Complexity in Protein Folding: Simulation Meets Experiment

Amedeo Caflisch[1,*] and Peter Hamm[2,*]

[1]*Department of Biochemistry, University of Zürich, Switzerland;* [2]*Institute of Physical Chemistry, University of Zürich, Switzerland*

Abstract: We review our joint experimental-theoretical effort on the folding of photo-switchable α-helices. The folding kinetics of these peptides is profoundly non-exponential, which is attributed to a partitioning of the unfolded state into several misfolded traps. These traps are connected to the folded state in a hub-like fashion with folding barriers of different heights. Molecular dynamics simulations reveal a semi-quantitative agreement with the complex response observed in the experiment, allowing one to discuss the process in unprecedented detail. It is found that the non-exponential response is to a large extent introduced by the photo-linker used to initiate folding. Hence, folding of these cross-linked peptides emulates formation of a helical segment in the context of a globular protein rather than folding of an isolated peptide.

Keywords: Molecular dynamics simulations, Time-resolved infrared spectroscopy, Protein folding, Cross-linked alpha-helix, Misfolded traps, Free energy surface of folding.

INTRODUCTION

Protein folding is an inherently complex problem [1-7]. Undoubted, the many soft degrees of freedom of a polypeptide chain allows for an uncountable number of conformations (microstates). Nevertheless, many proteins seem to fold in a two-state fashion, that is, the free energy surface can be partitioned into two basins, the folded and the unfolded state, which are separated by one dominating barrier. Two-state folding implicitly assumes that inter-conversion of microstates within each of these basins is much faster than hopping across the folding barrier, so that the microstates in a basin are kinetically homogeneous. One of the strongest experimental tests of two-state folding is the comparison of kinetic and equilibrium data, ideally measured with different experimental methods. If the Chevron plots are linear, the kinetics are exponential with rates that agree with the equilibrium constants from unfolding transitions within experimental error, then using a kinetic model more complex than two states cannot be justified based on the data [8].

On the other hand, there is growing evidence from both molecular dynamics (MD) simulations [9,10] and experiments [11-13] that the free-energy surface of structured peptides and small proteins is more complex than a simple two-state model. On the experimental side, the problem often is the very limited amount of observables (typically only one, e.g., an overall frequency shift of the amide I vibration or the fluorescence lifetime of one chromophore), combined with the fact that it needs quite strong deviations from two-state behavior in order to obtain detectable non-exponential kinetics [14]. For example, in a kinetic scheme in which the folded state acts as a hub and various misfolded traps are connected to it in a star-like manner (see e.g. Fig. **11** below), it requires a spread of folding rates by a factor of ≈ 10 from the individual misfolded traps to obtain stretched-exponential kinetics (i.e. a kinetics following $\exp[-(t/\tau)^{\beta}]$) with a stretching factor of $\beta = 0.8$. Such a weak deviation from exponential kinetics requires a pretty good signal-to-noise ratio to actually be able to detect it. In equilibrium folding experiments, NMR spectroscopy provides sufficient number of observables to resolve complex behavior [11]. However, NMR spectroscopy is too slow for kinetic studies, at least for fast folders that can also be studied computationally. Infrared (IR) spectroscopy in combination with site-selective isotope labeling, on the other hand, is the method of choice since it is a fast spectroscopy (time resolution typically 1 ps) and at the same time reveals site-selective information for chemically identical molecules [12, 13, 15, 16]. This is the approach we used as well.

On the theoretical side, the problems are computer time limitations and the question of how to reduce the huge amount of information contained in a MD trajectory into one (or few) physically meaningful reaction coordinates. Despite significant advances on hardware and simulation methods it is still not possible to simulate the reversible folding of globular proteins so that the computational studies have focussed on structured peptides.

The number of degrees of freedom (about 3×200 for the helical peptides of about 15 residues investigated here) hinders a straightforward analysis of the MD trajectories. Three methods have emerged for reducing the complexity of the free-energy surface: network analysis [10], Markov state models [17, 18], and the minimum cut-based free-energy profile (cFEP) [19]. Applications of these procedures to MD simulations of reversible folding of structured peptides have demonstrated that the very simple picture of protein folding (e.g., one, at most two barriers between the denatured and

*Address correspondence to these authors at the Department of Biochemistry, University of Zürich, Switzerland; Tel: 0041 44 635 55 21; Fax: 0041 44 635 68 62; E-mail: caflisch@bioc.uzh.ch and Institute of Physical Chemistry, University of Zürich, Switzerland; Tel: 0041 44 635 44 31; Fax: 0041 44 635 68 38; E-mail: p.hamm@pci.uzh.ch

1877-9468/12 $58.00+.00 © 2012 Bentham Science Publishers

native state), often obtained by projecting the free energy on an arbitrarily chosen progress variable(s), is not consistent with the complexity of the actual free-energy surface [20-22]. The essential element common to network analysis, Markov state models, and the cFEP method is the identification of free-energy basins, not according to geometrical characteristics, such as the fraction of native contacts or radius of gyration, but rather according to the transitions sampled in (multiple, independent) MD simulations. The cFEP approach has the additional advantage that a meaningful one-dimensional projection of the free-energy surface is obtained. It provides the basins on the surface and the barriers between them. Unlike the standard projections, which lead to overlap of the basins that smooth out the barriers and can make them disappear (as shown for the β-hairpin of protein G in [19]), the progress variable used in the cFEP approach does not result in such overlap.

In a series of papers, we have addressed the protein folding problem in a combined experimental-computational approach, using photo-switchable α-helices as molecular systems [23-28]. This concept was initially introduced by Woolley and co-workers [29-31], who linked two sites of an Ala-rich sequence with high α-helical propensity by an azobenzene moiety. Light-induced ultrafast *cis-trans* isomerization poses a constraint to the peptide backbone that either stabilizes or destabilizes the helical content of the molecule [32, 33]. When the cross-linker is attached to two cysteines spaced 11 residues apart, the 3-turn α-helix conformation is stabilized by the *trans*-conformation of the azobenzene group, whereas in the *cis*-conformation the linker considerably diminishes the helicity of the peptide (see Fig. **1a**). Moreover, as the *cis* - *trans* isomerization of the linker is ultrafast [34], folding and unfolding can be triggered with high time-resolution [23, 35-37], and both direction can be studied independently by pumping the azobenzene moiety at different wavelengths. We have studied the following helical peptides: Ac-EACARE AAAREAACRQ-NH_2 [23] (abbreviated EAAAR in the following), Ac-AACARAAAARAAACRA-NH_2 (AAAAR) [24], Ac-AACAKAAAAKAAACKA-NH_2 (AAAAK) [25], and Ac-EMCAREMAAREMACRQ-NH2 (EMAAR) [26]. For these peptides, we obtained qualitatively similar results with slightly varying overall folding time [26].

