WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Parallel tempering molecular dynamics folding simulation of a signal peptide in explicit water

Siegfried Höfinger,[1]* Benjamin Almeida,[2] and Ulrich H. E. Hansmann[1,3]

[1] Department of Physics, Michigan Technological University, Houghton, Michigan, 49331-1295

[2] Novartis Institutes for BioMedical Research, Brunnerstraße 59, A-1235 Vienna, Austria

[3] John v. Neumann Institute for Computing, Forschungszentrum Jülich, 52425 Jülich, Germany

ABSTRACT

Parallel temperature molecular dynamics simulations are used to explore the folding of a signal peptide, a short but functionally independent domain at the N-terminus of proteins. The peptide has been analyzed previously by NMR, and thus a solid reference state is provided with the experimental structure. Particular attention is paid to the role of water considered in full atomic detail. Different partial aspects in the folding process are quantified. The major group of obtained structures matches the NMR structure very closely. An important biological consequence is that in vivo folding of signal peptides seems to be possible within aqueous environments.

Proteins 2007; 68:662–669.

Key words: *enhanced sampling; aldehyde dehydrogenase; structure determination; structure prediction; solvation.*

INTRODUCTION

Predicting the natural shape of proteins solely from sequence data is believed to form one of todays major challenges to biophysical research.[1–3] A particularly interesting case is the folding of signal peptides (SP), short hydrophobic stretches of amino acids at the N-terminal end of a protein sequence. Such SPs represent a well separated domain of independent biological functionality, that is targeting the active ribosome at early stages in the biosynthesis of proteins to the membrane.[4–7] Like all other proteins a SP needs first to adopt a certain specific shape before it can carry out its biological task (here guiding the translating ribosome to the membrane). The hydrophobic collapse[8–10] constitutes an important driving force into this shape. For this reason, SPs can be considered an ideal minimal system for the study of protein folding.

Molecular Dynamics (MD) simulations can be used to study the pathway of protein folding.[11–14] However, care must be taken that the simulation does not become trapped in one of the many local minima on the rough energy landscape corresponding to the proteins' vast conformational space. Powerful techniques have been devised to overcome that problem.[15–17] While encouraging results are reported, the predictive quality of computer simulations should be tested further. Most straightforward is the comparison of simulated structures to experimentally derived structures. The latter typically result from X-ray crystallographic analysis or NMR spectroscopy.[18–20]

In the present study, we use parallel tempering MD (PTMD) in explicit water for the simulation of the folding of the SP of rat liver aldehyde dehydrogenase. This structure has been resolved by NMR analysis previously.[21,22] The SP consists of 22 amino acids and contains five positively charged residues in the hydrophobic domain.[23] Its sequence is NH_2-Met_1-Leu_2-Arg_3-Ala_4-Ala_5-Leu_6-Ser_7-Thr_8-Ala_9-Arg_{10}-Lys_{11}-Gly_{12}-Pro_{13}-Arg_{14}-Leu_{15}-Ser_{16}-Arg_{17}-Leu_{18}-Leu_{19}-Ser_{20}-Tyr_{21}-Ala_{22}-$CONH_2$. NMR analysis reveals a helix-loop-helix motif where an N-terminal α-helical element (residues 3–10) is linked to a C-terminal

Abbreviations: $-CONH_2$, C-terminal end group; DSSP, a popular program for protein secondary structure analysis; EWALD, proper treatment of long range electrostatic interactions; k_B, Boltzmann constant 1.38066 10^{-23} J/K; MD, molecular dynamics; MOLDEN, popular visualization program; N/C-terminal, amino-terminal begin/end in primary sequence; NH_2-, N-terminal end group; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NpT, constant number of particles, constant pressure, constant temperature; pdb, protein data bank; PTMD, parallel tempering molecular dynamics; RMSD, root mean square deviation; SASA, solvent accessible surface area; SHAKE, no variation of bond lengths involving H-atoms; SP, signal peptide; SRP, signal recognition particle; TINKER/AMBER, popular modeling packages; TIP3P, a widely used water model; TM, transmembrane domain; X-ray, short wavelength radiation; xyz-format, plain Cartesian coordinate format; α-helical, secondary structural element in proteins.

The Supplementary Material referred to in this article can be found online at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsor: National Institutes of Health; Grant number: GM62838

*Correspondence to: Siegfried Höfinger, Michigan Technological University, Department of Physics, 1400 Townsend Drive, Houghton, MI 49931-1295. E-mail: shoefing@mtu.edu

Received 11 May 2006; Revised 23 August 2006; Accepted 21 September 2006

Published online 17 May 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21268

α-helical element (residues 14–21) via the flexible linker region (residues Lys_{11}-Gly_{12}-Pro_{13}).

