

Defining the Nature of Thermal Intermediate in 3 State Folding Proteins: Apoflavodoxin, a Study Case

Rebeca García-Fandiño^{1,2}, Pau Bernadó³, Sara Ayuso-Tejedor^{4,5}, Javier Sancho^{4,5}, Modesto Orozco^{1,6*}

1 Joint IRB BSC Program in Computational Biology, Institute for Research in Biomedicine, Barcelona, Spain, **2** Department of Organic Chemistry and Center for Research in Biological Chemistry and Molecular Materials, Santiago de Compostela University, Santiago de Compostela, Spain, **3** Centre de Biochimie Structurale, INSERM U1054, CNRS UMR 5048, Université Montpellier 1 and 2, Montpellier, France, **4** Departamento de Bioquímica y Biología Molecular y Celular, Facultad de Ciencias, Universidad de Zaragoza, Zaragoza, Spain, **5** Joint Unit BIFI-IQFR, CSIC, Spain, Biocomputation and Physics of Complex Systems Institute (BIFI), Universidad de Zaragoza, Zaragoza, Spain, **6** Departament de Bioquímica, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain

Abstract

The early stages of the thermal unfolding of apoflavodoxin have been determined by using atomistic multi microsecond-scale molecular dynamics (MD) simulations complemented with a variety of experimental techniques. Results strongly suggest that the intermediate is reached very early in the thermal unfolding process and that it has the properties of an “activated” form of the native state, where thermal fluctuations in the loops break loop-loop contacts. The unrestrained loops gain then kinetic energy corrupting short secondary structure elements without corrupting the core of the protein. The MD-derived ensembles agree with experimental observables and draw a picture of the intermediate state inconsistent with a well-defined structure and characteristic of a typical partially disordered protein. Our results allow us to speculate that proteins with a well packed core connected by long loops might behave as partially disordered proteins under native conditions, or alternatively behave as three state folders. Small details in the sequence, easily tunable by evolution, can yield to one or the other type of proteins.

Citation: García-Fandiño R, Bernadó P, Ayuso-Tejedor S, Sancho J, Orozco M (2012) Defining the Nature of Thermal Intermediate in 3 State Folding Proteins: Apoflavodoxin, a Study Case. *PLoS Comput Biol* 8(8): e1002647. doi:10.1371/journal.pcbi.1002647

Editor: Emad Tajkhorshid, University of Illinois, United States of America

Received: January 23, 2012; **Accepted:** June 18, 2012; **Published:** August 23, 2012

Copyright: © 2012 García-Fandiño et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Spanish Ministry of Science and Innovation (BIO2009-10964, BFU2010-16296 and Consolider E-Science), DGA-B89-2011, Instituto Nacional de Bioinformática, Scalalife EU-Grant and Fundación Marcelino Botín. RGF also thanks the Spanish Ministry of Science and Innovation for her postdoctoral fellowship and Juan de la Cierva contract. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: modesto.orozco@irbbarcelona.org

Introduction

In addition to the folded and unfolded states, many proteins may adopt stable conformations that display mixed properties of the native and denatured states. These conformations, usually known as intermediates, may appear under unusual external conditions (i.e. non-physiological pH, pressure or temperature), in the presence of high concentrations of certain cosolutes (denaturants, salts), or as a consequence of mutations [1–9], and are supposed to be populated during folding (and unfolding), especially in the case of medium or large proteins [10]. Such folding intermediates may be on-pathway, facilitating the reaction, or off-pathway, acting as traps that may lead to misfolding and even aggregation [11–12].

The occurrence of equilibrium intermediates is often associated with stress phenomena and can trigger pathological effects, such as spongiform encephalopathy and other types of amyloidosis [13–14]. This explains the existence of many physiological mechanisms designed to reduce their harmful effects, mainly by reducing the life time of these potentially dangerous conformations [15–17]. For some proteins, however, physiological roles have been postulated for their intermediate conformations and such possibility might be more common than originally believed [8,18–21]. All these reasons explain the interest in understanding the nature of intermediates and the atomistic details that favours the transition

