Entropy Calculations on the Molten Globule State of a Protein: Side-Chain Entropies of α-Lactalbumin

Heiko Schäfer,[1] Lorna J. Smith,[2] Alan E. Mark,[3] and Wilfred F. van Gunsteren[1]*
[1]*Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH Zentrum, Zurich, Switzerland*
[2]*Oxford Centre for Molecular Sciences, New Chemistry Laboratory, University of Oxford, Oxford, England*
[3]*Laboratory of Biophysical Chemistry, University of Groningen, Groningen, The Netherlands*

ABSTRACT **We present entropy estimates based on molecular dynamics simulations of models of the molten globule state of the protein α-lactalbumin at low pH. The entropy calculations use the covariance matrix of atom-positional fluctuations and yield the complete configurational entropy. The configurational entropy of the entire protein and of each of its side chains is calculated. Exposed side chains show a larger entropy compared to buried side chains. A comparison to data from rotamer counting is made and significant differences are found. Proteins 2002;46:215–224.**

INTRODUCTION

Apart from the native state of a protein and the random coil, the molten globule state has been the focus of much interest.[1–3] The molten globule state is compact and retains a great part of secondary structure while the side chains are disordered. It is believed that the examination of the molten globule might give new insights into protein folding pathways.[2]

Recently Smith et al.[4] tried to model the molten globule state of human α-lactalbumin at low pH using molecular dynamics (MD) simulations. The interconversion between the native state of α-lactalbumin and the molten globule is estimated to be in the order of seconds, which is beyond the time scale accessible by MD simulations. The molten globule state was, therefore, modelled by simulating different conformers of the molten globule. These conformers were generated by disordering the side chains in four different ways.

In the present work, the configurational entropy of two of the conformers of the molten globule are estimated and compared. A formula for the estimation of the entropy introduced by Schlitter,[5] which was recently tested and applied to MD trajectories of a β-heptapeptide in methanol,[6,7] is used.

METHODS

The system considered in this work was previously described in detail by Smith et al.[4,8] It consists of 4 different simulations of α-lactalbumin in water at pH = 2, under which conditions α-lactalbumin is known experimentally to form a molten globule state.[3] In order to reproduce the low pH conditions, the calcium ion was removed and the aspartate, glutamate, and histidine side chains and the carboxy terminus were protonated. Starting from the crystal structure of human α-lactalbumin determined at pH = 6.5,[9] the protein and 5,582 water molecules were equilibrated for 100 ps at pH = 2. At this point, the simulations labeled R1 and R4 in Smith et al.,[4] which differ in their treatment of the side chains, were branched off. In simulation R1, the side chain positions were left unchanged. In R4, all χ_1 torsion angles were changed; for residues with a χ_1 torsion angle of $-60°$ ($\pm 30°$), the χ_1 angle was set to 60°, while for all other residues χ_1 was set to $-60°$. The torsion angles of cysteine and proline residues were excluded, as were those of Phe-31, Trp-60, and His-107 as rotation of the side chains of these residues resulted in the side chains of Lys-5, Ser-56, and Trp-104, respectively, going through the aromatic ring concerned.[4] Both systems simulations were then extended for 2 ns. The R1 and R4 simulations represent two possible conformers of the molten globule state of α-lactalbumin, not the molten globule itself.

In order to estimate the side-chain configurational entropy S_{cf} of α-lactalbumin, a heuristic formula introduced by Schlitter[5] is used. Schlitter's formula approximates the absolute entropy S,

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

where k_B is Boltzmann's constant, T is the absolute temperature, e Euler's number, $\hbar$ is Planck's constant divided by 2π, M is the mass matrix that holds the masses belonging to the atomic Cartesian degrees of freedom of the molecule or part of the molecule (side chain) considered on the diagonal and has all off-diagonal elements zero, and σ is the covariance matrix of atom-positional fluctuations with the elements

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle. \qquad (2)$$

Here x_i are the Cartesian coordinates. The entropy S' calculated using eq. 1 was shown by Schlitter to be an

Grant sponsor: Schweizerischer Nationalfond; Grant number: 21-50929.97.

*Correspondence to: Wilfred F. van Gunsteren, Laboratory of Physical Chemistry, Swiss Federal Institute of Technology Zurich, ETH Zentrum, Universitatstrasse 6, CH-8092 Zurich, Switzerland. E-mail: wfvgn@igc.phys.chem.ethz.ch

Received 28 August 2000; Accepted 10 July 2001

DOI 10.1002/prot.1166

upper limit to the correct entropy S. Schlitter's formula was extensively tested[5,6] and was shown to give results with errors around 5% for condensed phase systems.

In order to estimate entropies from MD trajectories using eq. 1, it is necessary to calculate the covariance matrix σ. This is done directly from the atom Cartesian coordinates x_i and thus includes the overall rotation and translation of the protein. On the time scales reachable by MD simulations of proteins in solution, one cannot hope for sufficient sampling of the overall rotation and translation of the protein. In order to remove these slowly converging degrees of freedom from the calculations, a geometrical (translational and rotational) least-squares fit was used. This implies that the configurational entropy, instead of the absolute entropy, is calculated. Different sets of atoms were used in the least-squares fit: (1) all main chain backbone atoms (N, C, C_α) denoted the symbol *bb*; (2) main chain backbone atoms only in the three helical regions of α-lactalbumin (residues 5–11, 24–33, and 86–98) that are stable in the R1 and R4 simulations, denoted by the symbol *2nd*; (3) when the entropy per residue is calculated an individual fit on the N, C, C_α of each residue denoted by the symbol *fir* is performed.

The MD simulations for α-lactalbumin explicitly included water as the solvent.[4,8] In order to calculate the configurational entropy of the protein, the configurations are fitted using a least-squares fit and *only* the coordinates of the protein atoms are used in the calculation of the covariance matrix σ. The entropy of the solvent and the correlation between the solvent and protein degrees of freedom are thus ignored.

Generally an arbitrary set of atom coordinates can be used in the calculation of σ. However, when using subsets of the degrees of freedom of a system in the entropy calculation, one has to be aware that the entropy is, in principle, a nonadditive quantity. By decomposing the total entropy into parts, one is ignoring the correlation between the parts. As an example, for the calculation of the entropy per residue (see below) a separate covariance matrix for each residue is used. Any correlation between the residues is, therefore, ignored and the entropies per residue do not add up to the configurational entropy of the entire protein. The sum of the configurational entropies of all residues is larger than the total configurational entropy of the protein, because of the neglect of correlations that lower the entropy.