There are two central points of interest in our present study. The first one is to probe whether PTMD can reliably predict the structural motif delivered by NMR. A second query is related to the role of the environment in this particular case. The reason that experimentally derived structures for SPs are rare lies in the technical difficulty associated with their strong hydrophobic character. In regular aqueous dilutions SPs will immediately start to aggregate, agglomerate and precipitate, hence cannot be used for the actual experiment. For this reason, both referenced NMR studies had to be carried out in membrane-like environments, that is dodecylphosphocholine micelles. These experimental difficulties suggest that SPs (and other mainly hydrophobic domains) fold only correctly inside membranous environments. On the other hand, biological considerations rather suggest another scenario. When SPs emerge from the ribosome they are immediately recognized and bound by the signal recognition particle (SRP).[24] Such a specific recognition prevents SPs from interacting with each other, hence from aggregation and precipitation. Therefore, SPs need to be considered isolated and largely individual entities that mainly face the aqueous environment of the interior of the ribosome exit channel. But in order to facilitate their duties, SPs should have already folded properly once they exit the ribosome. Indeed, it has been shown that a transmembrane sequence (TM)—a structural element closely related to SPs—can adopt its α-helical shape already far inside the ribosome exit tunnel.[4] Based on these findings, a picture arises in which it is exactly the aqueous environment that short hydrophobic stretches need to fold properly. The authors of Ref. 4 have addressed that point by speculating that folding of the TM takes place at specific sites of increased hydrophobicity adjacent to the ribosome exit tunnel. However, the crystal structure of the large ribosomal subunit[25] does not show extended locations of increased hydrophobic character in the vicinity of the ribosome exit channel.

The present computer simulations may aid in answering the outlined discrepancy. Our PTMD simulations consider the case of a non-interfering protein embedded in water which closely resembles the biological situation of an evolving SP confined to the interior of the ribosome. If PTMD successfully predicts the correct fold, then the *in vivo* folding of SPs can really occur in aqueous environments, instead of requiring a membrane-like environment.

In the following we will present results obtained from PTMD simulations on the SP of rat liver aldehyde dehydrogenase, discuss their impact on biology and biophysics, outline briefly the principle of PTMD and give detailed technical information concerning the protocol employed.

METHODS

Parallel tempering molecular dynamics-PTMD

PTMD simulations are one possible way to address the rough character of the potential energy surface, where the risk is high to get trapped in one of the many local minima. The principle has been described in Ref. 26. The essential kernel of PTMD is to not only run a single long term simulation, but run the same simulation in parallel multiple times at varying temperatures. At periodic intervals all simulations are halted and exchanges of the current structures between neighboring temperatures i and j are pursued with probability

$$e^{(\beta_j-\beta_i)(V_j-V_i)} > \mathrm{RND} \qquad (1)$$

where V_j and β_j are total energy and Boltzmann factor of the simulation running at temperature T_j, while V_i and β_i are total energy and Boltzmann factor for the simulation running at temperature T_i. The Boltzmann factor is given by $1/k_B T$ with k_B the Boltzmann Constant (i.e. $k_B = 1.3807 \times 10^{-23}$ J K^{-1} and the j-simulation is assumed here to run at higher temperature T_j than the simulation with temperature at T_i. A random number RND between 0 and 1 is chosen and the exchange of coordinates will be performed only if the above expression holds true.

PTMD protocol and employed programs

Initial structure creation

Program protein.x from the TINKER modeling package version 4.2[27] has been used to convert the amino acid sequence corresponding to the SP of the rat liver mitochondrial aldehyde dehydrogenase, that is Met_1-Leu_2-Arg_3-Ala_4-Ala_5-Leu_6-Ser_7-Thr_8-Ala_9-Arg_{10}-Lys_{11}-Gly_{12}-Pro_{13}-Arg_{14}-Leu_{15}-Ser_{16}-Arg_{17}-Leu_{18}-Leu_{19}-Ser_{20}-Tyr_{21}-Ala_{22}, into a strictly linear protein structure in 3D xyz format. The AMBER force field[28] has been selected. The xyz formatted coordinate file was read with program MOLDEN[29] and a structure file was written in pdb file format[30] excluding all hydrogen atoms. The pdb formatted structure was read from program *xleap* coming with AMBER version 6.0 and thus all hydrogen atoms were re-assigned properly. Five Cl^- counter ions were added according to rough electrostatic criteria and the system was soaked in TIP3P water extending the bare protein structure by 10 Å in all directions. Resulting AMBER internal coordinate and topology files were written and stored. Resulting simulation cell dimensions were 45 Å × 72 Å × 91 Å.

Initial minimization and equilibration

About 2000 steps of steepest descent minimization were performed with program *sander* from the AMBER

package starting with the previously created pair of coordinate/topology files. Bond interactions involving H-atoms were considered by the SHAKE algorithm.[31] The partial optimized structure was taken up and became subject to an initial equilibration run consisting of 50,000 steps of MD (NpT ensemble) at 1 fs time steps using EWALD summation with a direct space cut off radius of 8 Å and a Berendsen thermostat set to 300 K simulation temperature.[32] SHAKE was switched on here again. After approximately 17 ps stably equilibrated property values were reached, that is target temperature of 300 K, target pressure of 1 atm, total energy of −59,500 kcal mol^{-1} and a macroscopic density of 0.99 g cm^{-3}. Resulting coordinate/velocity restart files were stored for further PTMD simulation.