from native to intermediate structures. Unfortunately, the study of intermediate conformations is much more difficult than that of native forms. Equilibrium intermediates can be detected *in vitro* as a deviation from two-state behaviour, i.e., non-coincident protein unfolding curves obtained with different experimental techniques [22], although their structures and energetic properties are more difficult to probe. Folding intermediates are elusive to X-ray crystallography and their normally small population and the extensive signal broadening compared to the native state difficult their analysis by means of NMR techniques [23–28]. As a consequence, structural information on intermediates is often obtained by using low-resolution techniques, often based on Φ -analysis [29,30–31], or low-resolution spectroscopic or scattering data, which can give clues on the general shape of the protein, but not atomistic information. This explains the need and frequent use of simulation techniques, particularly molecular dynamics (MD) to try to gain atomistic details that are unreachable to experimental techniques [32–44].

Flavodoxins are a family of proteins essential for the survival of many human pathogens that has become one of the most studied models for protein folding and unfolding. They are mono-domain α/β proteins, with a parallel five-stranded β -sheet surrounded by five α -helices, and they carry a non-covalently bound FMN group which can be reversibly removed [45]. Several experimental studies on apoflavodoxins from the *Anabaena* [46–47], *Azotobacter*

Author Summary

A simplistic view of protein structure tends to emphasize the opposition between the native state and the denatured ensemble of unfolded conformations. In addition to these extreme conformations, proteins subjected to a variety of perturbations often populate alternative partly unfolded conformations, some of which are close in energy to the native state and, accordingly, can be populated under native or quasi-native conditions. There is increasing evidence that these “perturbed” conformations participate in protein function or, in some cases, are related to the outcome of folding diseases. We have used the “state of the art” molecular dynamics combined with a variety of experimental techniques to characterize for the first time, to our knowledge, the thermal intermediate of a three-state folding protein (apoflavodoxin). Based on our results we have been able to suggest a general mechanism of thermal unfolding in complex proteins and to determine interesting links between thermal intermediates and partially unfolded proteins.

and several *Desulfovibrio* [48–50] strains have demonstrated the thermal unfolding of this protein follows a three-state mechanism, where a partly unfolded intermediate accumulates at moderately high temperatures. Using a variety of techniques applied to wild type and mutant proteins Sancho’s group arrived to a low resolution picture of the thermal intermediate of the apoflavodoxin from *Anabaena* PCC 7119 [51–52] finding evidences that the intermediate is in fact close to the native structure, with the two hydrophobic cores well preserved, and with distortions probably located mostly in the loops and in one of the β -strands [53]. The overall dimensions of this thermal intermediate were characterized by small-angle X-ray scattering analysis, which suggests that the intermediate is slightly more extended than the native form, but clearly far from the expected situation of a random coil [54].

In this paper, we present a massive molecular dynamics (MD) effort for the study of the early stages of thermal unfolding of apoflavodoxin from *Anabaena* and for the characterization of its thermal intermediate. The study is especially challenging, since the slow folding dynamics of this protein (average transition times in the order of 10^1 – 10^2 millisecond makes impossible the use of pure force-approaches based on atomistic potentials, which would require second-scale trajectories. Furthermore, many attempts to use coarse-grained potentials failed to sample structures reproducing experimentally known intermediate properties and unfolding pathways, while equilibrium dynamics obtained from coarse-grained potential seems stiffer, but qualitatively similar to that expected from MD simulations (data not shown but available upon request). Accordingly, we decided to use a hybrid approach, based on the use of microsecond scale atomistic MD, supplemented by low-resolution spectroscopic and scattering data and previously derived Φ -analysis. The approach allowed us to characterize with atomistic detail the ensemble of conformations that define the intermediate as experimentally detected in melting experiments. With this synergistic approach the mechanism that drives the transition from native to intermediate and most likely the early stages of the thermal unfolding of the protein were explored.