The folding dynamics of isolated α-helices has been studied for the most part using laser induced T-jump methods to perturb the equilibrium, and applying either time-resolved IR spectroscopy [16, 38-41], fluorescence [42, 43], or Raman scattering techniques [44] as a probe. It is established that helix folding is not a two-state process [45], even though it is marginally cooperative. The overall relaxation of an α-helical peptide after an unfolding perturbation has been found to occur between $\approx$100 ns and 1 μs [16,23, 38-43]. Folding has successfully been modeled by so-called initiation-propagation models (or kinetic zipper-models) [42, 46, 47]. In these models, each residue of the

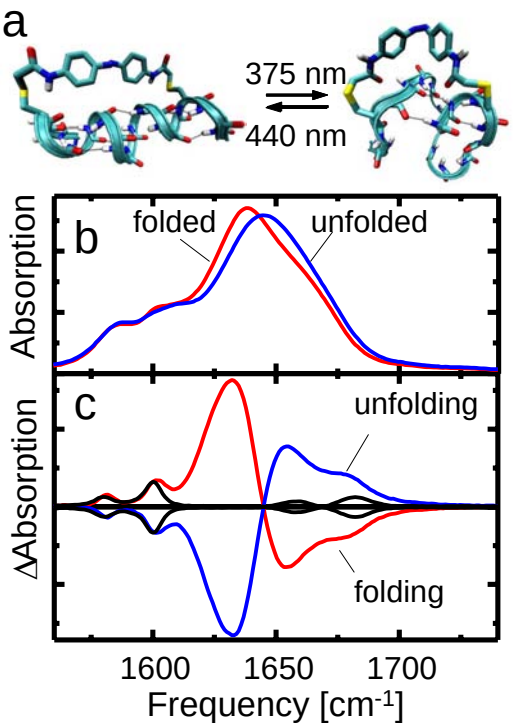

Fig. (1). (**a**) Example structures of the photo-switchable helix in the folded and the unfolded state, obtained from MD simulations. (**b**) Amide I′ absorption spectra of the AAAAR sequence and (**c**) difference spectra upon folding (red) and unfolding (blue). Adapted from Ref. [24] with permission.

polypeptide chain is considered to exist in one of two possible configurations -- coil-like or helical-like -- the latter having lower entropy because of the smaller number of possible conformations. Once the first helix turn has been formed, which is entropically costly, propagation of the helix is enthalpically favored due to the formation of hydrogen bonds. Consequently, the folding barrier is assumed to be entropic in nature. However, as we will see, initiation-propagation models break down completely for peptides cross-linked by the photo-switch. They fail because this photo-linker causes misfolded traps in the unfolded state, which are ignored in initiation-propagation models. Molecular dynamics simulations of small peptides [9,10,20,48,49] consistently demonstrate the existence of such free energy traps in the unfolded state, mostly due to non-native contacts. Hence, it turns out that the photoswitch adds complexity to α-helix folding that makes it comparable to that of a helical segment in the context of a globular protein. The system is thus ideally suited to study complexity in protein folding, since (a) it strongly deviates from two-state behavior, and since (b) folding is fast enough to connect the experimental results directly to all-atom MD simulations.

EXPERIMENTAL RESULTS

Fig. (**1b**) shows amide I′ spectra of the photoswitchable helix. When in the *trans* state, the photo-linker stabilizes the helical state, which results in an overall red-shift of the

amide I′ vibration. It is very well known that when a C=O mode gets involved in a hydrogen bond, it red-shifts by about 20 cm^{-1}, so the response of the amide I′ band reflect the in average stronger hydrogen bonds in the folded state. Fig. (**1c**) shows stationary difference spectra of an folding and an unfolding experiment, i.e., starting in the *cis* state (unfolded) and initiating the *cis-trans* isomerization by a 440 nm pulse, or *vice versa*. The amide I′ difference spectra are perfect mirror images of each other, proving that the conformational transition is fully reversible.

Figs. (**2** to **5**) show some of the key experimental results. Fig. (**2**) compares the folding and unfolding kinetics in an average sense by following the response of the amide I′ band at 1657 cm^{-1} as a function of temperature. Both processes accelerate with temperature since it is an activated process. The folding kinetics is essentially exponential at high temperatures but becomes stretched exponential at lower temperatures with a stretching factor of $\beta \approx 0.7-0.8$ (depending on the peptide sequence). The unfolding kinetics deviates much stronger from exponential with at least two phases, and it does so at all temperatures.

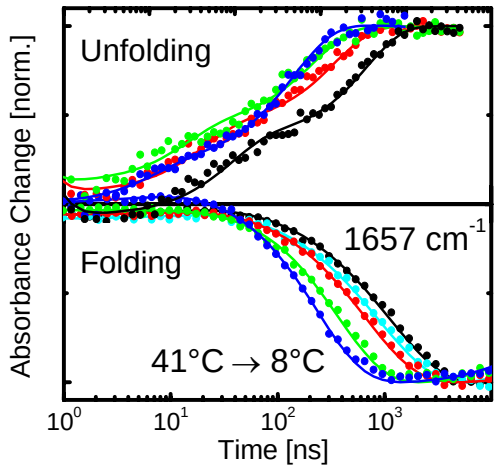

Fig. (2). Folding and unfolding kinetics of AAAAR measured at 1657 cm^{-1} for the temperature varied from 41 °C to 8 °C. Adapted from Ref. [24] with permission.