The off-diagonal elements of the covariance matrix σ stem from the correlation between the degrees of freedom considered in the entropy calculation. By using only the diagonal elements of σ, it is possible to ignore all correlations and thereby calculate the uncorrelated entropy. The difference between the uncorrelated entropy and the entropy calculated using all elements of σ, gives the decrease in entropy because of correlation between the degrees of freedom used in the calculation.

RESULTS

Configurational Entropy of α-Lactalbumin

The configurational entropy of the protein was calculated using eq. 1. Each configuration was geometrically fitted to the starting configuration of both the R1 and R4 simulations using a least-squares fit. Two different sets of atoms were used in the least-squares fit, as described above. The results are summarised in Table I and the configurational entropy as a function of simulation time is shown in Figure 1.

TABLE I. Configurational Entropy S_{cf} of α-lactalbumin at pH = 2 Calculated From the 2-ns R1 Simulation, Which Started From the Native Structure, and From the 2-ns R4 Simulation, Which Started With Almost All Side-Chain χ_1 Angles Changed by 120°†

	R1 (low pH)	R4 (maximal disorder)	Δ (R4 − R1)
Configurational entropy			
S_{cf}^{bb} (backbone fit)	27,545	28,270	725
(Uncorrelated)	80,660	83,356	2,696
S_{cf}^{2nd} (2nd struct. fit)	27,603	28,319	716
(Uncorrelated)	81,544	84,926	3,382
Backbone entropy			
S_{bb}^{bb} (backbone fit)	11,719	12,165	446
(Uncorrelated)	36,361	37,221	860
S_{bb}^{2nd} (2nd struct. fit)	11,764	12,206	442
(Uncorrelated)	36,886	38,075	1,189
Side-chain entropy			
S_{sc}^{bb} (backbone fit)	19,265	19,691	427
(Uncorrelated)	44,300	46,250	1,950
S_{sc}^{2nd} (2nd struct. fit)	19,294	19,720	426
(Uncorrelated)	44,724	46,876	2,152
Backbone/side chain correlation			
$\Delta S_{bb/sc}^{bb}$	3,439	3,586	147
$\Delta S_{bb/sc}^{2nd}$	3,455	3,607	152

†Two sets of atoms were used in the least-squares fit: all main-chain backbone atoms (^{bb}) and only main-chain atoms inside secondary structure (^{2nd}). The results of ignoring all correlations are indicated by *uncorrelated*. The configurational entropy of the backbone S_{bb}, side chains S_{sc}, and the decrease in entropy $\Delta S_{bb/sc}$ because of the correlation between backbone and side chains are also given. All values are in J K^{-1} mol^{-1}.

The effect of using different sets of atoms in the least-squares fit is minimal. Using secondary structure elements (less atoms in the fit) gives a slightly larger entropy than using all backbone atoms. The configurational entropy is certainly not converged after 2 ns and is still increasing at this time point. As could be expected, the R4 simulation where the side-chains were maximally disordered, has a greater entropy compared to the R1 simulation, where the side-chain conformations were not changed from native.

The atoms in the protein were split into two subgroups, the backbone atoms and the side-chain atoms. For these subsets the entropies were calculated using the same procedure as described above. Table I shows the results and Figure 2 the convergence of the entropies. The configurational entropy of the backbone converges more rapidly than that of the side chains. The ordering between the R1 and R4 simulations, i.e., R4 showing a larger entropy, is retained in the backbone and the side chain results. The side chains have a greater configurational entropy than

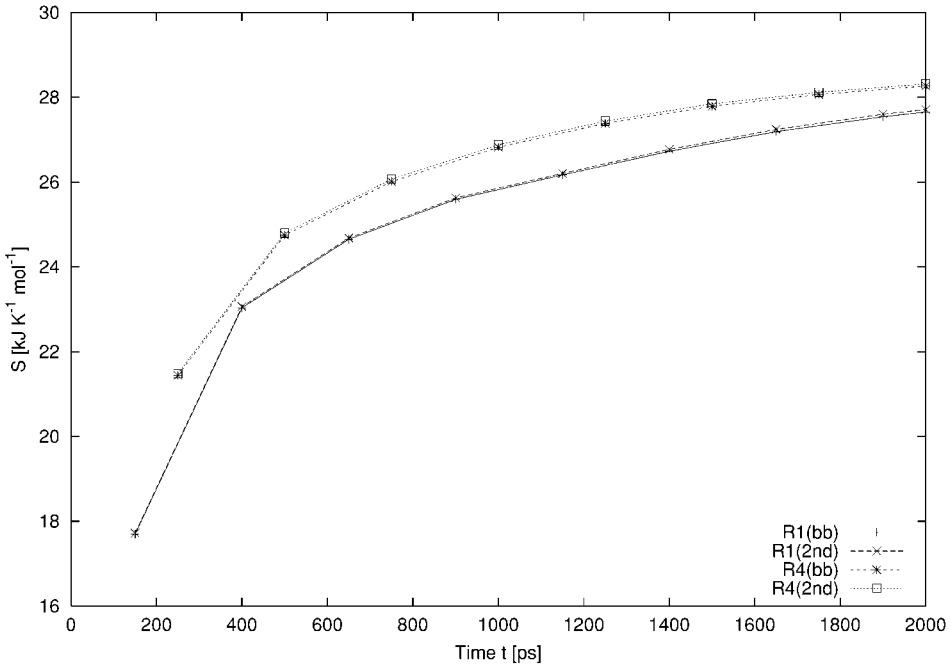

Fig. 1. Configurational entropy of the simulation R1, which started from the native structure, and of the simulation R4, which has maximally disordered side-chain conformations. Two different sets of atoms were used in the least-squares fit, backbone atoms (*bb*) and backbone atoms in regions of secondary structure (2nd).