Methods

Molecular dynamics simulations

The crystal structure of *Anabaena* apoflavodoxin deposited in the Protein Data Bank with reference 1FTG [51] was used as starting

conformation for our simulations. Crystallographic SO_4^{2-} was conserved and simulated together with the protein [in the X-ray structure of *Anabaena* apoflavodoxin (crystallized in high ammonium sulphate concentration), a sulphate ion is bound, mimicking the FMN phosphate, which opens the possibility that the native conformation in this region is a consequence of the binding of the ion], the rest of ions 24 Na^+ and 6 Cl^- which are needed to neutralize large values of electrostatic potential around the protein were added using CMIP calculations implementing Poisson-Boltzman potentials [55]. The resulting systems were then solvated by around 7600 TIP3P water molecules [56], partially optimized, thermalized to 300 K (Nose-Hoover thermostat) and equilibrated using our standard protocol [57], followed by additional 50 ns of post-equilibration. Ten randomly selected snapshots (separated by at least 1 ns) were selected from the last 20 ns of the equilibration trajectory to generate the starting coordinates of ten replicas of the protein in water at $T = 300$ K. To increase diversity in the ensemble of the native form velocities were randomized and each replica was re-equilibrated for 5 ns prior to 0.2 μs isothermic-isobaric production simulations ($T = 300$ K, $P = 1$ atm). The structure obtained at the end of the 50 ns equilibration at $T = 300$ K was heated slowly (0.5 ns) to 368 K and equilibrated at this temperature for additional 10 ns, followed by 2 μs simulation using isothermic-isobaric conditions ($T = 368$ K, $P = 1$ atm). Periodic boundary conditions and Particle Mesh Ewald calculations were used to deal with long-range effects [58]. RESPA (Multiple time step) [59] with a minimum time step of 1 fs was used in conjunction with RATTLE [60] algorithms for maintaining bonds involving hydrogen atoms at equilibrium distances.

Multi-microsecond trajectory at high temperature suggests that under the simulation conditions the unfolding trajectory reaches conformations which reproduce known properties of the thermal intermediate in less than 200 ns (see Results). Thus, to enrich our trajectories with the intermediate state we performed 50 independent simulations starting from 50 different snapshots of the solvated protein extracted every nanosecond during the first 50 ns of the long $T = 368$ K simulation. Velocities in each snapshot were randomized and after 5 ns re-equilibration the 50 independent trajectories were followed for 0.2 μs using identical simulation conditions, representing an aggregate time in the replicas of 12 μs . Such meta-trajectory was analyzed to determine the nature of the intermediate by confronting collected structures with experimental observables of the intermediate state.

All MD simulations were carried out with NAMD 2.6 [61–62] computer program using the CHARMM27 [63–64] force field using the *MareNostrum* supercomputer at the Barcelona Supercomputer Center.

Trajectory analysis

Snapshots were saved every picosecond and submitted to a large variety of analyses. Basic geometrical descriptors were determined using the *ptraj* module of AMBER9 [65–67], clustering was done in function of the RMSd of the clustered structures using the MMTSB Tool set [68] and representative structures of the clusters were determined as those closer to the centroid of each cluster. Secondary structure assignment and solvent accessibility of the representative structures of each cluster were calculated independently using the program PROCHECK [69]. Theoretical changes in the UV spectrum of the protein related to unfolding were determined by analysing the solvent accessible surface of the four Trp (SAS_{Trp}) and using four references: i) the crystal structure, ii) the ensemble obtained in MD simulations at room temperature, iii) four isolated Trp and iv) the protein after 50 ns of MD simulation at $T = 500$ K (where it reaches $\text{RMSd} > 15$ Å from X-

ray structure and all structural signatures are lost). SAS were computed using the NACCESS [70] program with standard values for protein and solvent particles.

Essential dynamics (ED) [71] was done to determine the nature of the easiest deformation movements in the native and intermediate states of the protein and to determine the overlap between the essential deformation modes of the protein and the native \leftrightarrow intermediate transition vectors. For this purpose covariance matrices were calculated for the native and intermediate ensembles (using a common reference system defined by the structurally conserved regions of the protein). Such covariance matrix was diagonalized to obtain a set of eigenvectors (the essential deformation modes) and the associated eigenvalues (the amount of variance associated to each eigenvector). The similarity between the essential space of native and intermediate was compared using Hess metrics [72–73] taking 50 eigenvectors as a common essential space (at least 90% of variance explained in each ensemble):

$$\gamma_{AB} = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n (v_i^A \cdot v_j^B)^2 \quad (1)$$

where n is the dimension of the essential space, A and B are two ensembles and v stands for the eigenvectors. Considering the relative size of the protein and the essential space, any $\gamma > 0.1$ signals a statistically significant similarity [74].