PTMD protocol

Following a previous example[33] but using explicit solvation instead,[34] we employed 24 different parallel temperatures, that is 250, 255, 260, 265, 273, 284, 298, 315, 333, 353, 373, 393, 413, 433, 454, 473, 493, 513, 533, 553, 573, 593, 613, and 633 all given in units of Kelvin. Other choices could have been made[35] and a fine-tuned pattern of optimized parallel temperatures is certainly subject to future work, but for this present case we prefer to follow a two stage strategy (see below), where we make an initial equilibration phase less sensitive to the actual choice of parallel temperatures by artifically increasing the exchange probability. The final structure obtained in the 300 K equilibration run became the starting structure to all these 24 parallel temperature simulations. Each incremental MD simulation was lasting for 5 ps. At first all parallel temperatures were considered for 800 successive, incremental steps of PTMD simulations (equilibration phase of PTMD). It is important to note that during this PTMD equilibration phase the exchange probability between adjacent temperature simulations was artificially increased by introducing a multiplicative scaling factor of ≈400 in the exponent of Eq. (1). The PTMD equilibration phase was followed by 800 steps of successive PTMD simulations without the artifical scaling factor in the exponent of Eq. (1) (PTMD production phase). Analogous MD run time options for the *sander* program were used in either PTMD phase similar to those taken earlier for the 17 ps initial equilibration, except that the frequency of heat bath coupling was increased (parameter TAUTP = 0.1. which means adjustments to the reference temperature performed every 0.1 ps). Trials of pairwise exchange of coordinates/velocities were invoked always at the end of each of the 5 ps intervals starting with the lowest temperature at first and comparing it to the second lowest, then comparing second lowest to third lowest and so forth. Coordinates/velocities were allowed to swap between several successive neighboring temperatures if application of Eq. (1) did allow so. The type of executed exchange was monitored and logged throughout the simulation. During equilibration PTMD 3% of the attempts of exchange were of obvious type, that is the total energy V_j at the higher temperature simulation lies deeper than V_i, the total energy of the lower temperature simulation. Thirty-three percent of all attempts during equilibration PTMD led to a rejection of interchange and 64% of the pursued swaps were successful according to the criterion given from Eq. (1). These ratios changed significantly in the production phase of PTMD, that is 0.8% of all cases were of obvious exchange type, 1.2% occurred because of successful application of Eq. (1) while the majority of attempts, 98.0%, was rejected. Intermediary structures after exchange were stored and processed with program *ptraj* from AMBER, that is conversion to pdb format and deletion of the water box were performed. Total energies including the explicit water molecules were recorded for each PTMD snapshot at the end of each 5 ps interval.

Data evaluation and processing

Clustering. For each of the parallel temperatures, all structures collected in the PTMD production phase are grouped into clusters of similar geometry. A root mean square deviation (RMSD) criterion of 1.0 Å was employed during pairwise comparisons of individual structures. Only the backbone atoms were taken into account and carbonyl oxygens were excluded for RMSD calculation. Each structure selected for pairwise comparison was (i) converted to xyz format (program *pdbxyz* TINKER), (ii) translated into the center of mass and rotated into the inertial frame (program *xyzedit* TINKER), (iii) subjected to superposition and RMSD calculation (program *superpose* TINKER). AMBER parameters were used for all these operations involving TINKER tools. Upon termination of the binning, only significantly populated clusters were taken, that is clusters containing at least 10 or more items, and a superposition of all these selected structures was pursued if the main conformation of the backbone did allow so (visual inspection). Final average structures were converted to pdb format with program MOLDEN.

Calculations of molecular properties. All snapshot structures in each of the parallel temperature simulations were used for calculation of the following properties: (i) vacuum potential energy (program *analyze* TINKER). (ii) solvation free energy obtained from the implicit solvent model by Eisenberg, McLachlan, and Wesson[36,37] (program *analyze* TINKER), (iii) solvation free energy obtained from the implicit solvent model by Ooi et al.[38] (program *analyze* TINKER), (iv) solvation free energy obtained from the implicit solvent model by Still et al.[39] (program *analyze* TINKER), (v) solvation free energy obtained from the implicit solvent model by Qiu et al.[40]

DOI 10.1002/prot

(program *analyze* TIN-KER), (vi) solvation free energy obtained from the implicit solvent model by Hawkins et al.[41] (program *analyze* TINKER), and (vii) solvent accessible surface area (SASA) and solvent excluded volume (program *spacefill* TINKER).

Calculation of the specific heat. Following the procedure devised at Ref. 33, we compute the heat capacities at constant pressure from finite differences, that is $C_p = [E(T_{n+1}) - E(T_{n-1})]/(T_{n+1} - T_{n-1})$, with $E(T_n)$ being the average total energy comprising potential as well as kinetic components at some temperature T_n. Only the PTMD production set of conformations is taken into account.

Radius of gyration calculations. Each structure of all parallel temperature simulations was converted from pdb format to xyz format (program *pdbxyz* TINKER), translated into the center of mass and rotated into the inertial frame (program *xyzedit* TINKER), and finally subjected to calculation of the radius of gyration (awk script, self-made).

Calculations of the fraction of helical content. All the pdb formatted structures of each parallel temperature simulation were processed with program DSSP[42]. The DSSP classification was analyzed and all atoms within helical signatures (classes H = alpha helix, G = 3-helix (3/10 helix), I = 5 helix (pi helix)) were counted and finally related to the total number of atoms resulting in values for the relative helical content.

