# Bulky Side Chains and Non-native Salt Bridges Slow down the Folding of a Cross-Linked Helical Peptide: A Combined Molecular Dynamics and Time-Resolved Infrared Spectroscopy Study

**Beatrice Paoli,[†] Michele Seeber,[‡] Ellen H. G. Backus,[§] Janne A. Ihalainen,[§] Peter Hamm,[§] and Amedeo Caflisch*[,†]**

*Department of Biochemistry and Department of Physical Chemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland, and Dulbecco Telethon Institute and Department of Chemistry, University of Modena and Reggio Emilia, Via Campi 183, I-41100 Modena, Italy*

*Received: November 27, 2008; Revised Manuscript Received: January 14, 2009*

Multiple 4-$\mu$s molecular dynamics (MD) simulations are used to study the folding process of the cross-linked α-helical peptide Ac-EACAR$^5$EAAAR$^{10}$EAACR$^{15}$Q-$NH_2$ (EAAAR peptide). The folding kinetics are single exponential at 330 K, while they are complex at 281 K with a clear deviation from single-exponential behavior, in agreement with time-resolved infrared (IR) spectroscopy measurements. Network analysis of the conformation space sampled by the MD simulations reveals four main folding channels which start from conformations with partially formed helical structure and non-native salt-bridges in a kinetically partitioned unfolded state. The independent folding pathways explain the comparable quality of models based on stretched exponential and multiexponential fitting of the kinetic traces at low temperature. The rearrangement of bulky side chains, and in particular their reorientation with respect to the cross-linker, makes the EAAAR peptide a slower folder at 281 K than a similar peptide devoid of the three glutamate side chains. On the basis of this simulation result, extracted from a total MD sampling of 1.0 ms, a mutant with additional bulky side chains (three methionines replacing alanines at positions 2, 7, and 12) is suggested to fold slower than the EAAAR peptide. This prediction is confirmed by time-resolved IR spectroscopy.

## I. Introduction

Protein folding from the ensemble of denatured conformations to the native state is a complex transition because of the many degrees of freedom and interaction centers involved.[1] In fact, folding is driven by a delicate balance of van der Waals and electrostatic forces between atoms in the protein, and protein−solvent interactions. To tackle the complexity of the folding process synergistic combinations of experimental and computational studies have been used successfully, in particular to characterize the transition state ensemble of two-state proteins,[2] and novel techniques are being developed. On the experimental side, small and ultrafast folding proteins have been discovered and further engineered.[3,4] Recently, time-resolved IR spectroscopic techniques have been used to monitor folding kinetics of helical peptides with a cross-linker from the picosecond to microsecond time range.[5] For triggering helical folding, the conformation of the cross-linker (azobenzene) is "switched" from cis to trans by using light of a given wavelength.[6] On the computational side, approaches based on complex network[7] and graph analyses[8] have been recently introduced for analyzing long MD simulations, which become more and more affordable because of the ever-increasing computer power. These graph-theoretical approaches have played an important role in showing that the surprisingly simple two-state picture of protein folding, often obtained by projecting the free energy onto an arbitrarily chosen progress variable, is not consistent with the complexity of the actual free-energy surface.[9]

Here, we investigate by MD the folding kinetics of the cross-linked peptide Ac-EACAR$^5$EAAAR$^{10}$EAACR$^{15}$Q-$NH_2$ (abbreviated hereafter as EAAAR), and compare the simulation results with the available time-resolved IR data.[5] In a previous work we had analyzed the kinetics of the Ac-AACAR$^5$-AAAAR$^{10}$AAACR$^{15}$A-$NH_2$ peptide (abbreviated as AAAAR).[10] One major difference between AAAAR and EAAAR is that only the latter can form salt bridges, which in the fully α-helical state involve the three pairs of side chains $Glu_n-Arg_{n+4}$ ($n$ = 1, 6, and 11). Therefore, we focus here on the influence of the sequence on the folding process. To also investigate the effect of the temperature, MD simulations of EAAAR folding are performed at 281 K (50 runs of 4 $\mu$s each) and at 330 K (50 runs of 0.5 $\mu$s each). The runs are started from pre-equilibrated cis ensembles at each temperature. The folding process upon ultrafast switching is emulated in the MD simulations by sudden switch of the dihedral potential of the N═N bond in the middle of the photoswitchable cross-linker. Two recently developed analysis techniques are applied to shed light on the kinetics of EAAAR folding and analyze its free-energy surface: The network of the conformation space (CS) and kinetic grouping analysis. The former is a very useful tool for the analysis and illustration of the actual (unprojected) free-energy surface and its dynamic connectivity[7,9,11] while kinetic grouping analysis is a recently introduced approach for identifying free-energy basins in long MD simulations.[12] Nonequilibrium MD simulations of an 8-residue cyclized peptide immersed in dimethylsulfoxide have been published recently[13] but the focus of that work was on the contributions to the frequency shift, while pathways and kinetics of folding were not investigated.

The main motivation of the present MD study was to provide the atomic-resolution picture needed to elucidate the effects

* To whom correspondence should be addressed. E-mail: caflisch@bioc.uzh.ch. Tel.: +41 44 635 55 21. Fax: +41 44 635 68 62.
[†] Department of Biochemistry, University of Zurich.
[‡] University of Modena and Reggio Emilia.
[§] Department of Physical Chemistry, University of Zurich.

10.1021/jp810431s CCC: $40.75 
Published on Web 03/03/2009

of the sequence on the folding kinetics of α-helical peptides, which has recently emerged from the time-resolved IR measurements. In this respect, the following three questions are particularly relevant: How do bulky side chains and (non-native) salt bridges influence the kinetics of α-helix formation? Can the simulation results be used to derive a simplified but quantitative description of the free-energy surface and its sequence dependence? Is it possible to use the atomic-resolution picture obtained by the MD study to predict mutants with different folding rates? The present work shows that the network analysis of the MD simulations provide a description of the free-energy surface of α-helix formation that not only captures the complex folding behavior and sequence dependence but can also be used to make suggestions for guiding experiments. In fact, a mutant of EAAAR with three methionine side chains (replacing three alanines) is predicted to fold slower on the basis of the MD simulation results, and this prediction is validated a posteriori by additional time-resolved IR measurements.