The relative similarity between two essential deformation spaces was computed using [73]:

$$\kappa_{AB} = 2 \frac{\gamma_{AB}}{(\gamma_{BB} \gamma_{AA})} \quad (2)$$

where the self-similarity indexes γ_{XX} were obtained by comparing two different parts of the ensemble. Relative similarity index corrects absolute metrics by the intrinsic noise of MD simulations. A κ value close or even greater than 1 indicates that considering the noise of the trajectories the two ensembles are identical.

The transition from the intermediate to the native states was obtained by taking the first eigenvector calculated by principal component analysis of a meta-ensemble obtained by mixing an equal number of snapshots of the intermediate and the native state. The overlap between the intermediate essential dynamics and the intermediate \rightarrow native transition vector was determined as:

$$Ov = \sum_{i=1}^n (r \cdot v_i^{\text{int}})^2 \quad (3)$$

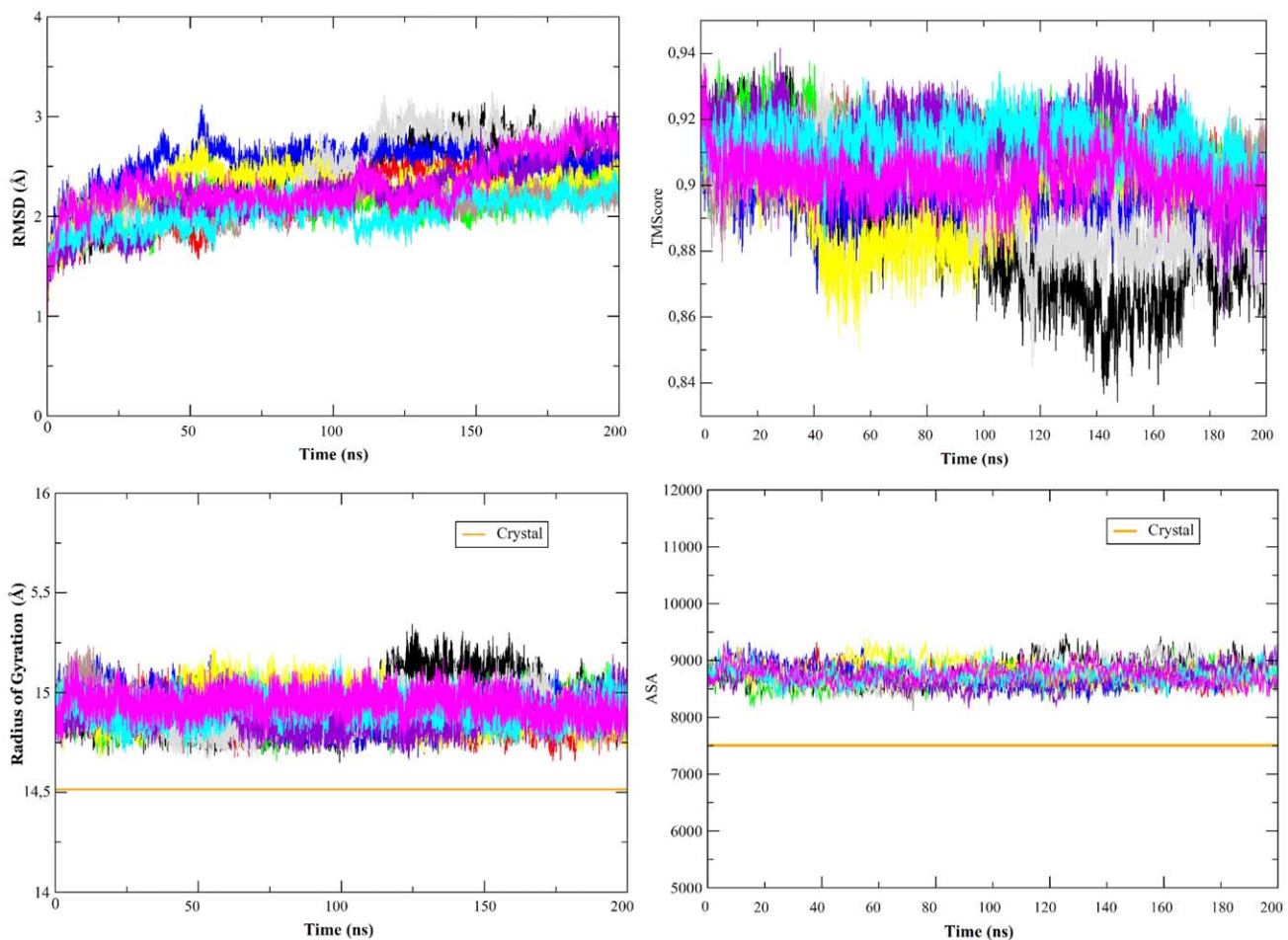


Figure 1. Evolution of different metrics along the 10 trajectories of apoflavodoxin at room temperature. TOP: All atoms root mean square deviation (LEFT) and TM-score^{44a} (RIGHT) from crystal structure (both in Å). BOTTOM: Radii of gyration (LEFT, in Å) and solvent accessible surface (RIGHT, in Å²). Orange lines represent the values found in the crystal structure (PDB code 1FTG). doi:10.1371/journal.pcbi.1002647.g001

where ov is the overlap (maximum equal to one), r is the transition vector and int stands here for the intermediate ensemble.

ϕ -values

Experimental ϕ -values profiles were taken from a previous work by Sancho's group [75,52,53]. Theoretical estimates were derived by individual ϕ_i^{calc} values (i stands for a residue) determined as the fraction of native contacts, N_i , made by that residue in the MD with respect to those found in the crystal structure, N_i^{nat} i.e., $\phi_i^{calc} = N_i/N_i^{nat}$ [76]. Comparison between experimental and simulated ϕ values was extended to all residues with $\phi_i < 1$ except for residues in helix 3, where experimental uncertainties in the determinations were large [75]. The ability of a structural ensemble to satisfy the experimental Φ -value profile was studied by analyzing the sum (over all residues) of the difference between predicted and simulated Φ -values:

$$error = \sum_i (\Phi_i^{exp} - \Phi_i^{teor}) \quad (4)$$

Small-angle X-ray scattering measurements and analysis

SAXS experiments were performed on the high brilliance beamline ID02 at the European Synchrotron Radiation Facility

(ESRF, Grenoble, France). An apoflavodoxin sample at 1 mg/ml concentration was prepared in 50 mM Mops buffer at pH 7. Several SAXS curves were acquired with a momentum transfer range of $0.07 < s < 0.31 \text{ \AA}^{-1}$ at a broad range of temperatures (6–67°C). Solutions were pushed in a capillary into the chamber where they were equilibrated for five minutes. An equivalent protocol was applied to measure buffer profiles. Ten successive frames of 1 s each were acquired for both sample and buffer. Each frame was inspected and the presence of protein damage was discarded. The different scans at each temperature were averaged and subtracted from their buffer counterpart using standard protocols with PRIMUS [77]. The forward scattering, $I(0)$, and the effective radius of gyration, R_g , was obtained from the scattering profiles using the Guinier's approximation [78] assuming that, at very small angles ($s < 1.3/R_g$), the intensity can be represented as $I(s) = I(0) \exp(-s^2 R_g^2/3)$.

SAXS curve measured at 26°C was used to evaluate MD trajectories in native conditions. The evaluation of trajectories in denaturing conditions was performed with the curve obtained from the Multivariate Curve Resolution by Alternating Least Squares (MCR-ALS) analysis of the SAXS dataset measured at the complete range of temperatures used to follow thermal denaturation of apoflavodoxin [54]. Principal Component Analysis (PCA) of the temperature variation SAXS dataset identified three components in the apoflavodoxin denaturation process that were assigned to the native the unfolded, and an intermediate states.