The source of the non-exponential response can be investigated using site-selective isotope labeling throughout the peptide sequence (one at a time). In this case, one observes essentially the formation of individual hydrogen bonds upon folding of the helix. It can be seen in Fig. (**3**) that individual sites form such hydrogen bonds on different time scales with rates that vary over almost one order of magnitude. Furthermore, the response of the individual sites is much closer (not exactly) to exponential. In fact, adding up the signals of the individual isotope labels, one regains the non-exponential response of the unlabeled amide I′ band (Fig. **4E**). Hence the response of the unlabeled amide I′ band is just a trivial average over essentially exponential processes, all of which with different time constants. Fig. (**4C**) summarizes these results as a function of temperature. At low temperatures, the individual rates differ strongly, but as the temperature is increased, they start to converge. A sum over individual processes, all of which with identical exponential rates, will of course still be exponential.

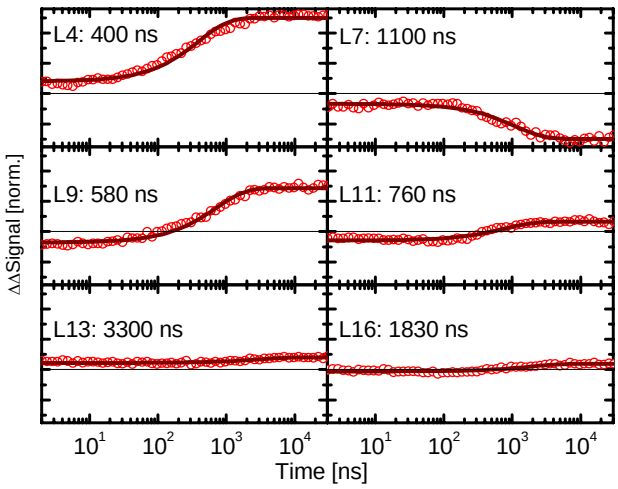

Fig. (3). Site specific folding signals of AAAAK at 19 °C for various isotope labels at positions 4, 7, 9, 11, 13, and 16. The time constants obtained from exponential fits are given. The folding signal of position 7 has opposite sign, indicating that its average hydrogen bond strength is weaker in the folded state. Adapted from Ref. [25] with permission.

It is interesting to note that one amino acid (site 7) actually shifts in the opposite direction (Fig. **3** and **4A**). This observation provides strong evidence that the unfolded state is not random, but consists of misfolded traps that are stabilized by hydrogen bonds. For site 7, it happens that these hydrogen bonds (in average) are stronger in the unfolded state than in the folded state. This becomes more apparent in 2D IR spectra [28], which better resolve the isotope-labeled carbonyl band (Fig. **5**). In this case, a doublet structure can be identified for that band which originates from the carbonyl group either being strongly hydrogen-bonded or essentially free. The intensity of the two sub-peaks shifts upon folding, however, the shift is opposite in L7 as compared to L4 (i.e., see the flip of the red-blue-red pattern for L4 in Fig. (**5**), right panel, into a blue-red-blue pattern for L7).

COMPUTATIONAL RESULTS

All MD simulations were carried out with a simple and very efficient implicit solvent model based on solvent accessible surface area [50]. To compare directly with the folding kinetics measured by time-resolved IR spectroscopy, out-of-equilibrium MD simulations were performed. Briefly, the system was first equilibrated in the *cis*-state by constant temperature MD. For each peptide sequence, 100 statistically independent snapshots were then taken from the *cis* ensemble and used as initial conformation for the out-of-equilibrium MD runs. These MD runs were started upon sudden switch of the N=N bond of the azobenzene moiety (switch of the minimum value of the corresponding dihedral energy term) as in the photochemical reaction. In this way,

the MD simulations generate an ensemble of folding trajectories. Atomistic simulations of short photo-switchable peptides have been performed by other research groups [35,51] but they did not report on the folding kinetics.

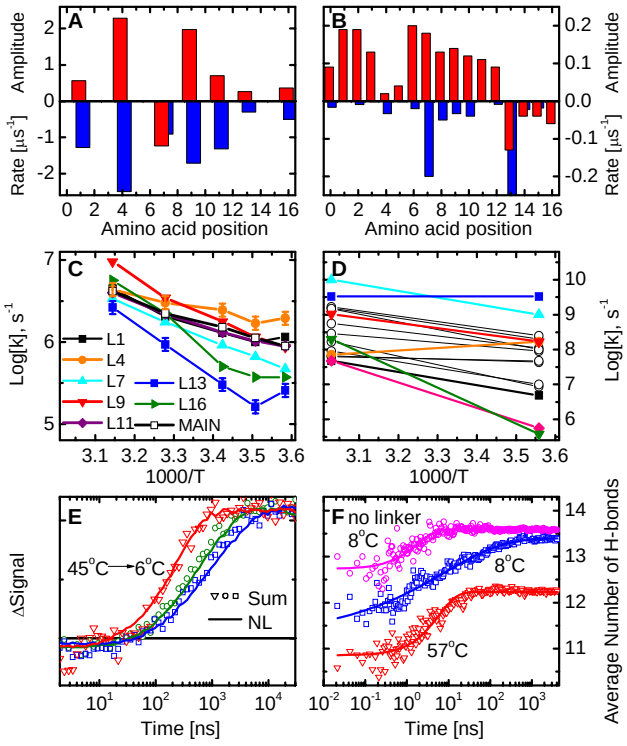

Fig. (4). Summary of the experimental (left column) and MD (right column) results. (**A**+**B**) Amplitudes (upwards, red) and corresponding rates (downwards, blue) at temperatures 19 ° C (**A**) and 8 ° C (**B**). Note that site 7 in the experiment, as well as sites 13-16 in the MD simulations, have inverted amplitudes. (**C**+**D**) Site selective folding rates as a function of inverse temperature. In (**D**) black circles are used for residues not measured experimentally. (**E**) Sum of all site signals (symbols) and amide I′ signal of the non-labelled peptide (solid lines) at 6, 19, and 45 ° C. (**F**) Average number of hydrogen bonds along the MD simulations at 8 ° C with stretched exponential fit (blue), at 57 ° C (red), and of the peptide without cross-linker at 8 ° C with single exponential fit (magenta). Adapted from Ref. [25] with permission.