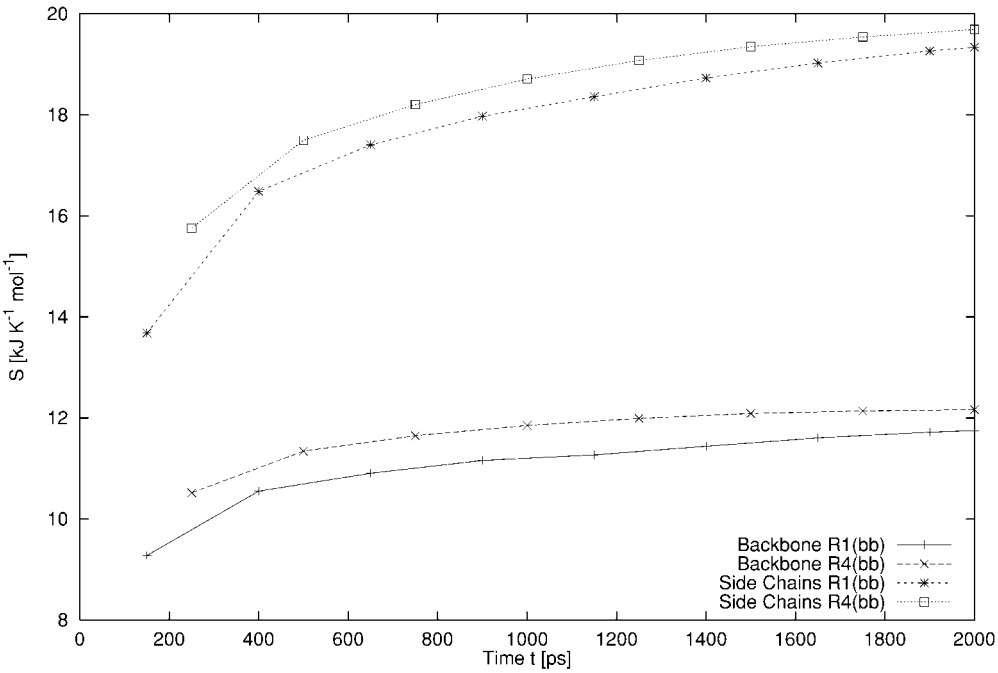

Fig. 2. Configurational entropy of the backbone and side-chain subsets for the R1 and R4 simulations. A least-squares fit on the backbone atoms was used to remove overall translation and rotation of the protein.

the backbone. The effect of using different sets of atoms in the least-squares fit is again minimal (see Table I).

Using only the diagonal elements of the covariance matrix, correlations between the degrees of freedom are ignored and a much larger value of the entropy is obtained (see Table I). The difference between the value obtained using the diagonal elements only and using the full covariance is due to the correlation between the atoms. Considering first the configurational entropy of the entire protein, it can be seen that the uncorrelated value is

TABLE II. Distribution of the Entropy Amongst the Backbone, Amongst the Side Chains, and the Correlation Between the Two

	R1 simulation	R4 simulation
Percentage of the correlation $\frac{\Delta S^{bb}_{bb/sc}}{S^{bb}_{cf}}$	13	13
Percentage of the entropy in the backbone $\frac{S^{bb}_{bb}}{S^{bb}_{bb}+S^{bb}_{sc}}\frac{S^{bb}_{cf}-\Delta S^{bb}_{bb/sc}}{S^{bb}_{cf}}$	33	33
Percentage of the entropy in the side chains $\frac{S^{bb}_{sc}}{S^{bb}_{bb}+S^{bb}_{sc}}\frac{S^{bb}_{cf}-\Delta S^{bb}_{bb/sc}}{S^{bb}_{cf}}$	54	54

roughly three times larger than the configurational entropy including all correlations. The same is true for the backbone entropy (slightly above three times). The side chains, however, show a smaller decrease of the entropy due to correlation; the uncorrelated value is only 2.3 times the configurational entropy.

The correlation within the backbone, within the side chains and within the entire protein has been discussed. The correlation between the backbone and the side chains is also of interest. This can be obtained by taking the difference between the sum of the backbone and the side chain entropies $S_{bb} + S_{sc}$, and the configurational entropy of the entire protein S_{cf}. While S_{cf} contains all correlations between the backbone and the side chain atoms, the values calculated using only either subsets do not. The difference therefore gives the decrease in entropy due to these correlations:

$$\Delta S_{bb/sc} = S_{bb} + S_{sc} - S_{cf} \quad (3)$$

Values for $\Delta S_{bb/sc}$(R1) and $\Delta S_{bb/sc}$(R4) are given in Table I.

The distribution of the entropy amongst the backbone and the side chains, and the decrease in entropy due to correlation are presented in Table II. The side chains dominate with over 50% of the entropy. The decrease in entropy because of the correlations between the backbone and the side chains is not negligible at over 10%. This distribution can be compared to the results from a similar calculation[7] on a β-heptapeptide in solution, where the correlation amounted to 17%, the backbone contribution to 28%, and the side-chain contribution to 55%. Even though the two systems considered differ widely, the distribution of the entropy over backbone and side chains is strikingly similar.

Entropy Per Residue

By using only the atoms of a particular residue in the calculation of the covariance matrix σ and using eq. 1, the entropy of a single residue can be estimated. This is of particular interest as there exists a large body of literature concerning side-chain entropies and their role in protein folding.[10–19]

Again, a least-squares fit is used to remove overall rotation and translation of the protein. Two fitting procedures were used: (1) a fit using all main chain backbone atoms and (2) a fit using only the main chain backbone atoms of each residue. The first procedure suffers from the fact that backbone motion will contribute to the entropy of a residue. The second procedure, in which each residue is fitted individually using its backbone atoms N, C, and C_α, is therefore preferred.

The entropy per residue S^{fir}_{res}, using the individual residue fits, as a function of residue number is shown in Figure 3. Figure 3 (top) shows the results for the R4 simulation and Figure 3 (bottom) the results for the R1 simulation. The entropies have been normalised by division through the number of atoms in the residue to make the comparison easier. Even though the helical regions, which are stable in both simulations (residues 5–11, 24–33, and 86–98), have a rather lower entropy, there is no pronounced correlation between the entropy per atom of a residue and its position in the protein structure.