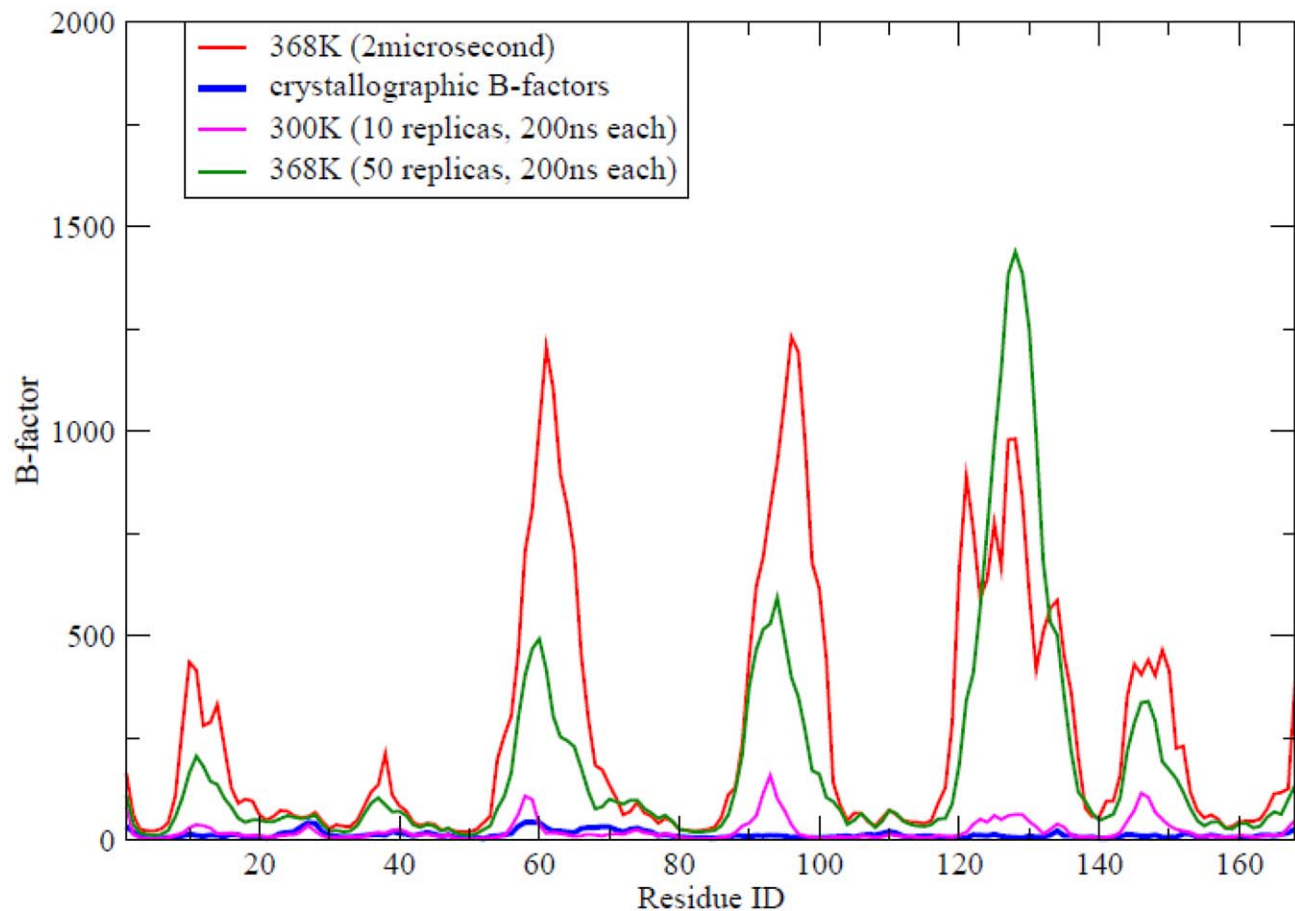


Figure 2. B-factors profile (Å) for apoflavodoxin obtained from simulations: i) the control meta-trajectory at 300 K, ii) the large unfolding trajectory at 368 K, iii) the meta-trajectory at 368 K and iv) the profile reported in the crystal structure (PDB code 1FTG). doi:10.1371/journal.pcbi.1002647.g002