The main results of the MD simulations are the kinetic partitioning of the unfolded state, the stretched-exponential behavior at low temperature, the sequence dependence of the folding rate, and the steric hindrance of the cross-linker. These observations are consistent with the picture of the folding kinetics that emerges from the time-resolved IR spectroscopy data. Importantly, the semi-quantitative agreement between experiments and simulations justifies the use of the latter to analyze the folding process at atomic level of detail.

Kinetic Partitioning of the Unfolded State

A complex network analysis [10] of the MD simulations of the EAAAR peptide shows that the unfolded state and folding pathways are much more heterogeneous at low than at high temperature (Fig. **6**). At 281 K, the subbasins in the unfolded state correspond to the fully unstructured peptide, unfolded N-terminal segment, and unfolded C-terminal segment (yellow, blue, and cyan basins in Fig. **6**, respectively). Notably, helix formation proceeds from each subbasin (or trap) in the unfolded state directly to the fully helical state which is a hub. In other words, the unfolded state is kinetically partitioned which emerges also from the coloring of the nodes according to the kinetic distance from the unfolded traps (Fig. **7**). Folding from each unfolded

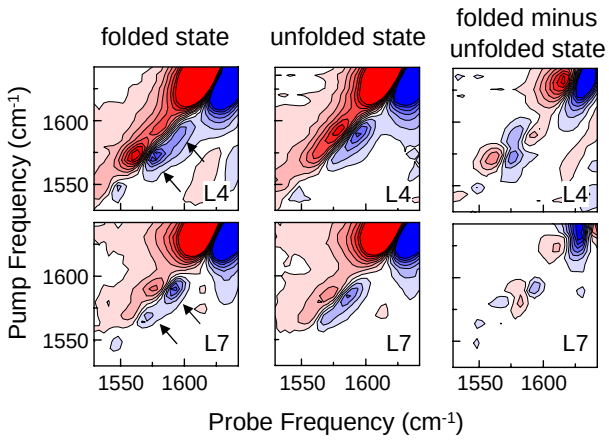

Fig. (5). Purely absorptive 2D-IR spectra of the AAAAK peptide isotope labeled at position 4 (top) and 7 (bottom) in the folded (left) and unfolded (middle) state taken at a temperature of 4 ° C. The right panel shows the spectra for the folded minus the unfolded state. Blue and red colors indicate negative and positive signals, respectively. The two arrows indicate the two bands originating from the isotope-labeled carbonyl band either being hydrogen bonded (lower frequency) of free (higher frequency). Adapted from Ref. [28] with permission.

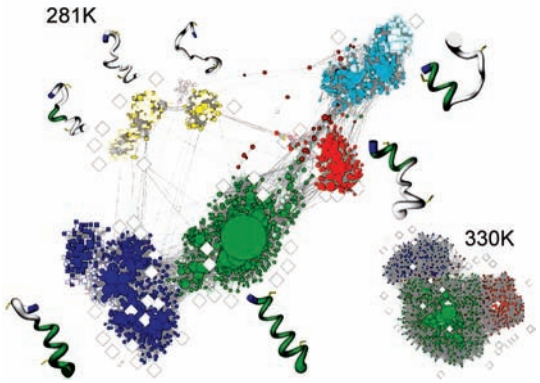

Fig. (6). Complex network analysis of the folding of the EAAAR peptide at 8 ° C and 57 ° C. Nodes and links are clusters of MD snapshots and transitions between them, respectively. The surface of each node is proportional to the number of snapshots within the node. Free-energy basins are emphasized by different colors with the fully helical state in green and unfolded states in yellow, cyan, red, and blue. The white diamonds are the starting points of the MD runs. Adapted from Ref. [26] with permission.

subbasin to the fully helical state is much faster than jumping between non-native subbasins. It has to be noted that folding from a kinetically partitioned denatured state is not consistent with the two-state behavior.

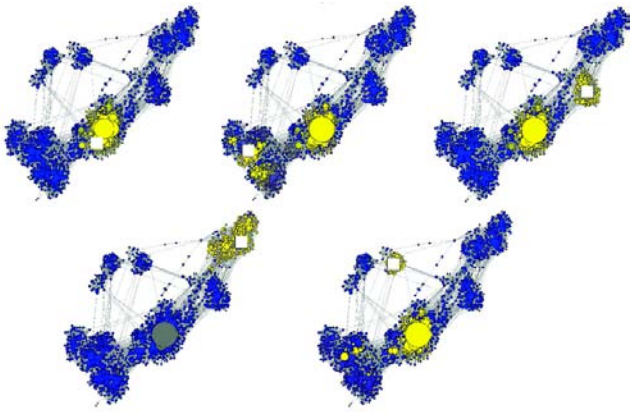

Fig. (7). The unfolded state of the EAAAR peptide is kinetically partitioned. Nodes are colored according to mean first passage times [52] from the most populated node (white square) of individual free-energy basins to all other nodes in the network. The timescale ranges from 0 (yellow) to 4 μ s (blue) except for helical state (network on top, left) where the range is 0-1 μ s. Most nodes within the basin of the starting node are visited relatively fast indicating rapid intrabasin transitions. 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. Adapted from Ref. [26] with permission.

Stretched-Exponential Kinetics of Folding and Sequence Effects

The kinetics of helix formation of the cross-linked peptides and the temperature dependence observed in the MD runs are consistent with the time-resolved IR spectroscopy data (Fig. **8**). At low temperature there is a high heterogeneity of the unfolded state and different barrier heights for different folding channels which explains the stretched-exponential time dependence, while at high temperature there are less substates in the unfolded state and the barriers to fold are of similar height which results in a single-exponential behavior (Fig. **6**).

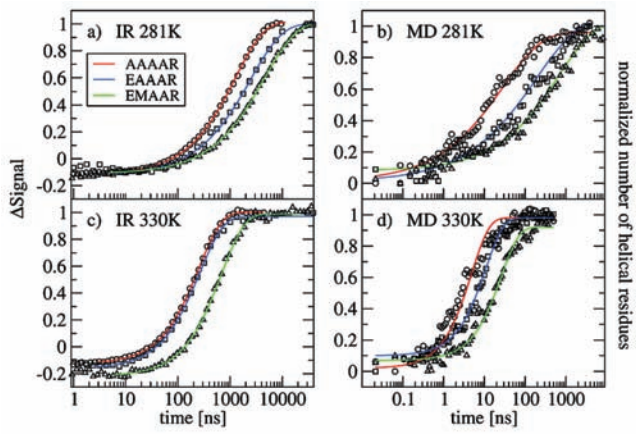

Fig. (8). Kinetic traces of three helical peptides measured by time-resolved IR spectroscopy (left) and monitored along the MD simulations (right). There is agreement in the rank-order of the folding rates as well as the time dependence which shows a stretched-exponential behavior at low temperature (top) and a single-exponential behavior at high temperature (bottom). Adapted from Ref. [26] with permission.