The convergence of the entropy per residue has been monitored for all residues. An illustration of the difference in time evolution of the entropy of the residues is given in Figure 4. The residues shown are Ala-22, Arg-70, Gln-54, Ile-21, and Lys-108. The behaviour ranges from a low constant entropy in the case of alanine, to the rapidly converging high entropy of arginine (and lysine), to the probably still not converged entropy of glutamine and isoleucine. The residues with long side chains show a stepwise increase in their entropy very similar to behaviour observed in our previous study of a folding β-heptapeptide.[7] Each of these steps corresponds to the sampling of a new region of configuration space. The convergence behaviour depends strongly on the degrees of freedom of the residue. Alanine, for example, will converge faster than the residues with longer side chains. There seems to be also a correlation with the solvent accessibility of a residue: the solvent accessibility as calculated using the NACCESS program[20] of the sample residues of Figure 4 are Ala-22 34%, Arg-70 79%, Gln-54 2%, Ile-21 11%, and Lys-108 68%. The more exposed residues appear to converge faster than the buried residues.

The only difference between the R1 and R4 simulations is the treatment of the side chains at the start of the simulations: while they were left in the native conformation for the R1 simulation, the χ_1 angle was maximally disordered in the R4 simulation. The difference in the entropy of the residues between the R4 and the R1 simulations is shown in Figure 5. The results shown by a dashed line in Figure 5 correspond to a fit on all backbone atoms, while for the solid line individual residue fits were used. They differ most in the region around residue 100 where there is a larger movement of the backbone in the R4 simulation. The values calculated using the individual residue fit are, therefore, used for further analysis. Surprisingly, there is little difference between the R1 and R4 simulations: around residue 105, the R4 simulation shows a larger entropy, while around residue 20 this trend is reversed.

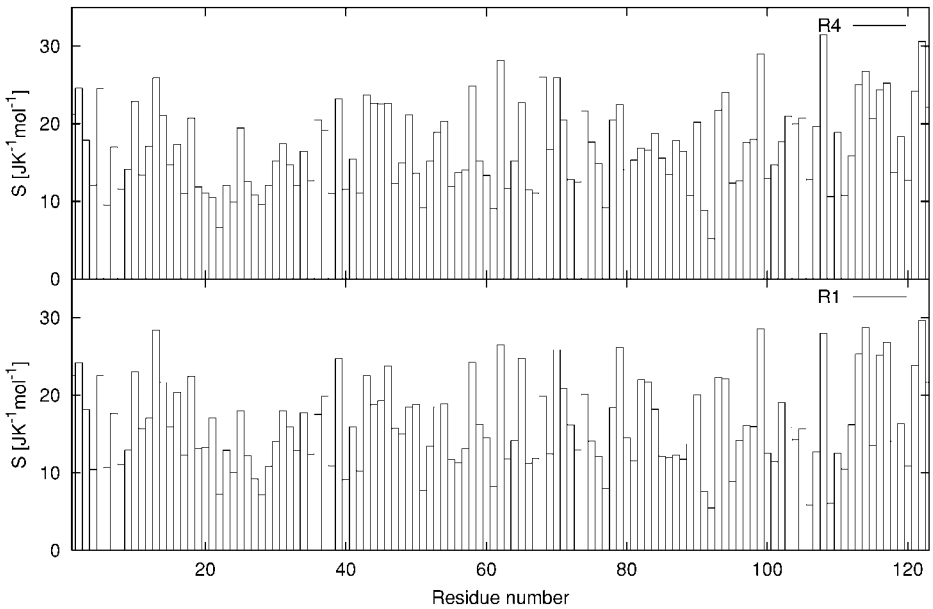

Fig. 3. Entropy per residue after an individual least-squares fit for each residue for the R1 simulation (S^{fir}_{res}(R1)) and the R4 simulation (S^{fir}_{res}(R4)). The entropy has been divided by the number of atoms in each residue.

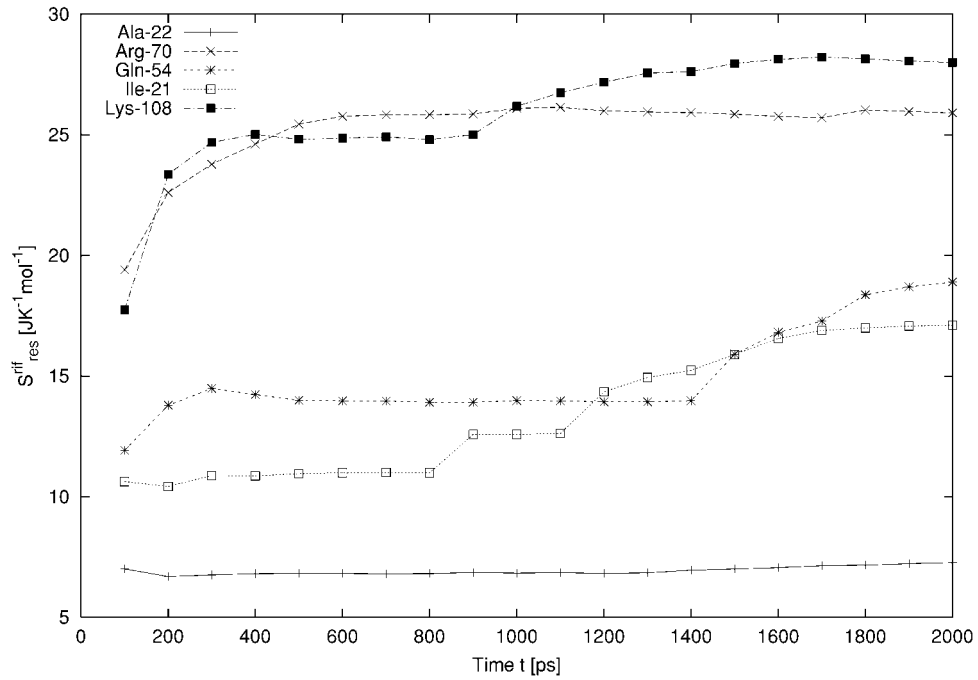

Fig. 4. Convergence of the configurational entropy per atom of some sample residues from the R1 simulation using a fit on individual residues. Shown are residues Ile-21, Ala-22, Gln-54, Arg-70, and Lys-108.

The residues were sorted according to their type and the entropy per residue (normalised by the number of atoms in the residue) was averaged over all residues of that specific type. The results are summarised in Table III. The normalised entropy per residue was also averaged over all residues and the results are given in the row labeled "All" in Table III. The spread in the values is large with entropies per atom ranging from 7 up to 26 J K^{-1} mol^{-1}. The averages are of different statistical quality as the number of residues from a specific type varies from 14 leucines to a single arginine.

