, which decreases with temperature. In the unfolded state, only a small difference between the residues remains because of the increased flexibility of the chain ends.

Comparing the entropy of the backbone with that of the side-chains, the latter contribute less entropy in absolute terms, but more per atom. The β-heptapeptide differs from proteins in that it contains fewer side-chain atoms than backbone atoms. In proteins the ratio is reversed. Normalizing the entropy through division by the number of atoms,

the side-chains contribute more entropy than the backbone does. Upon unfolding, the backbone contribution increases.

The correlation between the backbone and the side-chains decreases the entropy by around 17%. This percentage stays rather constant with simulation time and is rather insensitive to the temperature.

The difference in configurational entropy of the peptide between the folded and unfolded configurations is large. This results in a large free energy of folding for the peptide. The route to the free energy of folding via the configurational entropy of the peptide captures only the peptide part and ignores the change in the entropy and enthalpy of the solvent and the change in peptide–solvent correlation. This underlines the important role of the solvent in peptide folding. Using the GROMOS96 force field, the folding of the β-heptapeptide seems to be driven predominantly by entropy changes.

Entropy calculations clearly have an important role to play in understanding the underlying forces that drive peptide and protein folding. This study shows that consideration of the configurational entropy of the peptide alone is insufficient and that entropy calculations including solvent degrees of freedom are called for. In future work we will examine the influence of different solvents, different stable folds, and different chemical structure of non-natural peptides.

REFERENCES

1. Gō N, Scheraga HA. On the use of classical statistical mechanics in the treatment of polymer chain conformation. Macromolecules 1976;9:535–542.
2. Karplus M, Kushick JN. Method for estimating the configurational entropy of macromolecules. Macromolecules 1981;14:325–332.
3. Edholm O, Berendsen HJC. Entropy estimation from simulations of nondiffusive systems. Mol Phys 1984;51:1011–1028.
4. Di Nola A, Berendsen HJC, Edholm O. Free energy determination of polypeptide conformations generated by molecular dynamics. Macromolecules 1984;17:2044–2050.
5. Schlitter J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. Chem Phys Lett 1993;215:617–621.
6. Schäfer H, Mark AE, van Gunsteren WF. Absolute entropies from molecular dynamics simulation trajectories. J Chem Phys 2000; 113:7809–7817.
7. Dill KA. Dominant forces in protein folding. Biochemistry 1990;29: 7133–7155.
8. Bicout DJ, Szabo A. Entropic barriers, transition states, funnels, and exponential protein folding kinetics: a simple model. Protein Sci 2000;9:452–465.
9. Galzitskaya OV, Surin AK, Nakamura H. Optimal region of average side-chain entropy for fast protein folding. Protein Sci 2000;9:580–586.
10. Wrabl JO, Shortle D, Woolf TB. Correlation between changes in nuclear magnetic resonance order parameters and conformational entropy: molecular dynamics simulations of native and denatured staphylococcal nuclease. Proteins 2000;38:123–133.
11. Brady GP, Sharp KA. Entropy in protein folding and in protein–protein interactions. Curr Opin Struct Biol 1997;7:215–221.
12. Pickett SD, Sternberg MJE. Empirical scale of side-chain conformational entropy in protein folding. J Mol Biol 1993;231:825–839.
13. Yang AS, Honig B. Free energy determinants of secondary structure formation: I. α-helices. J Mol Biol 1995;252:351–365.
14. Daura X, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Reversible peptide folding in solution by molecular dynamics simulation. J Mol Biol 1998;280:925–932.
15. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH,

Krüger P, Mark AE, Scott WRP, Tironi IG. Biomolecular simulation: the GROMOS96 manual and user guide. Zürich, Switzerland: Vdf Hochschulverlag; 1996.

16. Scott WRP, Hünenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Krüger P, van Gunsteren WF. The GROMOS biomolecular simulation program package. J Phys Chem A 1999;103:3596–3607.

17. Seebach D, Ciceri PE, Overhand M, Jaun B, Rigo D, Oberer L, Hommel U, Amstutz R, Widmer H. Probing the helical secondary structure of short-chain β-peptides. Helv Chim Acta 1996;79:2043–2066.

18. Daura X, van Gunsteren WF, Mark AE. Folding-unfolding thermodynamics of a β-heptapeptide from equilibrium simulations. Proteins 1999;34:269–280.

19. Mark AE, van Gunsteren WF. Decomposition of the free energy of a system in terms of specific interactions. J Mol Biol 1994;240:167–176.

20. Smith PE, van Gunsteren WF. When are free energy components meaningful? J Phys Chem 1994;98:13735–13740.

21. Daura X, Mark AE, van Gunsteren WF. Peptide folding simulations: no solvent required? Comp Phys Commun 1999;123:97–102.

APPENDIX: ABSOLUTE ENTROPY IN TERMS OF THE COVARIANCE MATRIX OF ATOM-POSITIONAL FLUCTUATIONS

The formula for the entropy introduced by J. Schlitter[5] is based on the approximation that each degree of freedom can be represented by a quantum-mechanical harmonic oscillator. The frequency ω of each of these harmonic oscillators is determined by its classical variance $\langle x^2 \rangle_c$ through the equipartition theorem:

$$m\omega^2 \langle x^2 \rangle_c = k_B T \qquad (A1)$$

where m is the mass of the degree of freedom, k_B is Boltzmann's constant, T is the temperature and $\langle \dots \rangle$ indicated ensemble or trajectory averaging. Schlitter then proposes the following heuristic formula for the entropy of 1 degree of freedom:

$$S \leq S_{ho} < S' \equiv \frac{1}{2} k_B \ln\left[1 + \frac{k_B T e^2}{\hbar^2} \langle x^2 \rangle_c \right] \qquad (A2)$$

where e is Euler's number and $\hbar$ is Planck's constant divided by 2π.

The quantity S' as defined in eq. (A2) has a number of desired properties[5]:

1. It is an upper bound to the entropy S_{ho} of a quantum harmonic oscillator, and to the exact entropy.
2. The entropy S' approaches zero when the temperature goes to zero, which is the correct quantum-mechanical limit.
3. The entropy S' is proportional to $\ln T$ for T approaching infinity, which is the correct classical-mechanical limit.
4. Formula (A2) is easily generalizable to many degrees of freedom:

$$S < S' = \frac{1}{2} k_B \ln \det\left[\mathbf{1} + \frac{k_B T e^2}{\hbar^2} \mathbf{M}\boldsymbol{\sigma} \right] \qquad (A3)$$

in which $\mathbf{M}$ is the mass matrix, which holds the masses belonging to the degrees of freedom on the diagonal and has zero off-diagonal elements, and $\boldsymbol{\sigma}$ is the covariance matrix of atom-positional fluctuations

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \qquad (A4)$$

The cartesian coordinates of the atoms are denoted by $x_1 \dots x_{3N}$.

5. Schlitter's entropy formula has no singularity if the matrix $\boldsymbol{\sigma}$ is singular.

For a more detailed rationalization and test of formula (A3), we refer to refs. 5, 6.