One main advantage of the MD simulations is the spatial resolution i.e., the atomic details of the molecular system. In this study, the in-depth analysis of the MD simulations has unmasked the important role of bulky side chains and non-native salt bridges [26]. 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. Based on this observation it was predicted that the mutant Ac-EMCAREMAAREMACRQ-NH2 (EMAAR), 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 (Fig. **8**) [26].

Linker Effects: Steric Encumbrance and Restraint on Backbone

The folding process of the cross-linked peptide mimics the formation of a helical segment of a protein with a twofold effect of the cross-linker. The azobenzene cross-linker reduces the flexibility at both ends of the segment emulating the remaining parts of the polypeptide backbone. Moreover, the interactions between the peptide side chains and the cross-linker reflect the tertiary contacts between the side chains in the helical segment and other parts of the protein, respectively. Thus, MD simulations with a nonbulky cross-linker (harmonic restraint on the Cys3-Cys14 distance) have been carried out to focus on the reduced flexibility of the backbone without any steric hindrance effects [27].

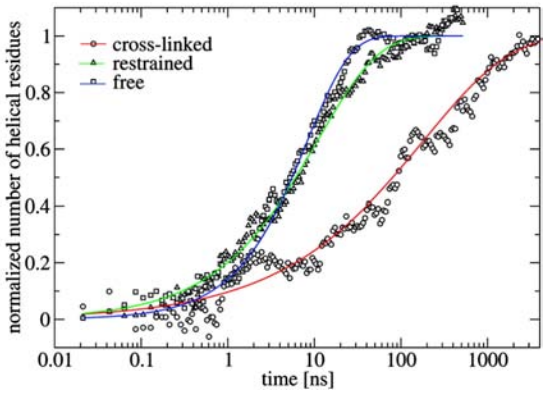

Fig. (9). Kinetic traces of helical content along the MD simulations. The plot shows the normalized difference in the number of helical residues (α -, π -, and 3_{10} -helix conformation) between initial and final states. Data points (symbols), from 50 folding runs for each system, are fitted with stretched-exponential functions (solid lines). The stretching factor is 0.4, 0.7, and 0.9 for the simulations with the cross-linker (red), nonbulky cross-linker (green), and free peptide (blue), respectively. Adapted from Ref. [27] with permission.

Two main observations have been reported. First, the folding kinetics with the distance restraint are almost as fast and simple (i.e., single-exponential) as for the free system, and both are much faster than in the presence of the azobenzene cross-linker (Fig. **9**). Thus, the slow kinetics of the cross-linked peptide originate mainly from the encumbrance of the atoms in the cross-linker rather than the restraining effect on the Cys3-Cys14 separation. In analogy, side chains rearrangements might slow down the folding of a helical segment in a globular protein, particularly if most of the protein assumes a compact conformation before folding of the helical segment.

Second, the helical ensemble of the cross-linked peptide consists of two free-energy subbasins separated by a barrier that reflects the slow transitions between helical conformers having opposite orientations of the Arg10 side chain with respect to the cross-linker (Fig. **10**). This observation indicates that the most populated state (r-orientation of Arg10, i.e., green basin in Fig. **10**) is only part of the helical ensemble. Therefore, simple geometric variables and experimental observables that monitor only the structure of the backbone (e.g., number of helical hydrogen bonds) do not capture the different subbasins within the fully helical ensemble.

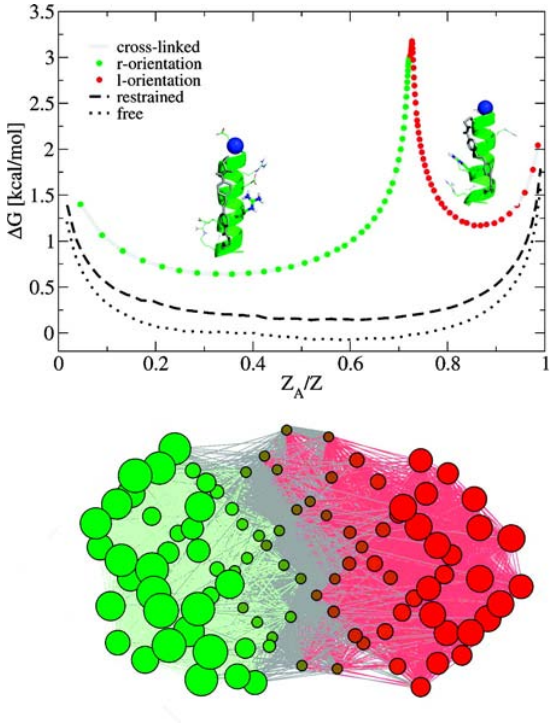

Fig. (10). The helical basin of the cross-linked EAAAR peptide consists of two subbasins. (Top) The cut-based free energy profile (cFEP) is a projection along the relative partition function Z_A/Z which preserves the barriers [19]. The MD snapshots were first clustered according to the value of $\cos(\theta)$, where θ is the angle that defines the orientation of the Arg10 side chain with respect to the cross-linker. The line with colored circles corresponds to the peptide with cross-linker, while the black dashed and dotted lines are the cFEP of the nonbulky cross-linker and free system, respectively. Adapted from Ref. [27] with permission.