As expected, residues with a short side chain, without side chain, or with a conformationally restricted side chain (Ala, Gly, Cys, Pro) have a low entropy while the amino acids with the largest entropies are Arg, Lys, Gln, and Glu, all of which have long side chains and are polar. The

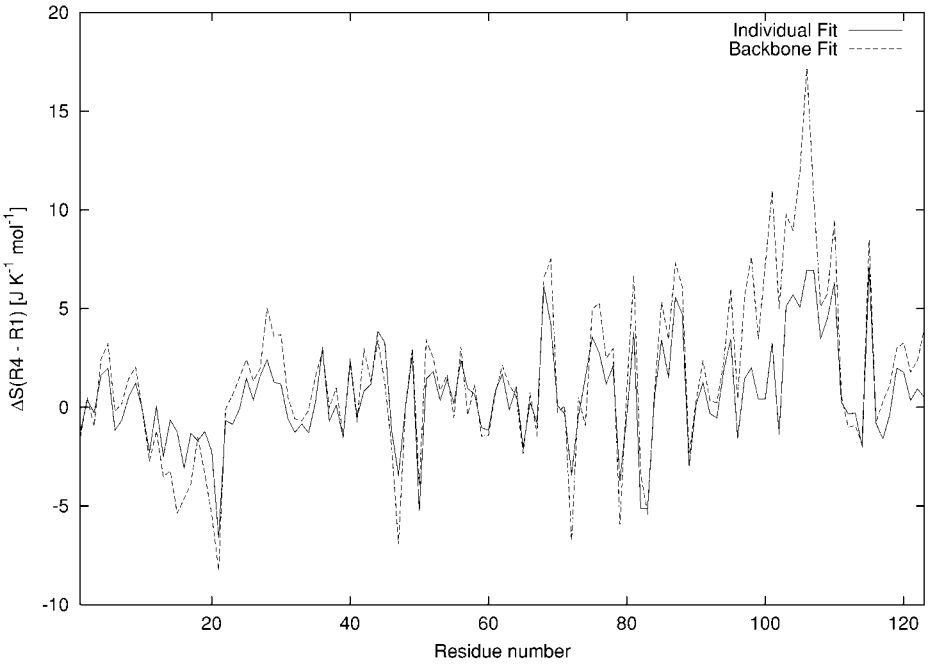

Fig. 5. The difference in entropy after an individual fit for each residue and after a backbone fit between the R4 and the R1 simulations: $\Delta S(\mathrm{R4} - \mathrm{R1}) = S_{res}^{fir}(\mathrm{R4}) - S_{res}^{fir}(\mathrm{R1})$.

TABLE III. Entropy S_{res}^{fir} Per Residue Type After an Individual Residue Fit[†]

Amino acid	Number of residues	R1 sim. S_{res}^{fir} ($\mathrm{J\,K^{-1}\,mol^{-1}}$)	R4 sim. S_{res}^{fir} ($\mathrm{J\,K^{-1}\,mol^{-1}}$)	Amino acid	Number of residues	R1 sim. S_{res}^{fir} ($\mathrm{J\,K^{-1}\,mol^{-1}}$)	R4 sim. S_{res}^{fir} ($\mathrm{J\,K^{-1}\,mol^{-1}}$)
Ala	5	6.8 ± 1.3	9.4 ± 2.9	Leu	14	14.6 ± 2.7	16.2 ± 3.4
Arg	1	25.9 ± 0.0	26.0 ± 0.0	Lys	12	25.8 ± 2.8	25.9 ± 3.2
Asn	3	17.6 ± 3.3	19.1 ± 3.7	Met	2	17.1 ± 3.0	17.7 ± 2.5
Asp	13	18.6 ± 3.2	18.8 ± 2.0	Phe	4	17.3 ± 1.6	17.1 ± 1.8
Cys	8	9.5 ± 1.9	10.3 ± 1.5	Pro	2	10.9 ± 0.9	10.5 ± 0.6
Gln	7	23.2 ± 2.6	23.6 ± 1.8	Ser	7	13.8 ± 2.1	14.3 ± 1.8
Glu	8	21.9 ± 3.1	22.2 ± 2.6	Thr	8	12.6 ± 1.9	13.3 ± 1.6
Gly	6	11.9 ± 1.9	11.5 ± 1.3	Trp	3	14.3 ± 0.2	15.7 ± 3.0
His	2	14.4 ± 1.6	17.2 ± 2.5	Tyr	4	18.7 ± 2.4	19.0 ± 3.1
Ile	12	13.5 ± 2.7	13.8 ± 2.5	Val	2	10.8 ± 0.5	11.3 ± 0.2
				All	123	16.3 ± 5.7	16.9 ± 5.5

[†]Results for the R1 and R4 simulations are shown.

hydrophobic amino acids Ala, Gly, Ile, Leu, and Val are all in the lower range of entropy values. The same is true for the cysteine residues that all form disulfide bridges.

The two simulations R1 and R4 show no significant difference in the entropy per residue. The most striking difference is observed for alanine where the starting structure was the same for R1 and R4 (there is no χ_1 angle in alanine). The difference can be traced back to two alanines (Ala-106, Ala-109) that show a markedly higher entropy in the R4 simulation. In the region of residues 100 to 110, the backbone shows a large motion in the R4 simulation and the higher entropy of the two alanines can be explained by larger backbone fluctuations.

As already pointed out above, polar residues generally show a larger entropy. The residues were, therefore, ordered according to their solvent accessibility calculated using the **NACCESS** program.[20] Residues were assigned to 10 bins ranging from 5 to 95% accessibility and the average entropy for these residues was calculated. The results for the R1 and R4 simulations are shown in Table IV. Again, the statistical accuracy differs widely with only three residues in the class of 75% accessibility while 36 residues fall into the 5% class.