MCR aims at finding the pure SAXS curves of these coexisting species in solution as well as the evolution of the relative concentration of these species upon environmental changes. The decomposition is obtained by solving the matrix equation

$$\mathbf{D} = \mathbf{C} \mathbf{S}^T + \mathbf{R} \quad (5)$$

where \mathbf{D} is the SAXS data matrix, \mathbf{C} is the matrix describing the contributions of the N components, \mathbf{S}^T is the matrix describing the instrumental responses of these N components, and \mathbf{R} accounts for the residuals of the fitting. Details of MCR-ALS approach and its application to SAXS data can be found in the original publications. [79–82]. Due to the post-processing nature of the SAXS profile of the intermediate, no experimental errors are associated to the derived intensities. A homogeneous 7% of error was assumed for each of the intensities of the curve. The agreement of SAXS profiles with three-dimensional structures of the MD trajectories was evaluated with CRY SOL [83] using default parameters. The χ -value of the fitting between experimental and theoretical curves is used as a measure of the quality of fitting (the smaller the χ -value, the better the agreement). Note that due to the de-convolution process and the use of a small homogeneous error in the intermediate, larger χ -values are expected in the fitting of the intermediate than to that of the native state.

UV spectra

Near-UV absorbance spectra of apoflavodoxin [51] at different temperatures were recorded from 250 to 310 nm in a *Chirascan* spectropolarimeter (from *Applied-Photophysics*) using 30 μM protein solutions in 50 mM Mops, pH 7 in a 4 mm path-length cuvette. The absorbance spectra of native, intermediate and unfolded *Anabaena* apoflavodoxin were then determined by deconvolution of spectra recorded at different temperatures, using equation:

$$Y(\lambda, T) = Y_N(\lambda, T)X_N(T) + Y_I(\lambda, T)X_I(T) + Y_U(\lambda, T)X_U(T) \quad (6)$$

where the observed absorbance value at a given wavelength and temperature, $Y(\lambda, T)$, is a linear combination of the values of the different states, $Y_i(\lambda, T)$ and of their populations, $X_i(T)$ [84]. On the other hand, the populations are calculated at each temperature from the free energy values ΔG_1 and ΔG_2 previously obtained by global fitting to the sequential three-state model of unfolding curves recorded using absorbance, fluorescence and circular dichroism [53].

Results

Equilibrium trajectories at room temperature

Ten independent 200 ns long MD simulations suggest that the equilibrium structure of the protein in solution is close to that