## II. Methods

**A. Simulation Methods and Protocols.** ***Force Field and Implicit Solvation Model.*** All simulations and most of the analysis of the trajectories were performed with the program CHARMM;[14,15] the rest of the analysis was done with the program WORDOM,[16] which is particularly efficient in handling large sets of trajectories. All heavy atoms were considered explicitely as well as the hydrogen atoms bound to nitrogen or oxygen atoms (PARAM19 force field). For the nonbonding interactions the default cutoff of 7.5 Å was used to be consistent with the parameters of the nonbonding energy terms of the force field which were determined using this cutoff value. A mean field approximation based on the solvent accessible surface area was used to describe the main effects of the aqueous solvent.[17]

***Parametrization of the Cross-Linker.*** The atom types for the cross-linker atoms were derived from the PARAM19 amide backbone and phenyl ring of Phe. For the double-bond of the two nitrogen atoms between the rings of the photoswitch the parameters for the dihedral angles were taken from ref 18. The parameters for bonds, angles, and impropers were derived per analogy with the corresponding proteinaceous fragments in PARAM19. Before starting the production runs, short MD simulations of the cross-linker alone were analyzed visually and showed reasonable qualitative behavior, for example, limited out-of-plane fluctuations of the *N*-acetylaniline moieties.

***Dihedral Function.*** To emulate the photoswitching process (i.e., isomerization of the C−N=N−C dihedral) the energy term

$$E^{\mathrm{N=N}}_{\mathrm{dihedral}} = 21.6(1 + \cos(2\theta - 180)) + 4.36(1 + \cos 4\theta) \quad (1)$$

(black curve in Figure 1) was modified by removing the minimum at $\theta = 0°$ and using a force constant of 60 kcal/mol to preserve the shape of the function from 90° to 180°

$$E^{\mathrm{N=N}}_{\mathrm{dihedral,trans}} = 60.0(1 + \cos\theta) + 4.36(1 + \cos 2\theta) \quad (2)$$

(red curve in Figure 1). The Langevin dynamics simulations of folding were run with the modified dihedral term starting from snapshots saved during equilibrium cis runs. The isomerization took place within a few picoseconds.

***REMD Simulations of the Peptide with Cross-Linker in the Cis Conformation.*** The equilibrium ensemble of the peptide in the cis conformation of the cross-linker was sampled by a replica exchange MD (REMD)[19] simulation of six replica at temperature values of 281, 304, 330, 358, 388, and 420 K. Temperature exchange attempts were performed every 20 ps (10000 MD steps) as in previous implicit solvent REMD simulations of helical and extended peptides,[10,20] and the acceptance ratio ranged between 0.27 and 0.33. Each replica was sampled for 18 and 27 μs for EAAAR and EMAAR, respectively. From the 18 μs (27 μs) of REMD sampling at 281 K, 50 (100) snapshots saved along time intervals of constant length were selected as starting structures for the EAAAR (EMAAR) folding runs at low temperature (see later and Table 1). The same procedure was used to select, from simulations at 330 K, the 50 and 130 starting conformations of the folding runs at high temperature of EAAAR and EMAAR, respectively.

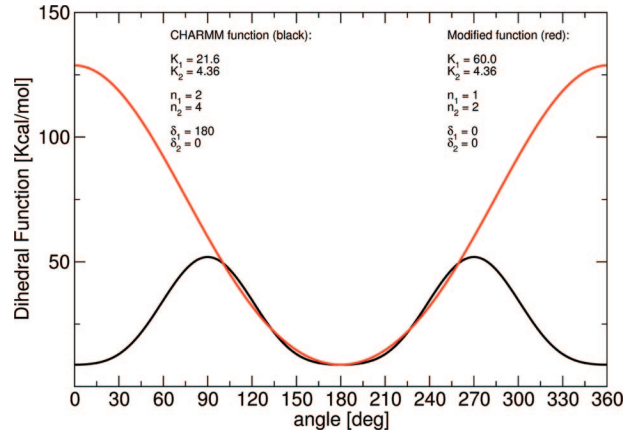



**Figure 1.** Emulation of the photoswitching by change in the dihedral energy term of the N=N bond. The dihedral energy term is $E^{\mathrm{N=N}}_{\mathrm{dihedral}} = K_1(1 + \cos(n_1\theta - \delta_1)) + K_2(1 + \cos(n_2\theta - \delta_2))$.

***MD Simulations of Folding.*** As in the time-resolved IR experiments, in the MD simulations the folding process is triggered by an ultrafast isomerization of the cross-linker within a few picoseconds. Therefore, the MD runs closely mimic the phototriggered α-helix formation. In a previous study we simulated the folding of the AAAAR peptide.[10] The same protocol is used here for the folding runs of EAAAR and EMAAR. At 281 K, 50 Langevin dynamics runs of 4 μs each were started from the equilibrium cis conformation of the EAAAR peptide previously sampled by REMD (see previous and Table 1). For the EMAAR peptide 100 runs of 8 μs each were performed at 281 K, and 130 runs of 0.5 μs each at 330 K. A friction coefficient of 1 $ps^{-1}$ was used in all simulations. This value is much smaller than the one of water (43 $ps^{-1}$ at 330 K) to allow for sufficient sampling within the μs time scale of the simulations. A time step of 2 fs was used and the coordinates were saved every 20 ps for a total of $2 \times 10^5$ snapshots for each 4-μs run. A series of 100 4-μs runs requires two months on a 100-CPU cluster (about 0.5 μs simulation time per week on a single Opteron 2.4 GHz CPU). Using explicit water simulations it would have been impossible to obtain the 1.0 ms of MD sampling (50 4-μs runs of EAAAR and 100 8-μs runs of EMAAR) required to collect a statistically significant number of folding transitions at low temperature, which is a necessary condition for the present analysis.

***Coarse-Graining.*** There are several ways for assigning snapshots (i.e., coordinate sets) to coarse-grained conformations (nodes and strings are used as synonyms in this paper) and different types of analysis might require different coarse-graining approaches.[7,21] The coarse-graining used in this work is based on secondary structure strings.[22] A conformation is a single string of secondary structure, for example, the most populated

**TABLE 1: Simulations Performed**

| | equilibrium cis state sampled by REMD | | folding runs[a] 281K | | 330K | | |
|---|---|---|---|---|---|---|---|
| | no. runs | length[b] [$\mu$s] | no. runs | length [$\mu$s] | no. runs | length [$\mu$s] | ref |
| AAAAR | 2 | 24 | 100 | 4 | 50 | 0.5 | 10 |
| EAAAR | 1 | 18 | 50 | 4 | 50 | 0.5 | this work |
| EMAAR | 3 | 27 | 100 | 8 | 130 | 0.5 | this work |

[a] An ensemble of snapshots, saved at time intervals of constant length along the REMD simulation segments at a given temperature, was selected as starting structures for the folding runs at that temperature. [b] These values refer to individual replicas and not to the cumulative time which is six times larger.