DISCUSSION AND CONCLUSION

Qualitatively, the picture that emerges from the MD simulations is shown in Fig. (**11**). In this kinetic scheme, which naturally explains all experimental observations, the folded state (FS) is a hub connected to a set of misfolded traps (T1-T5) in a star-like manner (compare Fig. **11a** with Fig. **6**). In the folding experiment (Fig. **11a**), all traps are initially populated and feed into the folded states (FS). For the non-labeled amide I′ band, the response, therefore, will be a non-specific sum of all contributions that results in stretched exponential relaxation, provided the rates from the individual traps are different. The individual rates are better resolved when measuring the response in a site-selective manner by isotope labeling (Fig. **3**). In the unfolding experiment (Fig. **11b**), in contrast, the process starts from the better defined folded state FS. Initially, it will feed only into kinetically, and not necessarily thermodynamically, favored states (T1 in the example of Fig. **11b**). Only after longer times, a thermodynamic equilibrium will be established. Hence, the ensemble of trajectories will initially follow a relatively specific pathway, in qualitative agreement with bi-phasic response of the unfolding experiment (Fig. **2** top).

The average rate increases with temperature and at the same time, the non-exponential response disappears (Fig. **4C**-**F**). In the framework of the model discussed here, this implies that the rates from the individual traps approach each other at higher temperatures, and they do so in a surprisingly small temperature range (indicated by the dotted lines in Fig. **11d**). In conventional transition state theory, this would happen only for $T \to \infty$ (i.e. $1/T = 0$). However, the average rate we observe is only one-to-two orders of magnitude away from the so-called *speed limit* of protein folding $k^{(0)}$ (about $(20\ \text{ns})^{-1}$ to $(200\ \text{ns})^{-1}$ for peptides of this size) [53-55]. Hence, above a certain temperature, which we do not reach in our experiment, but which is not too far away, the individual rates saturate at that speed limit (see Fig. **11d**). The free energy surface becomes relatively homogeneous at high temperatures (see Fig. **6**, insert). This implies that the individual free-energy barriers are effectively lowered by entropy, since the pre-exponential factor extracted from an Arrhenius fit of both the folding and the unfolding experiment (in the order of 100 fs to 1 ps) are significantly larger than that speed limit. Consequently, the trapped states are in fact mis-folded states and the various transition states that connect the misfolded states to the folded state are open structures, or better to say the "only truly unfolded structures" (see Fig. **11c**). Obviously, these open structures would be transition states with high entropy and high enthalpy (quite different from the common initiation-propagation models).

The complexity of protein folding has been (partially) neglected because simplified pictures have dominated the literature during the past three decades. Experimentalists have introduced the concept of two-state behavior while theoreticians have provided the funnel diagram which is a pictorial representation. These simplified views of protein folding are useful as long as they do not generate

misconceptions [56]. As an example, the funnel diagram does not illustrate the presence of the folding free-energy barrier because it mainly depicts the downhill profile of the effective energy, which is the free energy (that contains a barrier) minus the entropic contribution. It is exactly the incomplete compensation of the loss of conformational entropy by the favorable, but at the transition state not fully established, intra-protein interactions, which yields the free-energy barrier for folding.

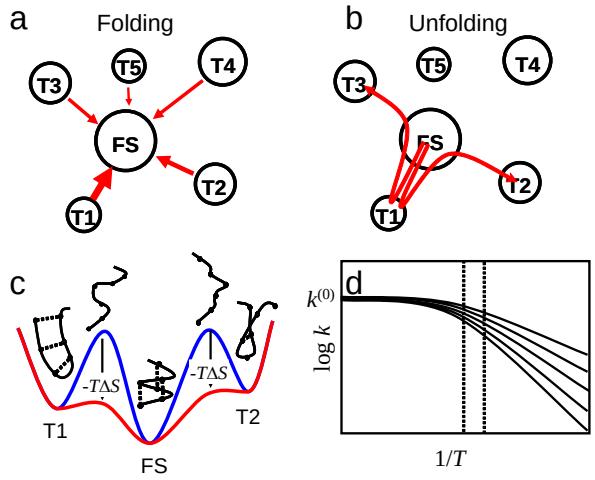

Fig. (11). Kinetic model with the folded state (FS) as a hub connecting various misfolded traps (T1-T5). (**a**) Flow of population in a folding experiment. The thickness of the arrows indicate the rate. (**b**) Flow of population in an unfolding experiment. Kinetically and not necessarily thermodynamically favored states (T1 in this case) are populated first. (**c**) Cut through the folding landscape explaining the temperature dependence. (**d**) Distribution of rates which leads to non-exponential response at low temperatures. The dotted lines indicate the small temperature range in which our experiments are done.

Two main conclusions emerge from our combined implicit solvent MD simulations and time-resolved IR spectroscopy analysis of the folding kinetics of cross-linked helical peptides. First, the free-energy surface of helical folding shows a hub-like character rather than a two-state profile. In other words, there are significant barriers within the unfolded state which is kinetically partitioned. A similar hub-like free-energy surface with an heterogeneous unfolded state, consisting of non-native enthalpic traps and entropically stabilized subbasins, has been observed in MD simulations of other peptides as well [10, 22, 57]. Second, the main effect of the cross-linker is its steric encumbrance. Therefore, instead of being a spurious (i.e., non-proteinaceous) element, it is the cross-linker itself that renders the folding of the photoswitchable peptides resembling more formation of a helical segment in the context of a protein rather than folding of an isolated helical peptide.

Finally, it has to be mentioned that the interest in protein folding has shifted in part towards protein aggregation since about ten years. Amyloid aggregation of about two dozens of (poly)peptide sequences is the cause of a range of incurable diseases, some of which are very frequent like Alzheimer's and Parkinson's diseases. We think that, as for protein folding, combined experimental-computational studies of peptide and protein (mis)folding and aggregation will play a key role in the understanding of these diseases at an atomic level of detail which might be essential to discover effective molecular-based therapies.

ACKNOWLEDGEMENTS

We wish to thank our coworkers in this project over the years, in particular Ellen Backus, Robbert Bloem, Jens Bredenbeck, Janne Ihalainen, Jand Helbing, Stefanie Muff, Beatrice Paoli, Riccardo Pellarin, Rolf Pfister, and Andrew Woolley. We thank Ben Schuler for interesting discussions. The work has been supported by the Swiss National Science Foundation (SNF).

REFERENCES

[1] Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins*, **1995**, *21*, 167-195.

[2] Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. Theory of Protein Folding: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.*, **1997**, *48*, 545-600.