RESULTS

The employed PTMD protocol delivers a coherent picture of converged simulation results

800 steps of equilibration PTMD simulation followed by 800 steps of production PTMD were carried out on the 22-amino acid SP of rat liver mitochondrial aldehyde dehydrogenase in explicit water starting from a fully extended linear fiber conformation. The covered temperature domain reached from 250 to 633 K and was split into 24 individual parallel temperatures. The focus was on the lowest temperature results (expected global minimum) and on room temperature results seen at 298 K. The set of parallel temperatures below 298 K was chosen to allow for a more detailed study of low-energy configurations. The trend of the potential energy of the isolated SP as a function of PTMD progress is shown in Figure 1. A significant drop of about −150 kcal/mol occurs after approximately 400 steps of PTMD equilibration indicating stabilization because of chain collapse. The relief in vacuum energy is seen at low temperatures (green and blue lines in Fig. 1) but not at high temperatures (red line in Fig. 1). Additionally, monitored molecular properties such as the radius of gyration (see Fig. 2). SASA and the solvent excluded volume (see Supplementary Material) show a similar evolution during PTMD simulation, indicating convergence after about 400 steps of equilibration PTMD. Again, while low temperature data are following comparable trends, the results at 633 K differ considerably. Together with the peak in the specific heat at around $T = 550$ K, shown in Figure 3(a), these results clearly demonstrate that our upper temperature of 633 K is well within the high temperature regime where the crossing of conformational barriers is possible. Attributing the initial 400 steps of equilibration PTMD to chain collapse, (see Supplementary Material) we yield a protein stabilization energy of about −28 kcal/mol (block average energies for always 100 steps PTMD at 250 K are −183 kcal/mol for steps 400–500. −216 kcal/mol for steps 500–600. −207 kcal/mol for steps 600–700, and −211 kcal/mol for steps 700–800, respectively).

The specific role of water in the folding process of SPs

Quantification of the energy coming from the environment in the folding process of the SP is extremely difficult. First of all it must be mentioned that the employed starting conformations, that is fully extended protein chains of almost linear shape, are very unphysical and must be considered far off the *in vivo* state. So, the initial phase of our simulation will most likely correspond to chain collapse rather than actual folding of the peptide. However, one of the explicit goals of this study was to examine whether plain PTMD simulations in explicit water will manage to get the kink between the two helices positioned at the correct site, that is at residues Lys_{11}, Gly_{12}, and Pro_{13}. Therefore, to avoid any initial bias we needed to start from fully extended chain conformations. This is particularly important because the conventional view of the shape of SPs is that of a single, linear, alpha helical element[6] but in this present case we clearly deal with a collapsed form of the shape of a helical hairpin. The second factor to take into account is that our PTMD protocol never really included any actual storage of explicit water coordinates. Rather the individual cycle files were processed on the fly and the explicit water shell was removed before the remaining coordinates corresponding solely to the SP were stored permanently on disk. A more elaborate desolvation analysis of the type reported in Ref. 43 is therefore impossible in our case. One can however also obtain semi-quantitative solvation data *a posteriori* to recording a trajectory of plain protein coordinates.[44] This is accomplished with the help of continuum solvation methods that deliver solvation free energies for each snapshot structure of the trajectory. Following along these lines, we apply five different implicit solvation models[36–41] to the series of PTMD snapshots. Similar to the trend in vacuum energy a common feature observed in all these continuum solvation studies is the achievement of a stable plateau value

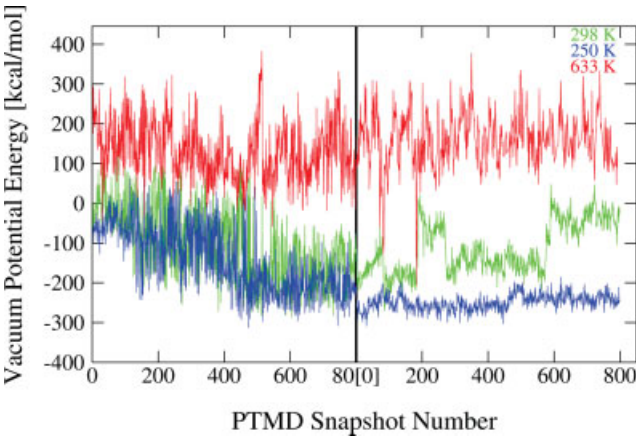

Figure 1

Evolution of vacuum potential energies during PTMD. The green line shows the trend for the 298 K simulation, while the blue line represents a parallel temperature run at 250 K and the red line shows results for 633 K. Switch from an equilibration interval to a production interval is indicated by a vertical black line.

after about 400 initial equilibration PTMD steps (see, e.g. Fig. 4 for the solvation model of Ref. 36 and Supplementary Material). All tested continuum solvation models render the protein solvation free energy negative in sign. This would indicate a favorable role of the aqueous environment in the folding process of the SP. Absolute numbers, however, differ considerably between individual models and are on an entirely different scale when regarding total energies of the full systems including the contributions from the water box (see Fig. 3(b) and Supplementary Material). Block average solvation free energies for always 100 steps PTMD at 250 K using the model of[36] are −93 kcal/mol for steps 400–500, −86 kcal/mol for steps 500–600, −89 kcal/mol for steps 600–700, and −90 kcal/mol for steps 700–800, respectively. Similarly, all the other continuum solvation models (see Supplementary Material) would also estimate the net solvation free energy for the actual folding process—that is PTMD step 400 onwards—to be in the range of +5 kcal/mol to +15 kcal/mol. This would lead us to a net free energy of folding of the SP in the range of −13 kcal/mol to −23 kcal/mol.