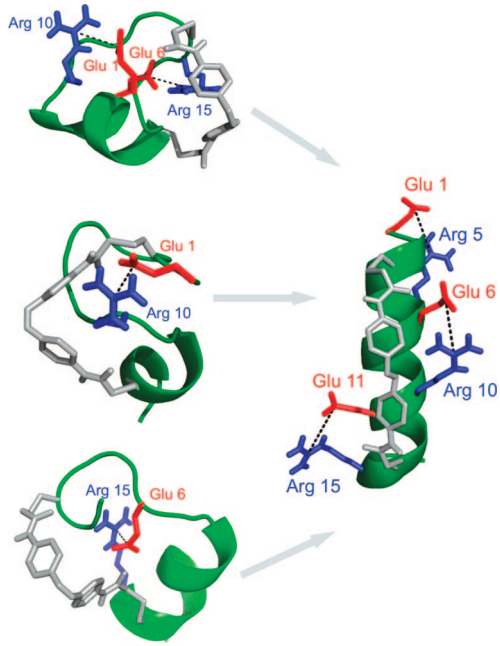



**Figure 2.** Folding of the cross-linked α-helical peptide EAAAR involves rearrangements of charged side chains. (Left) Three snapshots extracted from free-energy basins of the cis equilibrium simulations show non-native salt bridges. (Right) The most populated conformation with the cross-linker in the trans state is fully helical, and the three native, i.e., $n$, $n + 4$, salt bridges are populated at about 50%.

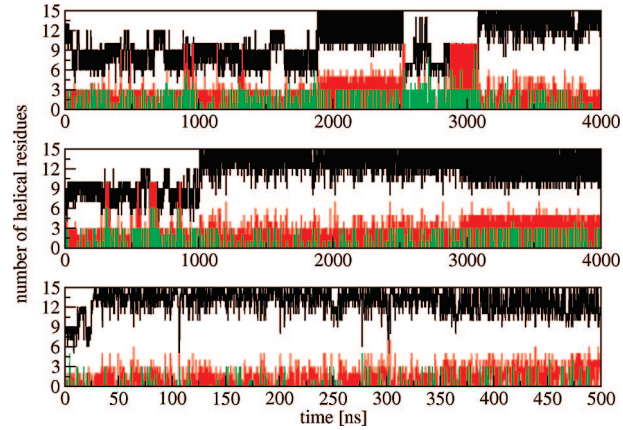
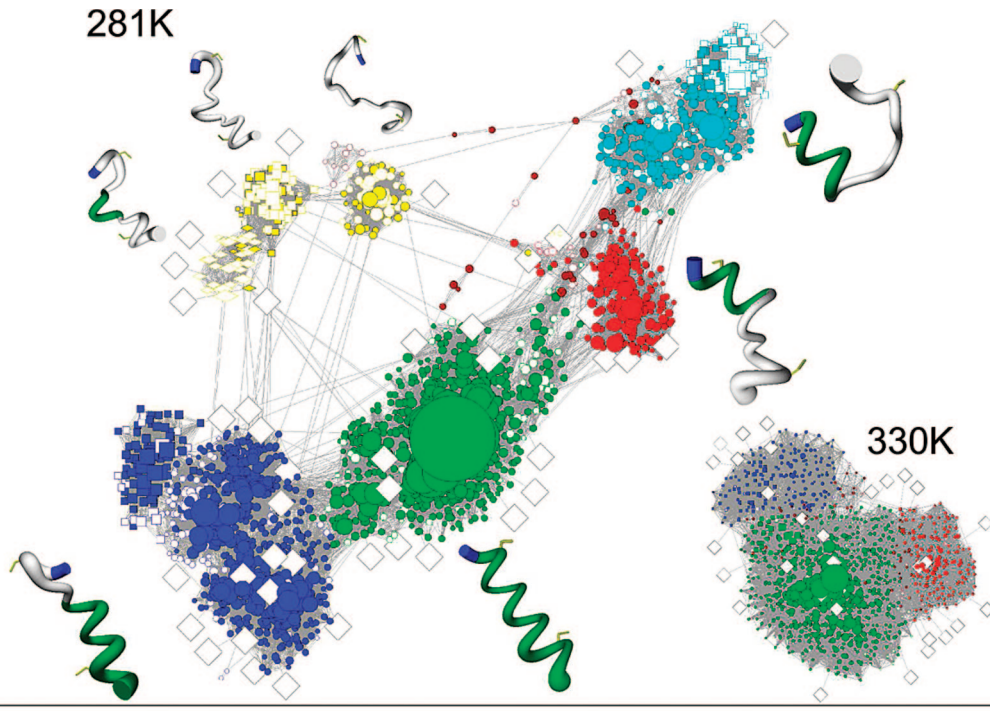



**Figure 3.** Increase of helical content in the MD simulations of the EAAAR peptide upon switching from the cis to the trans term of the dihedral energy of the N=N bond in the cross-linker. Time series of representative simulations are shown at 281 K (top and middle) and 330 K (bottom). The number of residues in stretches of α-helix, $\pi$-helix, and $3_{10}$-helix is in black, red, and green, respectively. The bottom time series shows only the first 500 ns to better resolve the folding event in the first 25 ns.

conformation of EAAAR with trans cross-linker is $-H_{15}-$. There are eight possible "letters" in the secondary structure "alphabet": "H", "G", "I", "E", "B", "T", "S", and "−", standing for α helix, $3_{10}$ helix, $\pi$ helix, extended, isolated $\beta$-bridge, hydrogen-bonded turn, bend, and unstructured, respectively. Since the N- and C-terminal residues are always assigned an "−",[22] an N-acetylated 16-residue peptide can in principle assume $8^{15} \simeq 10^{13}$ conformations. Note that the vast majority of these conformations are sterically forbidden. The secondary structure-based coarse-graining is appropriate for structured peptides without loops and has three advantages with respect to approaches based on root-mean-square deviation (rmsd) of atomic coordinates. First, it does not require the use of an arbitrarily chosen threshold value. Second, each node is uniquely defined by its secondary structure string which is a useful conformational "label". Third, the same type of secondary structure classification[22] is used for coarse-graining and analysis of structural properties (e.g., helical content in Figure 3).