[3] Brooks III, C. L.; Gruebele, M.; Onuchic, J. N.; Wolynes, P. G. Chemical physics of protein folding. *Proc. Natl. Acad. Sci. USA*, **1998**, *95*, 11037-11038.

[4] Karplus, M. Aspects of Protein Reaction Dynamics: Deviations from Simple Behaviour. *J. Phys. Chem. B*, **2000**, *104*, 11-27.

[5] Gruebele, M. Protein folding: the free energy surface. *Curr. Opin. Struc. Biol.*, **2002**, *12*, 161-168.

[6] Onuchic, J. N.; Wolynes, P. G. Theory of protein folding. *Curr. Opin. Struc. Biol.*, **2004**, *14*, 70-75.

[7] Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. The protein folding problem. *Annu. Rev. Biophys.*, **2008**, *37*, 289-316.

[8] Jackson, S. E.; Fersht, A. R. Folding of Chymotrypsin Inhibitor 2. 1. Evidence for a Two-State Transition. *Biochemistry*, **1991**, *30*, 10248-10435.

[9] Krivov, S. V.; Karplus, M. Hidden complexity of free energy surfaces for peptide (protein) folding. *Proc. Natl. Acad. Sci. USA*, **2004**, *101*, 14766-14770.

[10] Rao, F.; Caflisch, A. The protein folding network. *J. Mol. Biol.*, **2004**, *342*, 299-306.

[11] Sadqi, M.; Fushman, D.; Muñoz, V. Atom-by-atom analysis of global downhill protein folding. *Nature*, **2006**, *442*, 317-321.

[12] Brewer, S. H.; Vu, D. M.; Tang, Y.; L, Y.; Franzen, S.; 8 Raleigh, D. P.; Dyer, R. B. Effect of modulating unfolded state structure on the folding kinetics of the villin headpiece subdomain. *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 16662-16667.

[13] Brewer, S. H.; Song, B.; Raleigh, D. P.; Dyer, R. B. Residue Specific Resolution of Protein Folding Dynamics Using Isotope-Edited Infrared Temperature Jump Spectroscopy. *Biochemistry*, **2007**, *46*, 3279-3285.

[14] Chekmarev, S. F.; Krivov, S. V.; Karplus, M. Folding Time Distributions as an Approach to Protein Folding Kinetics. *J. Phys. Chem. B*, **2005**, *109*, 5312-5330.

[15] Huang, C.-Y.; Getahun, Z.; Wang, T.; DeGrado, W. F.; Gai, F. Time-Resolved Infrared Study of the Helix-Coil Transition Using 13C-Labeled Helical Peptides. *J. Am. Chem. Soc.*, **2001**, *123*, 12111-12112.

[16] Huang, C.-Y.; Getahun, Z.; Zhu, Y.; Klemke, J. W.; De- Grado, W. F.; Gai, F. Helix formation *via* conformation diffusion search. *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 2788-2793.

[17] Swope, W.; Pitera, J.; Suits, F. Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. *J. Phys. Chem. B*, **2004**, *108*, 6571-6581.

[18] Noe, F.; Fischer, S. Transition networks for modeling the kinetics of conformational changes in macromolecules. *Curr. Opin. Struc. Biol.*, **2008**, *18*, 154-162.

[19] Krivov, S. V.; Karplus, M. One-dimensional free-energy profiles of complex systems: Progress variables that preserve the barriers. *J. Phys. Chem. B,* **2006**, *110*, 12689- 12698.

[20] Caflisch, A. Network and graph analyses of folding free energy surfaces. *Curr. Opin. Struc. Biol.,* **2006**, *16*, 71- 78.

[21] Buchete, N.; Hummer, G. Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B,* **2008**, *112*, 6057-6069.

[22] Bowman, G. R.; Pande, V. S. Protein folded states are kinetic hubs. *Proc. Natl. Acad. Sci. USA,* **2010**, *107*, 10890-5.

[23] Bredenbeck, J.; Helbing, J.; Kumita, J. R.; Woolley, G. A.; Hamm, P. α-Helix formation in a photoswitchable peptide tracked from picoseconds to microseconds by time resolved IR spectroscopy. *Proc. Natl. Acad. Sci. USA,* **2005**, *102*, 2379-2384.

[24] Ihalainen, J. A.; Bredenbeck, J.; Pfister, R.; Helbing, J.; Chi, L.; van Stokkum, I. H.; Woolley, G. A.; Hamm, P. Folding and unfolding of a photoswitchable peptide from picoseconds to microseconds. *Proc. Natl. Acad. Sci. USA,* **2007**, *104*, 5383-5388.

[25] Ihalainen, J. A.; Paoli, B.; Muff, S.; Backus, E. H. G.; Bredenbeck, J.; Woolley, G. A.; Caflisch, A.; Hamm, P. α-Helix folding in the presence of structural constraints. *Proc. Natl. Acad. Sci. USA,* **2008**, *105*, 9588-9593.

[26] Paoli, B.; Seeber, M.; Backus, E. H. G.; Ihalainen, J. A.; Hamm, P.; Caflisch, A. Bulky side chains and non-native salt bridges slow down the folding of cross-linked helical peptides: a combined molecular dynamics and timeresolved IR spectroscopic study. *J. Phys. Chem. B,* **2009**, *113*, 4435-4442.

[27] Paoli, B.; Pellarin, R.; Caflisch, A. Slow folding of crosslinked α-helical peptides due to sterical hindrance. *J. Phys. Chem. B,* **2010**, *114*, 2023-2027.

[28] Backus, E. H. G.; Bloem, R.; Donaldson, P. M.; Ihalainen, J. A.; Pfister, R.; Paoli, B.; Caflisch, A.; Hamm, P. 2D-IR Study of a Photoswitchable Isotope-Labeled α-Helix. *J. Phys. Chem. B,* **2010**, *114*, 3735-3740.

[29] Rohl, C. A.; Baldwin, R. L. Deciphering rules of helix stability in peptides. *Meth. in Enzymol.,* **1998**, *295*, 1-26.

[30] Ferguson, N.; Fersht, A. Early events in protein folding. *Curr. Opin. in Struct. Biol.,* **2003**, *13*, 75-81.

[31] Gruebele, M. In *Protein Folding Handbook*; Buchner, J., Kiefhaber, T., Eds.; Wiley VCH: New York, 2005; pp 454-490.