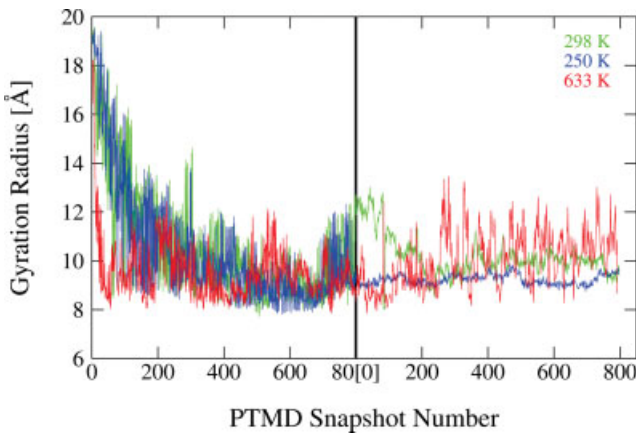

Figure 2

Evolution of the radius of gyration during PTMD. The green line shows the trend for the 298 K simulation, the blue line represents a parallel temperature run at 250 K and the red line shows results for 633 K. Equilibration intervals are separated from production intervals by a vertical black line.

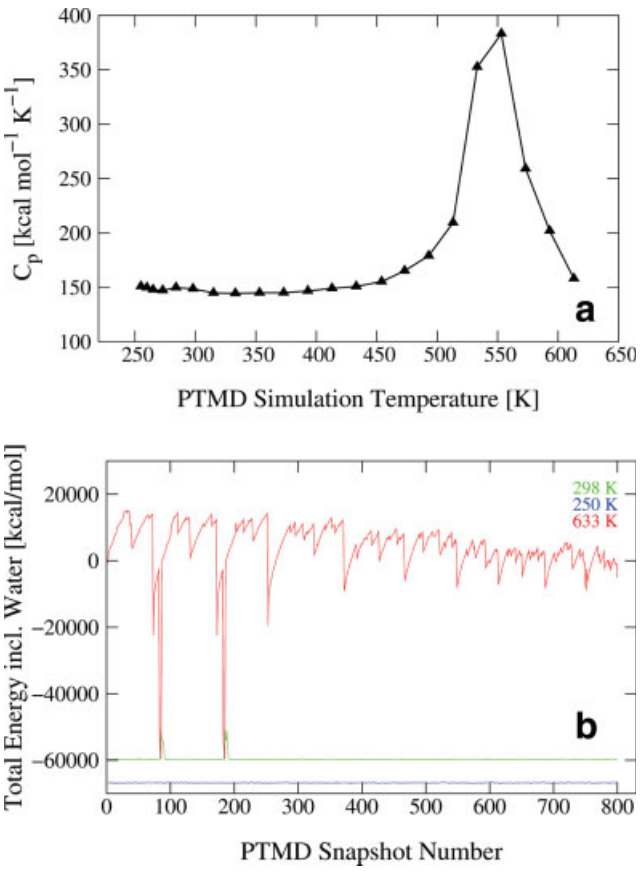

Figure 3

*(**a**) Calculated heat capacities at constant pressure as a function of the PTMD temperature. (**b**) Evolution of total energies (EWALD sums) in the production interval of PTMD. The green line shows the trend for the 298 K simulation, while the blue line represents a parallel temperature run at 250 K and the red line shows results for 633 K.*

PTMD simulations reproduce the NMR structure very well

Structural snapshots obtained during PTMD production phase are clustered corresponding to a criterion of RMSD of <1.0 Å between all backbone atoms of a pair

DOI 10.1002/prot

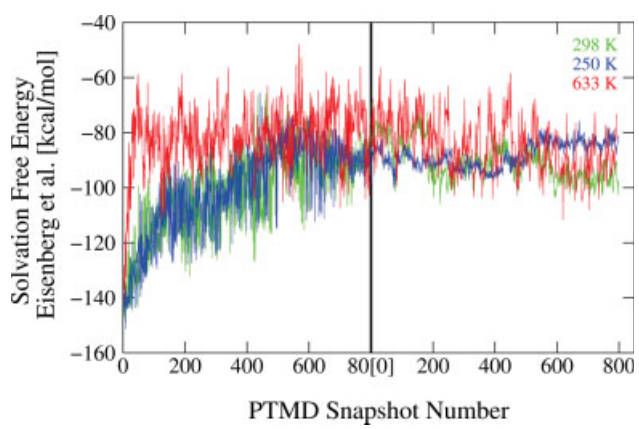

Figure 4

Evolution of solvation free energies due to Eisenberg et al.[36] during PTMD. The green line shows the trend for the 298 K simulation, the blue line represents a parallel temperature run at 250 K and the red line shows results for 633 K. Equilibration intervals are separated from production intervals by a vertical black line.

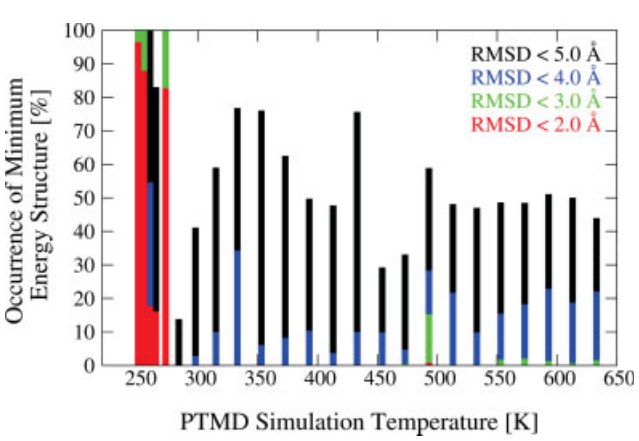

Figure 6

Frequency of occurrence of the putative minimum energy structure (seen at 250 K) for the SP of rat liver mitochondrial aldehyde dehydrogenase in all parallel temperatures. Similarity is measured in terms of RMSD. The minimum energy structure is frequently found only at low temperatures.