***CS Network.*** Conformations (i.e., secondary structure strings) are nodes of the CS network and the direct transitions between them are links.[7] Given the number $\tilde{w}$ of snapshots with a unique secondary structure string, the statistical weight $w$ of a node is equal to $w = \tilde{w}/N$, where $N$ is the total number of snapshots sampled in the folding runs. At 281 K, the 20 ps saving frequency and 50 4-$\mu$s folding runs yield $N = 10^7$ for EAAAR. The CS networks are illustrated using only nodes with $\tilde{w} \geq 50$ to avoid overcrowding. Two nodes (in the CS network of "heavy" nodes) are connected by a link if they either include a pair of snapshots that are visited within 20 ps or they are separated by one or more nodes with less than 50 snapshots each. It is important to underline that the CS network of "heavy" nodes is used solely for illustrative purposes, whereas all of the quantitative analysis in this work, including kinetic grouping and fitting of average number of helical residues, was performed using the complete trajectory, that is, all snapshots and all nodes.

***Kinetic Grouping Analysis.*** In long MD runs the complete information about the native and non-native free energy basins is present in the trajectory. Therefore, peptide conformations can be grouped into free energy minima according to rapid transitions at equilibrium. This approach, called kinetic grouping analysis, has been developed previously and used to analyze the differences in the free energy surface of a 20-residue three-stranded $\beta$-sheet peptide and its single-point mutant Trp10Val,[12] both simulated at folding−unfolding equilibrium. Kinetic grouping analysis is based on the observation that if two conformations interconvert rapidly they are not separated by a barrier and therefore belong to the same basin. The method requires only one parameter, the commitment time $\tau_{commit}$, which is a typical relaxation time within the basins of the system. Different values of $\tau_{commit}$ allow one to analyze different levels of ruggedness of the free energy surface.[12] Values of $\tau_{commit}$ ranging from 1 to 20 ns were tested in this study. The final analysis was performed with $\tau_{commit} = 10$ ns and $\tau_{commit} = 1$ ns at 281 and 330 K, respectively, because these values allow for intrabasin relaxation and are much shorter than the time required for interbasin




| Most populated node in individual basins | Percentage of unfolded | $\tau_f$ (ns) Node | Basin | Helical content in basins H | H+G+I | color |
|---|---|---|---|---|---|---|
| -HHHHHHHHHHHHHHH- | – | – | 6 | 13.5 ± 1.7 | 14.7 ± 1.7 | green |
| -HHHHHHIII-IIIII- | 15 | 84 | 93 | 8.0 ± 2.8 | 13.0 ± 1.7 | red |
| -BSSB-HHHHHHHHHH- | 49 | 843 | 601 | 8.0 ± 2.0 | 9.5 ± 1.3 | blue circles |
| --TT--HHHHHHH-TT- | 5 | 389 | 710 | 6.9 ± 0.6 | 7.0 ± 1.2 | blue squares |
| -HHHHHHII-S-TTS-- | 17 | 1772 | 1480 | 5.8 ± 1.2 | 7.6 ± 1.1 | cyan circles |
| -IIHHHHHI-S-TTS-- | 6 | 3292 | 3260 | 4.2 ± 1.5 | 7.5 ± 1.7 | cyan squares |
| -TTT-S-TT-IIIII- | 4 | 922 | 840 | 1.9 ± 2.6 | 6.1 ± 1.9 | yellow circles |
| --S--GGGII-IIIII- | 3 | 540 | 525 | 2.9 ± 3.1 | 8.1 ± 1.6 | yellow squares |
| -BSSB-HHHH-IIIII- | 1 | 273 | 313 | 5.3 ± 3.2 | 8.1 ± 2.1 | yellow diamonds |

**Figure 4.** The CS network of the folding runs of EAAAR at 281 K shows multiple channels to the folded state (green nodes). Each node (i.e., conformation) of the network represents a secondary structure string and a link is a direct transition observed in the MD runs. The surface of each node is proportional to its statistical weight and only the 2218 nodes with at least 50 snapshots (98.7% of the total sampling) are shown to avoid overcrowding. Conformations sampled only in the folding runs and not in the trans equilibrium simulation are shown by white nodes with colored rims. White diamonds indicate the starting points of 45 of the 50 folding runs while the remaining 5 runs reached directly the most populated node and are not shown. The free-energy basins identified by kinetic grouping analysis[12] are shown with different colors or symbols, and their characteristics are listed in the table where folding times ($\tau_f$) are average values for snapshots in the most populated node of individual basins (column Node) or the entire basin (column Basin). Brown nodes were not assigned by kinetic grouping analysis. Representative structures of individual basins are shown by flexible tubes of variable diameter reflecting conformational disorder, with α-helical segments in green, N-terminus in blue, and cysteine side chains in yellow for emphasizing the position of the cross-linker. The CS network of the folding runs at 330 K is also shown (bottom right). This figure was made using visone (www.visone.de) and MOLMOL.[27]

transitions. The probability $p_{\text{commit}}(i \rightarrow j)$ to observe a transition from node $i$ to node $j$ within a given $\tau_{\text{commit}}$ is an asymmetric, directed measure of the *kinetic similarity* of nodes $i$ and $j$. Once the $p_{\text{commit}}$-matrix has been calculated for highly populated nodes (the 500 heaviest nodes in this work), pairs of nodes $(i, j)$ are grouped together if $p_{\text{commit}}(i \rightarrow j) \geq 0.5$. Lighter nodes are then assigned in a postprocessing step.[12] This way of grouping leads to a partitioning into disjoint sets, that is, a disconnected CS network, whose subgraphs correspond to different basins. Although the kinetic grouping analysis has been introduced to analyze long equilibrium simulations, it can be also applied to nonequilibrium simulations. In fact, since kinetic grouping analysis takes into account only the rapid, local interconversion between conformations within a basin, there is no requirement for global equilibrium sampling. Therefore, also basins visited during kinetic simulations can be isolated, provided that the system equilibrates locally.

**B. Experimental Methods.** The experimental setup for the time-resolved measurements has been described previously.[5,10,23]