The entropy generally increases with increasing solvent accessibility. This can also be seen in Figure 6 where α-lactalbumin is shown with the residues coloured according to their entropy: the scale ranges from blue for low entropy values to red for high entropies. Obviously, residues on the outside of the protein are higher in entropy than buried residues. From Table IV it can be seen that there is a peak in the entropy per residue at an accessibility of 75%. The three residues in this class are Glu-43, Arg-70, and Lys-122. The latter two are the top-ranking residues in entropy terms and the peak may be an effect of

TABLE IV. Entropy S_{res}^{fir} Per Residue as a Function of Solvent Accessibility After an Individual Residue Fit†

Accessibility (%)	Number of residues	R1 sim. S_{res}^{fir} (J K^{-1} mol^{-1})	R4 sim. S_{res}^{fir} (J K^{-1} mol^{-1})
5.0	36	12.5 ± 3.4	13.6 ± 3.6
15.0	15	13.3 ± 3.5	14.1 ± 2.9
25.0	11	15.9 ± 4.1	17.7 ± 4.4
35.0	12	17.4 ± 5.4	16.2 ± 4.9
45.0	6	17.1 ± 3.9	16.5 ± 3.6
55.0	15	19.0 ± 6.2	19.7 ± 5.2
65.0	13	22.2 ± 5.6	22.2 ± 5.3
75.0	3	26.1 ± 2.9	26.8 ± 2.9
85.0	9	17.9 ± 4.3	19.1 ± 5.7
95.0	3	19.2 ± 3.5	19.5 ± 3.1

†Results for the R1 and R4 simulations are shown.

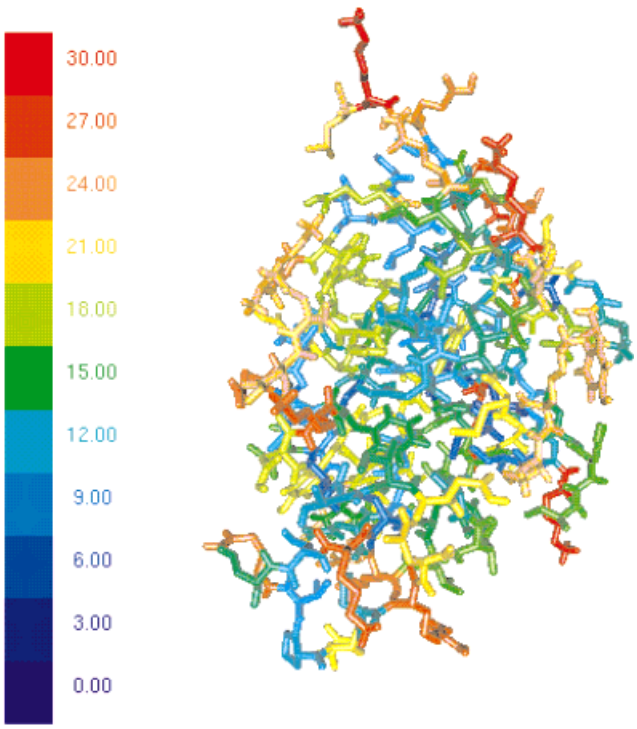

Fig. 6. The starting structure of α-lactalbumin of both, the R1 and R4 simulations. The residues are coloured according to their entropy S_{res}^{fir}, which was calculated using an individual fit for each residue and divided through the number of atoms in the residue. The colour scale ranges from low entropy values (blue) to high values (red). Entropy values are given in J K^{-1} mol^{-1}.

the poor statistics for residues with high solvent accessibility. Again, there is hardly any difference between the R1 and R4 simulations.

Finally, the residues were split into two groups, those within the stable helices (residues 5–11, 24–33, and 86–98) and the remaining ones. The entropy per residue was averaged over these two groups and the results can be found in Table V. As expected, the residues inside the helices have a lower average entropy per residue compared to the rest of the residues. There is again no significant difference between the R1 and R4 simulations.

TABLE V. Average Entropy Per Residue (Normalised by division through the number of atoms) for Residues Inside the Three Helices (Residues 5–11, 24–33, and 86–98) and the Remaining Residues

	$\langle S_{res}^{fir}(\mathrm{R1})\rangle$ (J K^{-1} mol^{-1})	$\langle S_{res}^{fir}(\mathrm{R4})\rangle$ (J K^{-1} mol^{-1})
Inside helices	14.2 ± 4.7	14.9 ± 4.7
Outside helices	17.0 ± 5.8	17.6 ± 5.5

DISCUSSION

The configurational entropy for the entire protein and for the backbone and side chains is not yet converged after 2 ns (Figs. 1 and 2). The configurational entropy of the R4 simulation, where the χ_1 angles of the side chains were maximally disordered at the beginning of the simulation, is indeed larger than the entropy of the R1 simulation. Yet, the increase in entropy cannot be traced to the side chains. The entropy is spread quickly throughout the entire protein and can be found in the backbone as well as the side chains. This view is strengthened when looking at the entropy per residue: again, there are no significant differences between the R1 and R4 simulations. Looking closer at Figures 1 and 2, it becomes clear that the difference between the R4 and the R1 simulation decreases with time. It can be expected that both simulations will eventually converge to the same entropy. They both represent possible conformers of the molten globule state. The only difference between the two simulations is that, due to the reassignment of the χ_1 angles, R4 initially samples a part of configuration space more quickly than R1. This also fits with the observation of Smith et al.[4] that the fluctuations of the χ_1 angle are only larger for the R4 simulation in the first 100 ps.

COMPARISON TO SIDE-CHAIN ENTROPIES FROM ROTAMER COUNTING

The loss of conformational entropy of the side chains upon folding is important for the folding process.[10,21] Therefore, many attempts have been made to estimate the conformational entropy of the side chains in different environments. One assumption in most of these calculations is the distinction between conformational and configurational entropy. A side chain is assumed to stay in a conformational minimum for some time, undergoing only vibrations. This gives rise to the vibrational contribution to the entropy S_{vib}. When the side chain jumps into another conformation (another rotamer), the conformational entropy S_{conform} increases. A common assumption is that both contributions are *additive* in which case their correlation is ignored.[12] We are not aware of any evidence to justify this approximation. The vibrational entropy S_{vib} is commonly assumed to dominate the side-chain entropy. Another common assumption is that S_{vib} does not significantly change between the folded and unfolded states of a protein.[21] In the discussion of side-chain entropy changes upon folding it is, therefore, often ignored.

The conformational entropy S_{conform} is often calculated using rotamer counting in databases of protein structures.