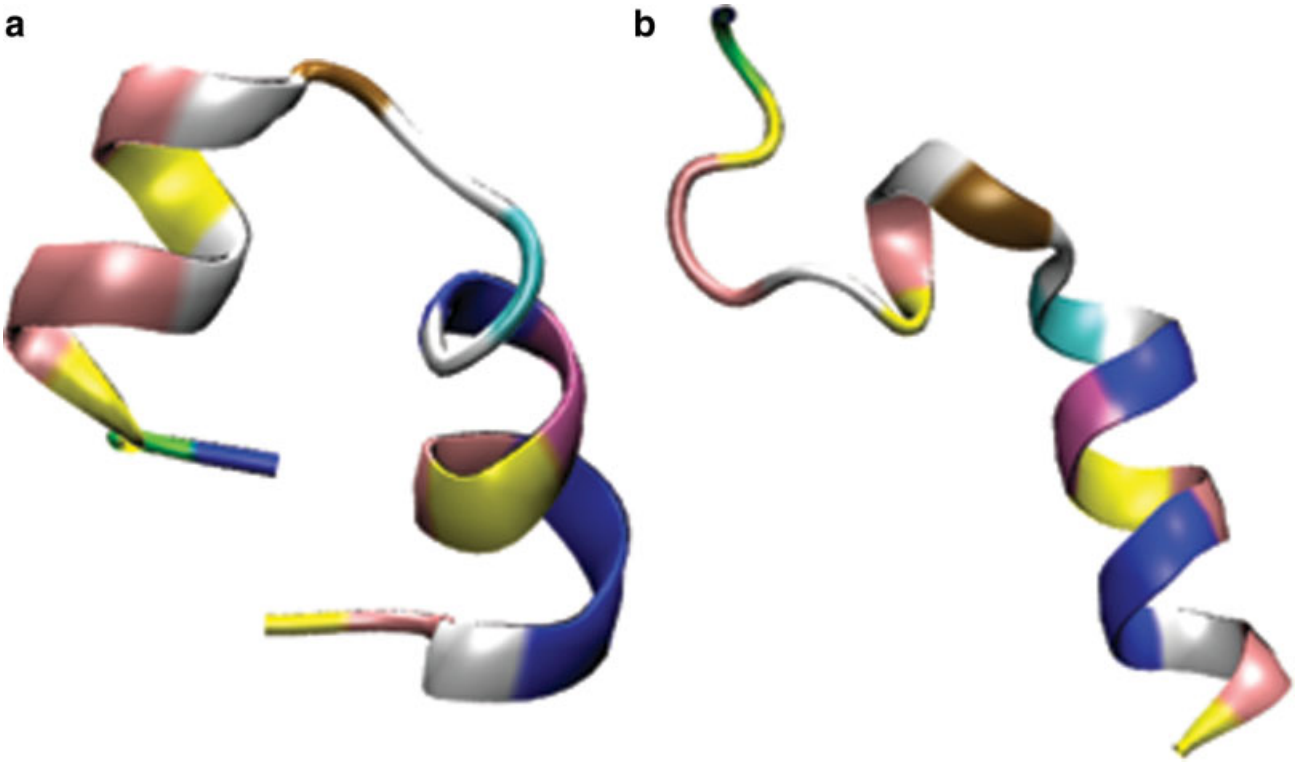

Figure 5

*Major structural motif of the SP of rat liver mitochondrial aldehyde dehydrogenase obtained upon clustering the production PTMD data of the parallel temperature run at 250 K (**a**) and 298 K (**b**). The kink occurs at residues Lys_{11} (cyan), Gly_{12} (white), and Pro_{13} (brown). The N-terminal Met_1 is found at the lower right corner (yellow). Leu and Thr are shown in pink, Arg and Gly are given in white, Ala is held blue, Ser is colored yellow, and Tyr shown in green. A helical hairpin structure is adopted with perfect α-helical fold of three full turns at the N-terminus and two full turns at the C-terminus for the 250 K run, and partial C-terminal unfolding at the 298 K simulation.*

Table I
Confrontation of Structural Features in the Major Structural Motif Obtained from PTMD at 250 K to the NMR Data Reported in Karslake et al.[21]