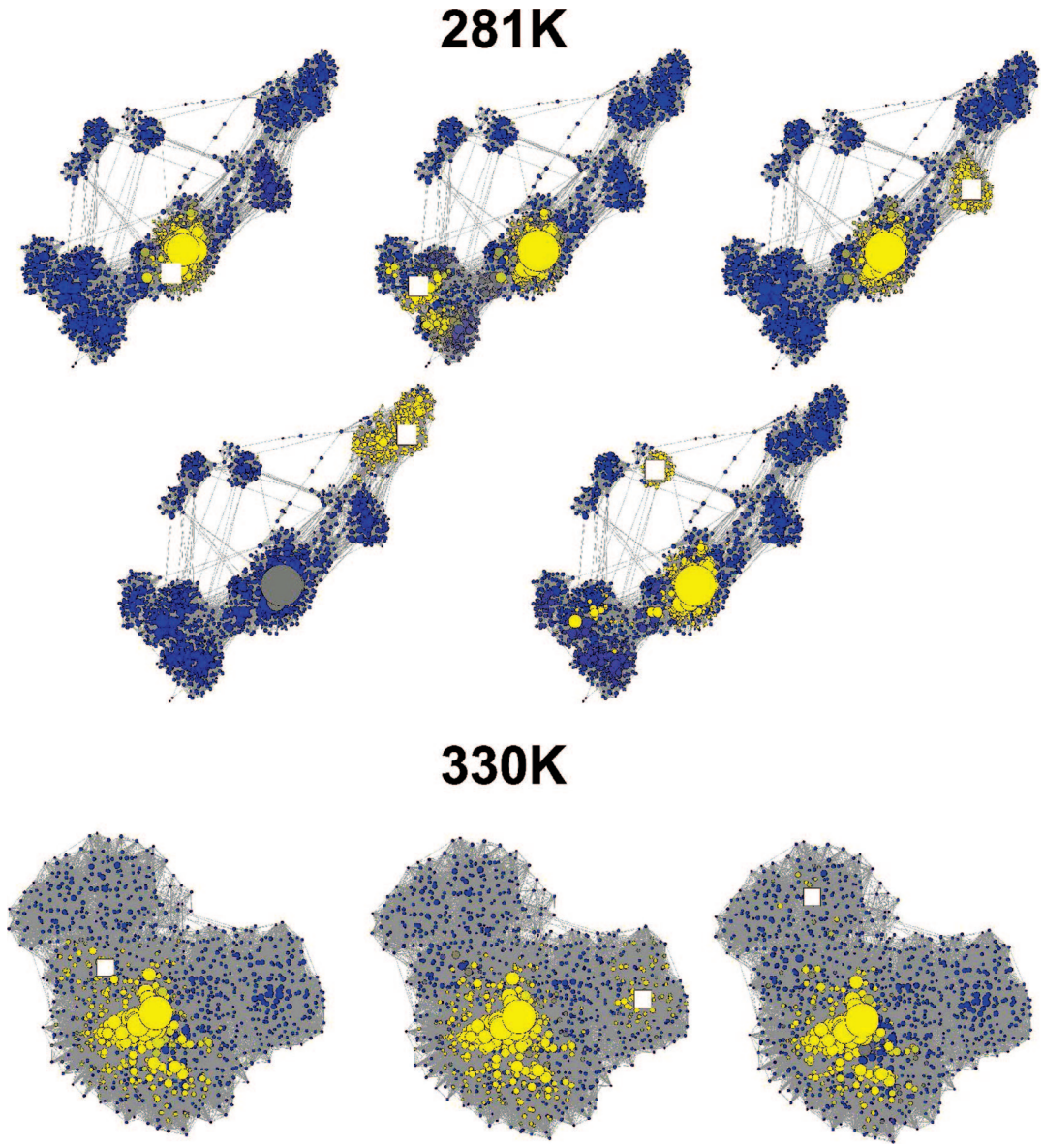



**Figure 5.** The unfolded state of the EAAAR peptide is kinetically partitioned. Nodes are colored according to mean first passage times[12] from the most populated node (white square) of individual free-energy basins to all other nodes in the CS network. (Top and middle) Folding runs at 281 K. The time scale ranges from 0 (yellow) to 4 $\mu$s (blue) except for the top, left network where the range is 0−1 $\mu$s. The 13th most populated node of the native basin was used as "starting node" in the top, left network. Most nodes within the basin of the starting node are visited relatively fast (yellow), indicating rapid intrabasin transitions and supporting the results of the kinetic grouping analysis (Figure 4). Visits to unfolded basins different from the starting one is much slower (blue) than reaching the folded state (olive or yellow) which shows that the unfolded state is kinetically partitioned. In other words, the folded state is a hub.[7,12,23] (Bottom) Folding runs at 330 K. The time scale ranges from 0 (yellow) to 0.1 $\mu$s (blue).

Briefly, two electronically synchronized femtosecond laser systems were used. The output of the one laser system was frequency doubled to generate pump pulses at 420 nm which switches the azobenzene from the cis to the trans state. The output of the second laser system was used to pump an OPA to obtain IR probe pulses (100 fs, center frequency 1620 $cm^{-1}$, bandwidth 240 $cm^{-1}$ fwhm), that could continously be delayed with respect to the pump-pulses from about 10 ps to 40 $\mu$s. The probe pulses were frequency dispersed in a spectrometer and imaged onto a 32 pixel HgCdTe detector array. A "folding signal" was generated from the amide I band, which frequency-shifts upon the strengthening of the backbone hydrogen bonds in the helical state. The sample was circulated in a closed-cycle flow cell optimized for small sample volumes. To prepare a well-defined initial condition, unfolded peptides with the azobenzene moiety in the cis-conformation were produced by constantly illuminating the sample with an excess of light at 366 nm, generated either by a cw-Argon-ion laser or a high-power LED.

## III. Results and Discussion

The equilibrium cis and trans ensembles of the EAAAR peptide are presented before the folding runs of which they represent the initial and final states, respectively (Table 1).

**A. Unfolded State (Cross-Linker in Cis Conformation) of the EAAAR Peptide.** In the equilibrium cis simulations, the helicity (i.e., sum of residues in $\alpha$-, $3_{10}$- or $\pi$-helical conformation as measured by the program DSSP)[22] of the EAAAR peptide is about 65% at both temperature values. Different conformations with partial helical content are observed but predominant structures are not present. At 281 K, several free energy basins are identified by kinetic grouping analysis and

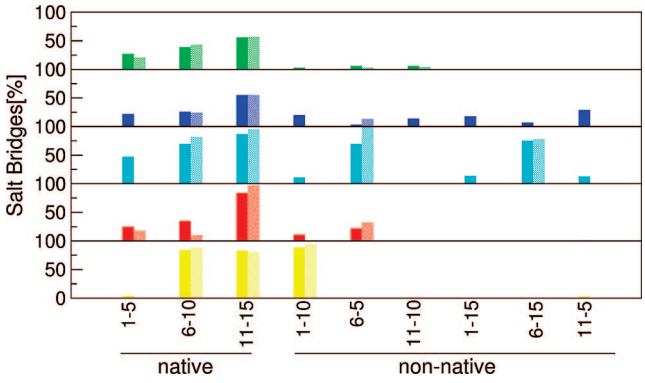



**Figure 6.** Salt bridge frequencies in individual free-energy basins of folding runs of the EAAAR peptide at 281 K. Values for entire basin and most populated node are displayed by filled and hatched bars, respectively. Colors correspond to those used in Figure 4.