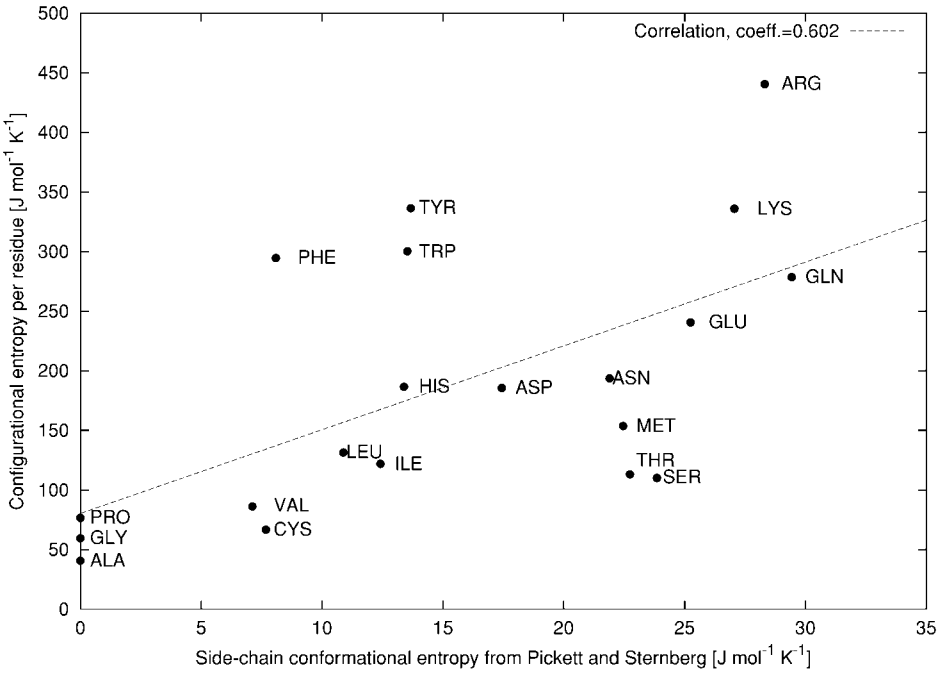

Fig. 7. Correlation between the side-chain conformational entropy obtained by rotamer counting using a database of 50 different protein structures (Pickett and Sternberg[11]) and the configurational entropy per residue from the R1 simulation.

One of the original studies, that of Pickett and Sternberg,[11] reports conformational entropy changes upon folding for all 21 amino acids obtained from rotamer counting. They assume that in the folded state of a protein, the side chains stay in a single conformation and that their conformational entropy in the folded state is thus zero. Their numbers can, therefore, also be interpreted as absolute conformational entropies. More recent studies, such as that of Creamer,[18] are based on slightly different assumptions but yield similar results.

The comparison of the conformational entropy of folding per residue to the configurational entropies per residue presented in Table III is not straightforward:

- Eq. 1 using a least-squares fit yields the configurational entropy, i.e., the vibrational part, the conformational part, and the correlation between both.
- In our work, a model of the molten globule and not the folding process is investigated. S_{res}^{fir} is the configurational entropy of the residues in two possible conformers of the molten globule.
- Only α-lactalbumin is considered in this work, while Pickett and Sternberg,[11] for example, scan a database consisting of 50 different protein structures.

Nevertheless, a comparison between both sets of entropies obtained by completely different approaches is of some interest. Figure 7 shows a correlation plot between the data of Pickett and Sternberg[11] and the results from the R1 simulation (Table III). For the R1 simulation, the entropy per residue *not* divided by the number of atoms is taken. A linear regression yields a poor correlation (correlation coefficient of 0.6). The configurational entropy from the 2 ns R1 simulation is an order of magnitude larger than the conformational entropy of Pickett and Sternberg.[11] This confirms the dominating role of the vibrational part of the entropy.[21] There are marked differences in the ranking of amino acids in both methods. The rotamer counting procedure of Pickett and Sternberg[11] ranks the hydroxyl bearing amino acids serine and threonine much higher than the entropy calculations based on MD trajectories. On the other hand, the entropy of the aromatic amino acids phenylalanine, tryptophane, and tyrosine is underestimated by the same rotamer approach. Different workers using different rotamer definitions obtain a different rank order. The values for serine and threonine given by Abagyan and Totrov,[13] for example, are approximately 30% lower than those given by Pickett and Sternberg.[11] Uncertainties in regard to these extremes show shortcomings in the very simple rotamer counting methods. Pickett and Sternberg[11] appear to have overestimated the entropy in the torsional angles of the rather short side chains of serine and threonine. In the case of the aromatic amino acids, the underestimation is probably due to the neglect of vibrational motions of the aromatic rings.

The entropy per residue, which was normalised by dividing it through the number of atoms in each residue, was also correlated against the data of Pickett and Sternberg[11] (see Fig. 8). The reasoning behind this is that the vibrational part of the entropy is certainly proportional to the number of atoms and thus a better correlation can be expected. The correlation coefficient from a linear regression is indeed higher (0.79). Still, the ranking of amino acids in both methods is very different.

Side-chain configurational entropies can also be estimated using rotamer counting in combination with force-

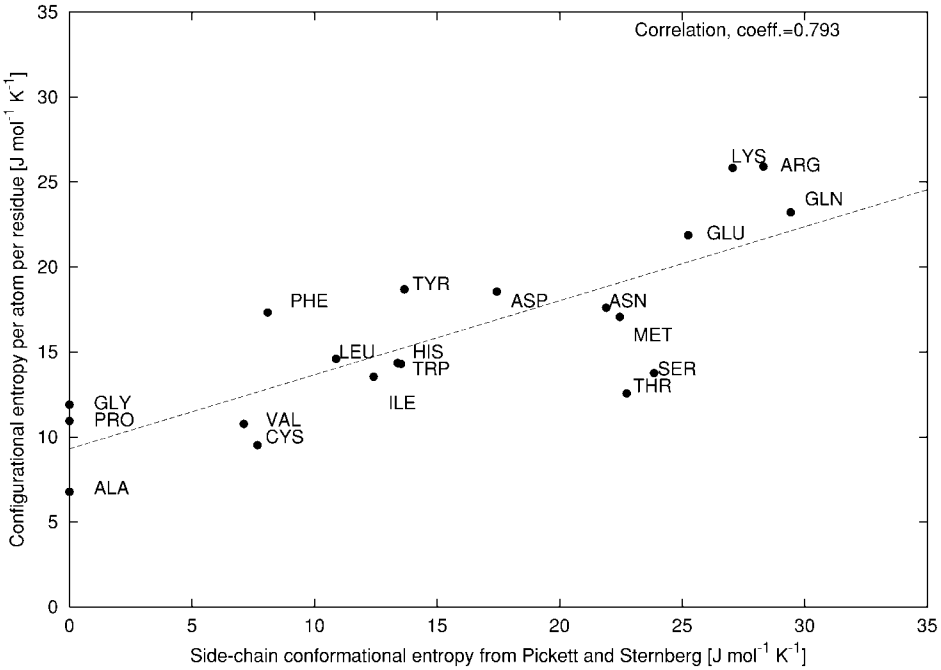

Fig. 8. Correlation between the side-chain conformational entropy obtained by rotamer counting using a database of 50 different protein structures (Pickett and Sternberg[11]) and the configurational entropy per residue divided through the number of atoms in the residue from the R1 simulation.