NOE pair/NMR finding	Separation in PTMD motif (Å)	Verification of NMR-experiment
Arg_{14} ($N_\alpha H$)-Leu_{15} ($N_\alpha H$)	2.868	Y
Leu_{15} ($N_\alpha H$)-Ser_{16} ($N_\alpha H$)	2.496	Y
Ser_{16} ($N_\alpha H$)-Arg_{17} ($N_\alpha H$)	2.559	Y
Arg_{17} ($N_\alpha H$)-Leu_{18} ($N_\alpha H$)	2.596	Y
Leu_{18} ($N_\alpha H$)-Leu_{19} ($N_\alpha H$)	2.713	Y
Leu_{19} ($N_\alpha H$)-Ser_{20} ($N_\alpha H$)	2.832	Y
Ser_{20} ($N_\alpha H$)-Tyr_{21} ($N_\alpha H$)	3.134	Y
Ala_5 ($N_\alpha H$)-Leu_6 ($N_\alpha H$)	2.132	Y
Leu_6 ($N_\alpha H$)-Ser_7 ($N_\alpha H$)	2.854	Y
Ser_7 ($N_\alpha H$)-Thr_8 ($N_\alpha H$)	2.804	Y
Thr_8 ($N_\alpha H$)-Ala_9 ($N_\alpha H$)	2.670	Y
Ala_9 ($N_\alpha H$)-Arg_{10} ($N_\alpha H$)	2.848	Y
Arg_3 ($C_\alpha H$)-Leu_6 ($N_\alpha H$)	3.185	Y
Ser_{16} ($C_\beta H$)-Leu_{19} ($N_\alpha H$)	5.600	N
Arg_3 ($C_\alpha H$)-Ser_7 ($N_\alpha H$)	3.576	Y
Ser_{16} ($C_\beta H$)-Arg_{17} ($N_\alpha H$)	2.758	Y
Ser_{20} ($C_\beta H$)-Tyr_{21} ($N_\alpha H$)	4.103	Y/N
Leu_{18} ($N_\alpha H$) reduced exchange rate	Visual check, very buried	Y
Leu_{19} ($N_\alpha H$) reduced exchange rate	Visual check, very buried	Y
Leu_{15} ($N_\alpha H$) increased exchange rate	Visual check, less buried	Y

of structures excluding carbonyl oxygens of the peptide bond. Average structures are deduced in all main clusters and superimposed again amongst each other when the major structural motif is found to be identical. The PTMD result for the simulation at 250 K is shown in Figure 5(a). It is representative of more than 95% of all configurations collected at this temperature (pdb coordinates provided in Supplementary Material). Small structural flexibility is seen only in the exact orientation of the sidechains of the basic residues (see Supplementary Material). An N-terminal α-helical element comprising residues Leu_2 to Lys_{11} is separated from a second α-helical element extending from residues Pro_{13} to Ser_{20} at the C-terminus. Both helical elements are connected via the linker residues Lys_{11}, Gly_{12}, and Pro_{13} which are shown in cyan, white, and brown in Figure 5. The identified structure at 250 K is in very good agreement with the experimental data reported in Ref. 21. Specific comparison to the NMR data is shown in Table I. With increasing simulation temperature the ensemble of recorded structures becomes increasingly disperse (see Table I in Supplementary Material). The C-terminal helix starts to unravel and either terminus tends toward a minimal helical motif of about two and a half full turns [see for example a frequently observed structural motif obtained for 298 K in Fig. 5(b) and Supplementary Material]. Calculation of the relative α-helical content yields a percentage of 40–50% (see Supplementary Material) for the 298 K run which is very close to the CD data reported in Ref. 21. Occurrence of the helical hairpin structure from the 250 K simulation—which we consider to be a global minimum—in all the remaining parallel temperatures is represented in Figure 6. RMSD values with respect to the 250 K structure are computed for all the conformations stored in all parallel temperature runs and corresponding fractions of similarity are calculated. Thus the minimum energy structure is only found at low temperatures. In summary, comparing our results to the NMR data of Ref. 21 where two helical segments from residues 3 to 10 and from residues 14 to 21 were observed, we find a very satisfactory overall agreement with experimental data.

DISCUSSION AND CONCLUSIONS

A full *ab initio* PTMD folding simulation of the 22-amino acid SP of rat liver mitochondrial aldehyde dehydrogenase in explicit water has yielded a structural motif very similar to the NMR structure as the global minimum. This is not only an encouraging affirmation of this technique but also has important biological implications. Contrary to *in-vitro* knowledge, the folding of an isolated hydrophobic SP can take place in an aqueous environment thus backing the early structure exhibition model of Johnson and coworkers[4] for confined ribosomal compartments. While the experimental structure is detected here as the global minimum at low temperatures, an increase in thermal flexibility shows an alteration of this motif to become relevant for room temperature. Most notably are the increase in ensemble dispersion of individual structures as well as an increased probability of unfolding of the C-terminal helix and enhanced re-partitioning of the two helical branches into an equal-sized hairpin of two and a half full turn for each of the branches. All this might reflect a basic requirement for a very subtle fine balance of different environmental effects in the case of true *in vivo* folding of SPs.

In conclusion, the presented work is an interesting case of successful protein folding simulation, but also demonstrates that the role of water in the process of hydrophobic collapse is far from obvious.

REFERENCES

1. Socolich M, Lockless SW, Russ WP, Lee H, Gardner KH, Ranganathan R. Evolutionary information for specifying a protein fold. Nature 2005;437:512–518.
2. Snow CD, Sorin EJ, Rhee YM, Pande VS. How well can simulation predict protein folding kinetics and thermodynamics? Annu Rev Biophys Biomol Struct 2005;34:43–69.
3. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M. Simultaneous determination of protein structure and dynamics. Nature 2005;433:128–132.
4. Woolhead CA, McCormick PJ, Johnson AE. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. Cell 2004;116:725–736.