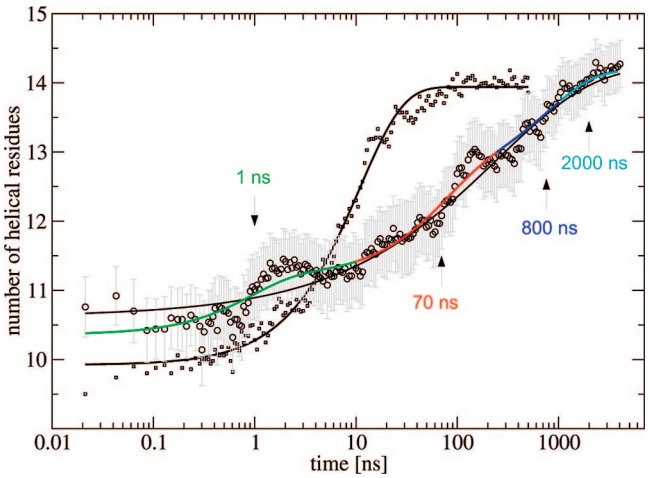



**Figure 7.** Kinetic traces of folding of the EAAAR peptide from MD simulations at 281 (circles) and 330 K (squares). The four-exponential fitting curve at 281 K [$14.2 - 0.8\exp(-x/1) - 1.4\exp(-x/67) - 1.4\exp(-x/755) - 0.2\exp(-x/1770)$, $\chi^2 = 3.3$] is colored according to the folding "channels" identified by kinetic grouping and CS network analyses, i.e., as in Figure 4, and arrows denote the four time constants. The stretched-exponential fitting of the data at 281 K [$14.3 - 3.9\exp(-x/185)^{0.40}$, $\chi^2 = 3.5$], and the single-exponential fitting of the data at 330 K [$13.9 - 4.0\exp(-x/11)$, $\chi^2 = 1.9$] are shown by black solid lines. Error bars are not shown for the data points at 330 K to avoid overcrowding and because the single-exponential fitting curve is always within the error bars.

only three of them have a statistical weight larger than 10%. The most populated basin (containing about one-quarter of the snapshots) has an unstructured N-terminal segment and two α-helical turns at the C-terminal segment. Individual basins have homogeneous distributions of helical content and salt bridges. In several basins non-native salt bridges are more frequent than the three salt bridges between residues $Glu_n$ and $Arg_{n+4}$ (with $n = 1$, 6, and 11) termed *native* salt bridges hereafter (Figure 2).

**B. Helical State (Cross-Linker in Trans Conformation) of the EAAAR Peptide.** In the equilibrium trans simulations, the helicity of the EAAAR peptide is about 90% and 80% at 281 and 330 K, respectively. These values can be compared with the helicity measured by CD: 93% and 64% at 281 and 322 K, respectively.[5] The agreement is good if one takes into account the 10% error of the CD measurements. On the other hand, the influence of the temperature is less pronounced in the MD simulations which is due mainly to the lack of explicit temperature-dependence in the solvation model. Therefore, in the comparison with the experiments we focus on qualitative trends rather than the quantitative comparison of absolute values.

The fully α-helical structure (secondary structure string $-H_{15}-$) is the most populated conformation (i.e., node) in the trans state of the EAAAR peptide with a statistical weight of 38% and 29% at 281 and 330 K, respectively. The corresponding free-energy basin has a population of about 90% at 281 K. In the fully α-helical state, the three native salt bridges are formed about 25−55% and 20−45% of the time at 281 and 330 K, respectively, while non-native salt bridges are sporadic (Supporting Information). Therefore, the main differences between the equilibrium cis and trans ensembles are that the former has a significantly less pronounced helical content and a much higher frequency of non-native salt bridges.

**C. Folding Kinetics and Complexity of Free-Energy Surface of the EAAAR Peptide.** At each temperature value, 50 snapshots saved with constant frequency along the corresponding equilibrium cis simulation were used as starting conformations for the kinetic runs (see Methods and Table 1). The time series of the helical content during the 281 K runs show that the folding transition takes place after time intervals of different length (Figure 3). Furthermore, folding at 330 K is much faster than at 281 K.

***CS Network Analysis Reveals Kinetic Partitioning.*** The network representation of the folding runs at 281 K shows the presence of multiple folding channels and time scales (Figure 4). Seventeen of the 50 folding runs start from the free-energy basin with unfolded N-terminal segment (blue region in Figure 4). The slowest folding channel starts from a free-energy basin with α-helical N-terminal segment and residues 7−16 unstructured (cyan basin). It is useful to color the CS network of the folding runs according to values of mean first passage time (mpft) from individual free-energy basins (Figure 5, top). Folding from the non-native basins proceeds directly to the fully helical state without passing through other basins. The barriers between individual free energy basins in the unfolded state and the fully helical state are lower than barriers within the unfolded state. The observation of kinetic partitioning of the unfolded state by kinetic grouping analysis provides further evidence to the fact that the native basin of EAAAR acts as a hub.[7,12,23] Such centrality of the native state has a significant influence on the folding kinetics and their complexity at low temperature (see also subsection Kinetic Models and Fitting of Helical Traces . These simulation results are consistent with recent experimental reports on competing folding routes and kinetic partitioning in ubiquitin[24] and the DNA-binding domain of p53.[25] Importantly, the present results provide evidence that folding pathways depend crucially on the unfolded state, that is, the starting ensemble. Therefore, in vitro folding experiments under strongly unfolding conditions (e.g., high concentration of chemical denaturants) do not necessarily give insights into the folding process under physiological conditions.

The CS network at 330 K consists of only the blue, red, and green (i.e., folded) basins (inset of Figure 4). Moreover, the unfolded state is kinetically partitioned (Figure 5, bottom), and the folding barriers are small and of similar height. The significant differences in folding kinetics at 281 and 330 K are indicative of the complexity of the system and the role of the entropy. Note that the implicit solvation model used in the MD simulations does not have an explicit temperature dependence so that only qualitative conclusions on temperature effects are possible.