[32] Kumita, J. R.; Smart, O. S.; Woolley, G. A. Photocontrol of helix content in a short peptide. *Proc. Natl. Acd. Sci. USA,* **2000**, *97*, 3803-3808.

[33] Flint, D. G.; Kumita, J. R.; Smart, O. S.; Woolley, G. A. Using an Azobenzene Cross-Linker to Either Increase or Decrease Peptide Helix Content upon *trans*-to-*cis* Photoisomerization. *Chem. Biol.,* **2002**, *9*, 391-397.

[34] Nägele, T.; Hoche, R.; Zinth, W.; Wachtveitl, J. Femtosecond photoisomerization of *cis*-azobenzene. *Chem. Phys. Lett.,* **1997**, *272*, 489-495.

[35] Spörlein, S.; Carstens, H.; Satzger, H.; Renner, C.; Behrendt, R.; Moroder, L.; Tavan, P.; Zinth, W.; Wachtveitl, J. Ultrafast spectroscopy reveals subnanosecond peptide conformational dynamics and validates molecular dynamics simulation. *Proc. Natl. Acad. Sci. USA,* **2002**, *99*, 7998-8002.

[36] Bredenbeck, J.; Helbing, J.; Sieg, A.; Schrader, T.; Zinth, W.; Renner, C.; Behrendt, R.; Moroder, L.; Wachtveitl, J.; Hamm, P. Picosecond conformational transition and equilibration of a cyclic peptide. *Proc. Natl. Acad. Sci. USA,* **2003**, *100*, 6452-6457.

[37] Chen, E.; Kumita, J. R.; Woolley, G. A.; Kliger, D. S. The kinetics of helix unfolding of an azobenzene crosslinked peptide probed by nanosecond time-resolved optical rotatory dispersion. *J. Am. Chem. Soc.,* **2003**, *125*, 12443-12449.

[38] Williams, S.; Causgrove, T. P.; Gilmanshin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. Fast Events in Protein Folding: Helix Melting and Formation in a Small Peptide. *Biochemistry,* **1996**, *35*, 691-697.

[39] Huang, C.-Y.; Klemke, J. W.; Getahun, Z.; De-Grado, W. F.; Gai, F. Temperature-Dependent Helix-Coil Transition of an Alanine Based Peptide. *J. Am. Chem. Soc.,* **2001**, *123*, 9235-9238.

[40] Petty, S. A.; Volk, M. Fast folding dynamics of an α-helical peptide with bulky side chains. *Phys. Chem. Chem. Phys.,* **2004**, *6*, 1022-1030.

[41] Werner, J. H.; Dyer, R. B.; Fesinmeyer, R. M.; Andersen, N. H. Dynamics of the Primary Processes of Protein Folding: Helix Nucleation. *J. Phys. Chem. B,* **2002**, *106*, 487-494.

[42] Thompson, P. A.; Eaton, W. A.; Hofrichter, J. Laser temperature jump study of the helix-coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model. *Biochemistry,* **1997**, *36*, 9200-9210.

[43] Thompson, P. A.; Muñoz, V.; Jas, G. S.; Henry, E. R.; Eaton, W. A.; Hofrichter, J. The helix-coil kinetics of a heteropeptide. *J. Phys. Chem. B,* **2000**, *104*, 378-389.

[44] Lednev, I. K.; Karnoup, A. S.; Sparrow, M. C.; Asher, S. A. α-Helix peptide folding and unfolding activation barriers: a nanosecond UV resonance Raman study. *J. Am. Chem. Soc.,* **1999**, *121*, 8074-8086.

[45] Schwarz, G.; Seelig, J. Kinetic properties and the electric field effect of life helixcoil transition of poly(γ-benzyl 9 L-glutamate) determined from dielectric relaxation measurements. *Biopolymers,* **1968**, *6*, 1263-1277.

[46] Naganathan, A. N.; Doshi, U.; Fung, A.; Sadqi.; Muñoz, V. Dynamics, Energetics, and Structure in Protein Folding. *Biochemistry,* **2006**, *45*, 8466-8475.

[47] Henry, E. R.; Eaton, W. A. Combinatorial modeling of protein folding kinetics: free energy profiles and rates. *Chem. Phys.,* **2004**, *307*, 163-185.

[48] Hummer, G.; Garcia, A. E.; Garde, S. Helix Nucleation Kinetics From Molecular Simulations in Explicit Solvent. *Proteins,* **2001**, *42*, 77-84.

[49] Mu, Y. G.; Nguyen, P.; Stock, G. Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principle Component Analysis. *Proteins,* **2005**, *58*, 343-357.

[50] Ferrara, P.; Apostolakis, J.; Caflisch, A. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins: Struct. Funct. Bioinform.,* **2002**, *46*, 24-33.

[51] Nguyen, P. H.; Stock, G. Nonequilibrium molecular dynamics simulation of a photoswitchable peptide. *Chem. Phys.,* **2006**, *323*, 36-44.

[52] Muff, S.; Caflisch, A. Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a beta-sheet miniprotein. *Proteins,* **2008**, *70*, 1185-1195.

[53] Bieri, O.; Wirz, J.; Hellrung, B.; Schutkowski, M.; Drewello, M.; Kiefhaber, T. The speed limit for protein folding mesured by triplet-triplet energy transfer. *Proc. Natl. Acad. Sci. USA,* **1999**, *96*, 9597-9601.

[54] Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci. USA,* **2000**, *97*, 7220-7225.

[55] Kubelka, L.; Hofrichter, J.; Eaton, W. A. The protein folding 'speed limit'. *Curr. Opin. Struc. Biol.,* **2004**, *14*, 76-88.

[56] Karplus, M. Behind the Folding Funnel Diagram. *Nature Chem. Biol.,* **2011**, *7*, 401-404.

[57] Krivov, S. V.; Muff, S.; Caflisch, A.; Karplus, M. One-Dimensional Barrier Preserving Free-Energy Projections of a beta-sheet Miniprotein: New Insights into the Folding Process. *J. Phys. Chem. B,* **2008**, *112*, 8701-8714.

Received: March 31, 2011 Revised: August 28, 2011 Accepted: September 21, 2011