DOI 10.1002/prot

5. Van den Berg B, Clemons WM, Collins I, Modis Y, Hartmann E, Harrison SC, Rapoport TA. X-ray structure of a protein-conducting channel. Nature 2004;427:36–44.
6. von Heijne G. Protein transport: life and death of a signal peptide. Nature 1998;396:111–113.
7. Mitra K, Schaffitzel C, Shaikh T, Tama F, Jenni S, Brooks CL, III, Ban N, Frank J. Structure of the *E. coli* protein-conducting channel bound to a translating ribosome. Nature 2005;438:318–324.
8. Despa F, Fernández A, Berry RS. Dielectric modulation of biological water. Phys Rev Lett 2004;93:228104.
9. Chandler D. Interfaces and the driving force of hydrophobic assembly. Nature 2005;437:640–647.
10. Liu P, Huang X, Zhou R, Berne BJ. Observation of a dewetting transition in the collapse of the melittin tetramer. Nature 2005;437: 159–162.
11. Snow CD, Nguyen H, Pande VS, Gruebele M. Absolute comparison of simulated and experimental protein-folding dynamics. Nature 2003;420:102–106.
12. Duan Y, Kollmann PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 1998;282:740–747.
13. Karplus M, Kuriyan J. Molecular dynamics and protein function. Proc Natl Acad Sci USA 2005;102:6679–6685.
14. Baron R, Bakowies D, van Gunsteren WF. Carbopeptoid folding: effects of stereochemistry, chain length, and solvent. Angew Chem Int Ed 2004;43:4055–4059.
15. Laio A, Parrinello M. Escaping free-energy minima. Proc Natl Acad Sci USA 2002;99:12562–12566.
16. Boczko EM, Brooks CL, III. First-principles calculation of the folding free energy of a three-helix bundle protein. Science 1995; 269:393–396.
17. Lin CY, Hu CK, Hansmann UHE. Parallel tempering simulations of HP-36. Proteins 2003;52:436–445.
18. Hendrickson WA. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science 1991;254:51–58.
19. Wüthrich K. The way to NMR structures of proteins. Nat Struct Biol 2001;8:923–925.
20. Wüthrich K. NMR of proteins and nucleic acids. New York: Wiley; 1986.
21. Karslake C, Piotto ME, Pak YK, Weiner H, Gorenstein DG. 2D NMR and structural model for a mitochondrial signal peptide bound to a micelle. Biochemistry 1990;29:9872–9878.
22. Thornton K, Wang Y, Weiner H, Gorenstein DG. Import, processing, and two-dimensional NMR structure of a linker-deleted signal peptide of rat liver mitochondrial aldehyde dehydrogenase. J Biol Chem 1993;268:19906–19914.
23. Goder V, Spiess M. Topogenesis of membrane proteins: determinants and dynamics. FEBS Lett 2001;504:87–93.
24. Spanggord RJ, Siu F, Ke A, Doudna JA. RNA-mediated interaction between the peptide-binding and GTPase domains of the signal recognition particle Nat Struct Mol Biol 2005;12:1116–1122.
25. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science 2000;289:905–920.
26. Hansmann UHE. Parallel tempering algorithm for conformational studies of biological molecules. Chem Phys Lett 1997;281:140–150.
27. Ponder JW. TINKER software tools for molecular design, Version 4.2. Copyright 1990–2004, Jay William Ponder, 2004.
28. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr., Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995;117:5179–5197.
29. Schaftenaar G, Noordik JH. Molden: a pre- and post-processing program for molecular and electronic structures. J Comput Aided Mol Des 2000;14:123–134.
30. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28:235–242.
31. van Gunsteren WF, Berendsen HJC. Algorithms for macromolecular dynamics and constraint dynamics. Mol Phys 1977;34:1311–1327.
32. Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
33. Pitera JW, Swope W. Understanding folding and design: replica-exchange simulations of "Trp-cage" miniproteins. Proc Natl Acad Sci USA 2003;100:7587–7592.
34. Zhou R, Berne BJ, Germain R. The free energy landscape for beta hairpin folding in explicit water. Proc Natl Acad Sci USA 2001;98:14931–14936.
35. Trebst S, Troyer M, Hansmann UHE. Optimized parallel tempring simulations of proteins. J Chem Phys 2006;124:174903.
36. Eisenberg D, McLachlan AD. Solvation energy in protein folding and binding. Nature 1986;319:199–203.
37. Wesson L, Eisenberg D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. Protein Sci 1992;1:227–235.
38. Ooi T, Oobatake G, Nemethy G, Scheraga HA. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proc Natl Acad Sci USA 1987;84:3086–3090.
39. Still WC, Tempczyk A, Hawley RC, Hendrickson T. A semiempirical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112:6127–6129.
40. Qiu D, Shenkin PS, Hollinger FP, Still WC. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate born radii. J Phys Chem A 1997;101:3005–3014.
41. Hawkins GD, Cramer CJ, Truhlar DG. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. J Phys Chem 1996;100:19824–19839.
42. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features. Biopolymers 1983;22:2577–2637.
43. Garcia AE, Onuchic JN. Folding a protein in a computer: An atomic description of the folding/unfolding of protein A. Proc Natl Acad Sci USA 2003:100:13898–13903.
44. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini Q, Cieplak P, Srinivasan J, Case DA, Cheatham TE, III. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 2000;33:889–897.