***Salt Bridges.*** As mentioned above, the EAAAR peptide has Glu and Arg side chains at positions $n$ and $n + 4$, respectively, along the sequence. Strikingly, the stability of non-native salt bridges correlates with the length of the characteristic folding

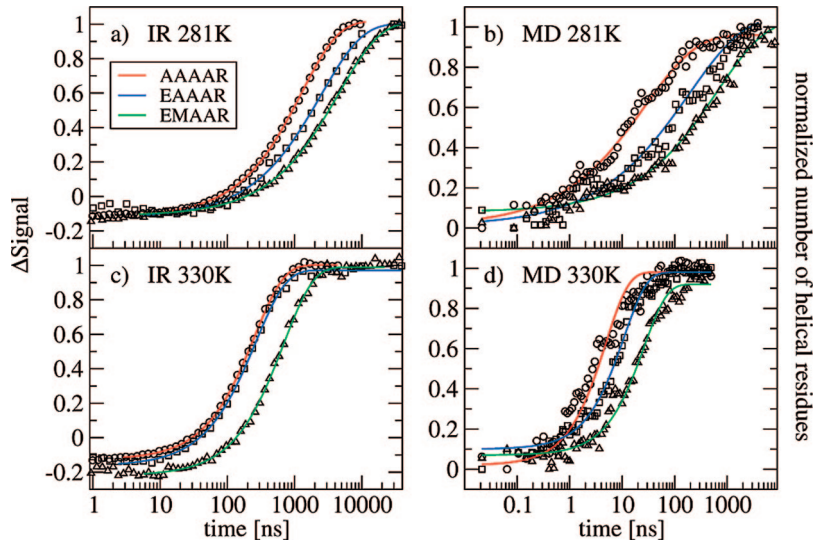



**Figure 8.** Comparison of normalized kinetic traces measured by time-resolved IR spectroscopy (a and c) and helical content, i.e., normalized number of residues in α, $\pi$, or $3_{10}$ conformation, along the MD simulations (b and d). Circles, squares, and triangles are used for AAAAR, EAAAR, and EMAAR, respectively. The fitting curve at 281 and 330 K is a stretched exponential (a and b) and a single exponential (c and d), respectively. The stretching factors at 281 K are in the range 0.4−0.5 and 0.7−0.8 for the MD and IR data, respectively. This discrepancy originates from the overestimation of the intrinsic barriers in the MD simulations, as previously observed in the large spread of folding rates of individual residues.[10] This overestimation is due in part to the predominance of intrasolute friction in the Langevin dynamics with implicit solvent and small friction coefficient (see Methods). Experimental data for AAAAR and EAAAR (a and c) are adapted from refs 5, 23.

times of the different channels (compare Figure 6 with Figure 4). In particular, the slowest channel, that is, folding from the cyan basin, is characterized by the stability of the non-native salt bridges $Glu_6-Arg_5$ and $Glu_6-Arg_{15}$, and two other slow channels by the presence of the salt bridges $Glu_1-Arg_{10}$ (yellow basin) and $Glu_{11}-Arg_5$ (blue basin). Moreover, visual analysis of some of the 281 K folding runs indicate that rearrangement of charged side chains is hindered by the presence of the atoms of the cross-linker. In this context, it is interesting to note that the Asp14Ala mutant of the five-helix bundle protein $\lambda_{6-85}$ was observed experimentally to fold faster than the wild-type[4] which is probably a consequence of the involvement of the wild-type Asp14 side chain in non-native salt bridges in the denatured state.

***Kinetic Models and Fitting of Helical Traces.*** It is interesting to monitor the average number of helical residues during the folding process and compare it to the time-resolved IR spectroscopy data. At 330 K, the MD kinetic trace can be fitted by a single exponential (Figure 7). On the other hand, single-exponential fitting is not appropriate at 281 K, and the data are best fitted by a multiple exponential fit or a stretched exponential. While the latter is in agreement with the time-resolved IR spectroscopy data,[5] the multiple exponential fit is consistent with the CS network, which shows multiple channels connecting individual unfolded basins to a hub-like helical state (Figure 4) and kinetic partitioning of the unfolded state (Figure 5).

Stretched exponential fitting is often preferred to multiexponential fitting because the former requires only two parameters (a time constant and a power of the stretching exponent) while the latter needs two parameters for each exponential term. On the other hand, if it is possible to relate the multiexponential behavior to a specific model, like the CS network of EAAAR at 281 K, then valuable insight into the folding process is obtained. Another interesting example of multiexponential fitting is the simple kinetic model used to describe the folding of ubiquitin.[26]

**D. Validation of the Atomistic Picture by Prediction of a Mutant Peptide.** The time series of helical content during folding of the EAAAR and AAAAR peptides can be fitted by a stretched exponential at 281 K (Figure 8b) with time constants of 200 ± 50 and 30 ± 2 ns, respectively. As shown in Figure 2, the folded (i.e., fully helical) state has a unique arrangement of the Glu and Arg side chains and their orientation with respect to the cross-linker. Since the unfolded state consists of basins with non-native salt bridges, the slower folding of the EAAAR peptide with respect to AAAAR is mainly due to rearrangement of salt bridge pairing and entanglement of bulky side chains with the cross-linker. For this reason, we thought that a mutant with additional "long" side chains would show slower kinetics of folding, and therefore we decided to investigate a peptide with three additional Met side chains, Ac-EMCAR[5]EMAAR[10]EMACR[15]Q-$NH_2$ (EMAAR peptide). One-hundred 8-$\mu$s simulations of folding of EMAAR from the equilibrium cis ensemble at 281 K confirmed our prediction (Figure 8b). In particular, the fitting of helical content by a stretched exponential yields a time constant of 650 ± 75 ns, which is slower than those of EAAAR and AAAAR. Note that the slower folding of EMAAR is not due to interactions between Met side chains, because the $\pi$-helical arrangment, which would bring the $n, n+5$ side chain pairs in close contact, is sporadic (Supporting Information).

To obtain an experimental validation of our prediction and the MD results, the three-point mutant EMAAR was synthesized (with exactly the same cross-linker as in AAAAR and EAAAR) and its photoswitched folding was investigated by time-resolved IR spectroscopy. At 281 K the folding of EMAAR is slower than EAAAR according to the IR traces in agreement with the MD simulation results (Figure 8a,b). The folding rates obtained by stretched-exponential fitting of the IR kinetic traces at 281 K are 4600 ns, 2500 ns, and 1300 ns for EMAAR (this study), EAAAR (ref 5), and AAAAR (ref 23), respectively. Note that at high temperature the kinetics are much faster and single exponential according to both MD simulations and time-resolved IR spectroscopy traces (Figure 8c,d). Furthermore, the rank-order of the time constants is the same as at 281 K in both simulations and experiments except for the AAAAR and EAAAR peptides that have essentially identical kinetic traces at high temperature in the experiments. In conclusion, the EMAAR mutant

was designed on the basis of the MD analysis of the AAAAR and EAAAR peptides to further investigate the mechanism of folding and the role of the sequence. Notably, the time-resolved IR data on the cross-linked α-helical peptides are consistent with and can be interpreted by the detailed picture that has emerged from the atomistic simulations.