field calculations.[12] The values obtained are of similar magnitude as those presented here. For a concise discussion of the approximate sizes of the contributions of various degrees of freedom to the configurational entropy of proteins, we refer to Karplus et al.[21]

Ideally, one would wish to compare the results from the simulations directly with data obtained experimentally.[22–24] For example, the change in nuclear magnetic resonance order parameters between the native and denatured states of some proteins has been empirically correlated with changes in the side-chain or backbone entropy on folding. The interpretation of the change in order parameters is, however, strongly model dependent and does not provide an absolute scale.[22–24] Ultimately, experimental verification of the results from simulation requires the ability to analyse a representative ensemble of conformations that will hopefully be available in the future.

CONCLUSION

We have shown that it is possible to estimate the configurational entropy of an entire protein using the formula of Schlitter (eq. 1). The time to obtain fully convergent results for the entire protein is probably in the range of 10 ns. The distribution between the backbone and side chains and the correlation between both atom sets is very similar in α-lactalbumin at low pH compared to the β-heptapeptide described in Schäfer et al.[7] This was unexpected but strengthens the assumption that the physical principles governing peptide and protein motion are the same.

No significant difference could be found between the two simulation models of the molten globule state of α-lactalbumin. The entropy introduced by changing almost all χ_1 angles by 120° quickly dissipates to all degrees of freedom of the system and both simulations seem to converge to the same result.

The configurational entropy per residue calculated using an individual fit to the backbone of each residue converges within the 2-ns simulations considered here. As expected, the exposed side chains possess more entropy than the buried side chains. There is generally a positive correlation between solvent accessibility and configurational entropy of a residue. Using the covariance matrix of atom-positional fluctuations from MD simulations to calculate the side-chain entropy is a much more direct method and involves less assumptions than the side-chain rotamer counting methods.

With so many protein simulations now covering a couple of ns, a systematic study of side-chain entropies would be of interest. This will give more statistics on different amino acids in different parts of the protein structure.

ACKNOWLEDGMENTS

L.J.S. is a Royal Society University Research Fellow.

REFERENCES

1. Dobson CM. Unfolded proteins, compact states and molten globules. Curr Opin Struct Biol 1992;2:6–12.
2. Ptitsyn OB. Molten globule and protein folding. Adv Protein Chem 1995;47:83–229.
3. Kuwajima K. The molten globule state of α-lactalbumin. FASEB J 1996;10:102–109.
4. Smith LJ, Dobson CM, van Gunsteren WF. Side-chain conformational disorder in a molten globule: Molecular dynamics simulations of the A-state of human α-lactalbumin. J Mol Biol 1999;286: 1567–1580.
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. Schäfer H, Daura X, Mark AE, van Gunsteren WF. Entropy calculations on a reversibly folding peptide: changes in solute free energy cannot explain folding behaviour. Proteins 2001;43:45–56.
8. Smith LJ, Dobson CM, van Gunsteren WF. Molecular dynamics simulations of human α-lactalbumin: Changes to the structural and dynamical properties of the protein at low pH. Proteins 1999;36:77–86.
9. Acharya KR, Ren JS, Stuart DI, Phillips DC, Fenna RE. Crystal structure of human α-lactalbumin at 1.7 Å resolution. J Mol Biol 1991;221:571–581.
10. Dill KA. Dominant forces in protein folding. Biochemistry 1990;29: 7133–7155.
11. Pickett SD, Sternberg MJE. Empirical scale of side-chain conformational entropy in protein folding. J Mol Biol 1993;231:825–839.
12. Lee KH, Xie D, Freire E, Amzel LM. Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. Proteins 1994;20:68–84.
13. Abagyan R, Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. J Mol Biol 1994;235:983–1002.
14. Doig AJ, Sternberg MJ. Side-chain conformational entropy in protein folding. Protein Sci 1995;4:2247–2251.
15. Stapley BJ, Doig AJ. Free-energies of aminoacid side-chain rotamers in alpha-helices, beta-sheets and alpha-helix N-caps. J Mol Biol 1997;272:456–464.
16. Brady GP, Sharp KA. Entropy in protein folding and in protein-protein interactions. Curr Opin Struct Biol 1997;7:215–221.
17. Pal D, Chakrabarti P. Estimates of the loss of main-chain conformational entropy of different residues on protein folding. Proteins 1999;36:332–339.
18. Creamer TP. Side-chain conformational entropy in protein unfolded states. Proteins 2000;40:443–450.
19. Galzitskaya OV, Surin AK, Nakamura H. Optimal region of average side-chain entropy for fast protein folding. Protein Sci 2000;9:580–586.
20. Hubbard SJ, Thornton JM. NACCESS computer program. Departments of Biochemistry and Molecular Biology, University College London, 1993.
21. Karplus M, Ichiye T, Pettitt BM. Configurational entropy of native proteins. Biophys J 1987;52:1083–1085.
22. Yang D, Mok YK, Forman-Kay JD, Farrow NA, Kay LE. Contributions to protein entropy and heat capacity from bond vector motions measured by NMR spin relaxation. J Mol Biol 1997;272: 790–804.
23. Lee AL, Kinnear SA, Wand AS. Redistribution and loss of side-chain entropy upon formation of a calmodulin-peptide complex. Nat Struct Biol 2000;7:72–77.
24. Wrabl JO, Shortle D, Woolf TB. Correlation between changes in nuclear magnetic resonance order parapemters and conformational entropy: molecular dynamics simulations of native and denatured staphylococcal nuclease. Proteins 2000;38:123–133.