## IV. Conclusions

Multiple MD simulations of the folding of the cross-linked helical peptide EAAAR have been performed to shed light on pathways and kinetics, with the aim of obtaining a detailed (i.e., atomistic) interpretation of the available time-resolved IR spectroscopy data. The cross-linker restricts the accessible conformational space. Yet, the following three observations provide strong evidence that (a large part of) the complexity of protein folding is still present. First, with respect to an unlinked peptide, folding of the cross-linked peptide might resemble more the structuring of a protein segment in the context of the remaining of the polypeptide chain.[10] Second, there is a significant dependence of the folding kinetics on the temperature in both MD simulations and time-resolved IR spectroscopy measurements. Third and most importantly, different sequences show different folding rates indicating that the primary structure has a strong influence on the free-energy surface.

Previous MD simulations have shed light on the complex kinetics of folding of the cross-linked AAAAR peptide,[10] which in contrast to EAAAR cannot form salt bridges. The multiple folding routes and kinetic partitioning of the unfolded state, which were both revealed by the CS network and kinetic grouping analyses, suggested that the kinetics of helix folding are determined by the exiting rate(s) from traps rather than nucleation of the first helical turn. At low temperature different barrier heights were observed for different folding channels, which explained the multiexponential kinetics. The role of the side chains was not investigated.

Here, the folding kinetics of the cross-linked EAAAR peptide and their temperature dependence observed in the MD runs are not only consistent with the time-resolved IR spectroscopy data[5] but provide atomic detail explanations of the complex kinetics. In particular, analysis of MD simulations unmasks the important role of bulky side chains and non-native salt bridges. The charged side chains involved in non-native contacts in the unfolded state have to rearrange because in the fully α-helical state only the three salt bridges between residues $Glu_n$ and $Arg_{n+4}$ are stable. This rearrangement is hindered by the presence of the cross-linker. On the basis of this observation it was predicted that the EMAAR mutant, which has three additional Met side chains, should fold even slower than the EAAAR peptide. This prediction was substantiated by a comparison of the MD simulations of EAAAR and EMAAR, and validated experimentally.

In conclusion, a detailed picture of the folding mechanism has emerged from the MD simulations of cross-linked α-helical peptides and has been used to interpret the kinetic traces measured experimentally. The combined MD simulation and time-resolved IR spectroscopy study of the three peptides indicates that non-native interactions among the charged side chains and entanglement of bulky side chains with the cross-linker are responsible for the stretched exponential kinetics at low temperature and rank order of folding rates (AAAAR faster than EAAAR faster than EMAAR). Note that the cross-linked peptides used in the present study represent the situation of a helical segment inside a protein because the cross-linker reduces the backbone flexibility, and the interactions between the side chains and the cross-linker reflect the tertiary contacts between the helical segment and other parts of the protein, respectively. Since the three peptides AAAAR, EAAAR, and EMAAR have the same cross-linker, the "tertiary contacts" have a stronger influence on the folding kinetics than the reduced backbone flexibility.

**Acknowledgment.** We thank Stefanie Muff for interesting discussions and comments to the manuscript. We also thank our co-workers of refs 5 and 23, Jens Bredenbeck, Jan Helbing, Rolf Pfister, and Andrew Woolley, who significantly contributed to the experimental data of AAAAR and EAAAR (Figure 8a,c). We are grateful to Armin Widmer for the program WITNOTP which was used for visual analysis of the MD trajectories. The MD simulations were performed on the Matterhorn cluster of the University of Zurich, and we gratefully acknowledge the support of C. Bolliger and A. Godknecht. This work was supported by Swiss National Science Foundation grants to A.C. and P.H.

**Supporting Information Available:** Formation of native and non-native salt bridges. This material is available free of charge via the Internet at http://pubs.acs.org.

## References and Notes

(1) Karplus, M. *J. Phys. Chem. B* **2000**, *104*, 11.
(2) Daggett, V.; Fersht, A. R. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 497.
(3) Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sat, S.; Jas, G. S.; Freund, S. M. V.; Alonso, D. O. V.; Daggett, V.; Fersht, A. R. *Nature* **2003**, *421*, 863.
(4) Yang, W.; Gruebele, M. *Nature* **2003**, *423*, 193.
(5) Bredenbeck, J.; Helbing, J.; Kumita, J. R.; Woolley, G. A.; Hamm, P. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2379.
(6) Kumita, J. R.; Smart, O. S.; Woolley, G. A. *Proc. Natl. Acd. Sci. U.S.A.* **2000**, *97*, 3803.
(7) Rao, F.; Caflisch, A. *J. Mol. Biol.* **2004**, *342*, 299.
(8) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14766.
(9) Caflisch, A. *Curr. Opin. Struct. Biol.* **2006**, *16*, 71.
(10) Ihalainen, J. A.; Paoli, B.; Muff, S.; Backus, E.; Bredenbeck, J.; Woolley, G. A.; Caflisch, A.; Hamm, P. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 9588.
(11) Gfeller, D.; De Los Rios, P.; Caflisch, A.; Rao, F. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1817.
(12) Muff, S.; Caflisch, A. *Proteins: Struct., Funct., Bioinf.* **2008**, *70*, 1185.
(13) Nguyen, P. H.; Gorbunov, R. D.; Stock, G. *Biophys. J.* **2006**, *91*, 1224.
(14) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(15) Brooks, B. R.; et al. *J. Comput. Chem.*; in press.
(16) Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. *Bioinformatics* **2007**, *23*, 2625.
(17) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins: Struct., Funct., Bioinf.* **2002**, *46*, 24.
(18) Carstens, H. Master's Thesis. LMU M374nchen, Faculty of Physics, 2004.
(19) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
(20) Cecchini, M.; Rao, F.; Seeber, M.; Caflisch, A. *J. Chem. Phys.* **2004**, *121*, 10748.
(21) Hubner, I. A.; Deeds, E. J.; Shakhnovich, E. I. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17747.
(22) Andersen, C. A. F.; Palmer, A. G.; Brunak, S.; Rost, B. *Structure* **2002**, *10*, 174.
(23) Ihalainen, J. A.; Bredenbeck, J.; Pfister, R.; Helbing, J.; Woolley, G. A.; Hamm, P. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 5383.
(24) Crespo, M. D.; Simpson, E. R.; Searle, M. S. *J. Mol. Biol.* **2006**, *360*, 1053.
(25) Butler, J. S.; Loh, S. N. *J. Mol. Biol.* **2005**, *350*, 906.
(26) Chekmarev, S. F.; Krivov, S. V.; Karplus, M. *J. Phys. Chem. B* **2006**, *110*, 8865.
(27) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics Modell.* **1996**, *14*, 51.

JP810431S