Major clusters 10 replicas T= 300 K

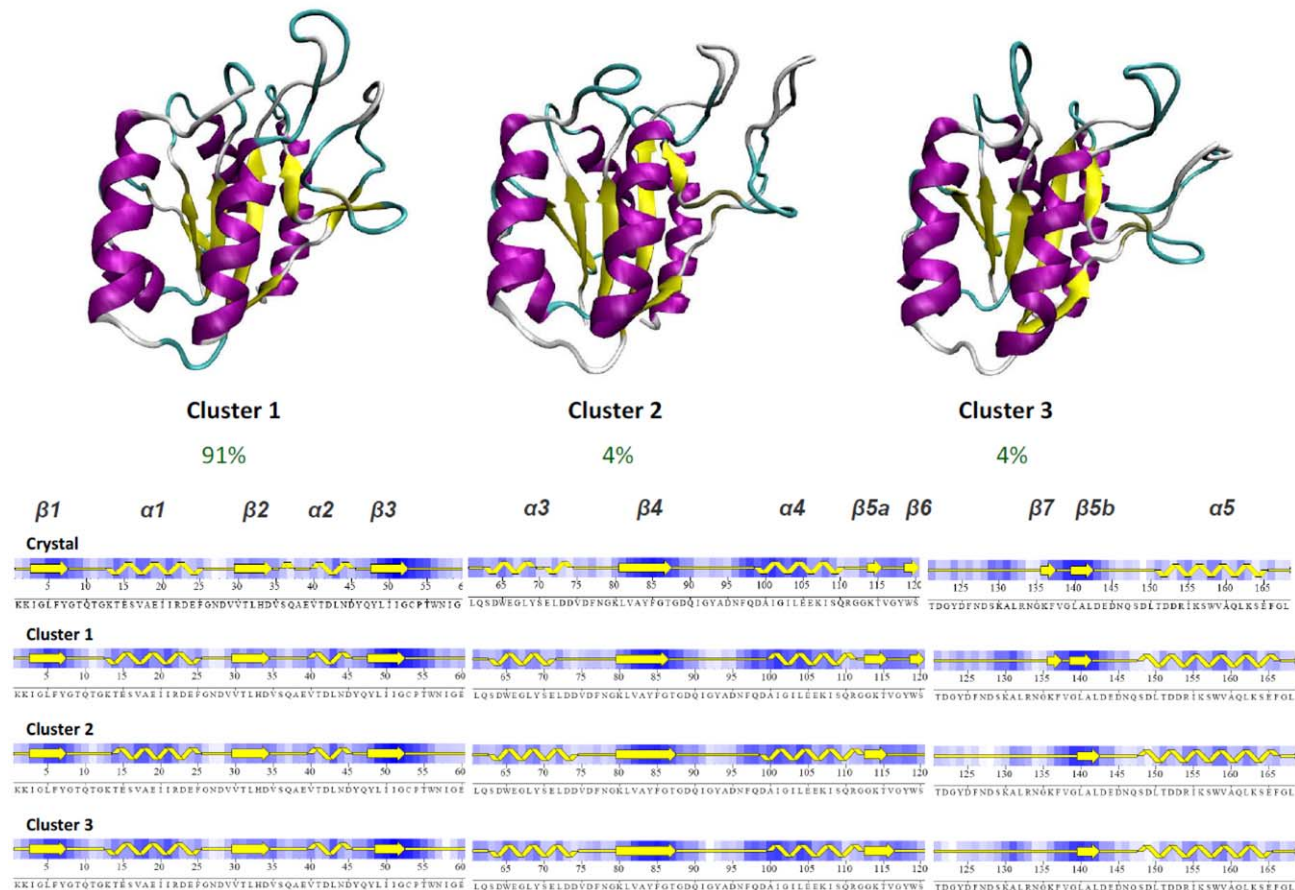


Figure 3. Major clusters obtained (cluster radii 2.5 Å; see Methods) from the meta-trajectory of apoflavodoxin at room temperature. The typical secondary structure content of each of cluster is shown, with the X-ray structure as reference. doi:10.1371/journal.pcbi.1002647.g003

found in the crystal, without any clear unfolding tendency (Figure 1). The RMSd of trajectory from the crystal structure are always below 3 Å for all replicas, and seems quite stable after the first 10–40 ns where protein relax from lattice contacts existing in the crystal structure (see Figure 1). The general shape of the protein and the structural core is fully maintained (see Tm-score plot in Figure 1) and the most significant deviation from crystal state is a small expansion of the protein as a result of the removal of lattice constraints, a behaviour very commonly found in massive MD simulations of the proteome [85–86], which is visible in a small (in average around 0.5 Å see Figure 1) increase in radii of gyration (which happens already in the post-equilibration phase) as well as in an increase around 13% in the solvent accessible surface without changes in the polar/apolar SAS (see Figure 1 and Suppl. Figure S1). This slight increase in the size of the protein when liberated from lattice constraints is reflected in a small increase in the Trp accessibility, a parameter that correlates with the UV spectra [87] of the protein (see Suppl. Figure S2). However, all changes in size and shape of the protein upon transferring from crystal to solution are small. Not surprisingly then, the scattering properties computed from the 2 μ s ensembles agree very well with the experimental SAXS curve, and also match the conformational preferences indicated by the X-ray structure (see Suppl. Figure S3).

Contacts between residues are massively preserved (see Suppl. Figure S1) and the few native contacts which are transiently lost are typically replaced by alternative contacts with neighbouring residues (see Suppl. Figure S1). Both α helices and β sheet elements remain fixed at crystal values, while there is a conversion of a portion of residues in β turn into coil conformations (see Suppl. Figure S1), which is localized in the loop regions. In fact, analysis of B-factors (Figure 2) obtained in the 2 μ s meta-trajectory reveals that the regions of larger flexibility are located around the loops of the protein. It is worth noting that most of these loops appear with large B-factors in the crystal structure. However, two loops which are flexible in the simulation (loop 90–100, contributing to binding the cofactor; and loop 120–135, characteristic of long-chain flavodoxins and involved in the binding of partner proteins) appear with low B-factors in the crystal. Analysis of different crystal structures of this protein in PDB (including 1FTG used here as starting conformation) reveals that in the crystal all these loops are directly or indirectly constrained by intermolecular packing contacts, which suggests that the largest mobility found in our simulations cannot be considered a simulation artefact (see Suppl. Figure S4).

Cartesian cluster analysis (Figure 3) reveals that around 91% of the time trajectories are sampling the same conformational

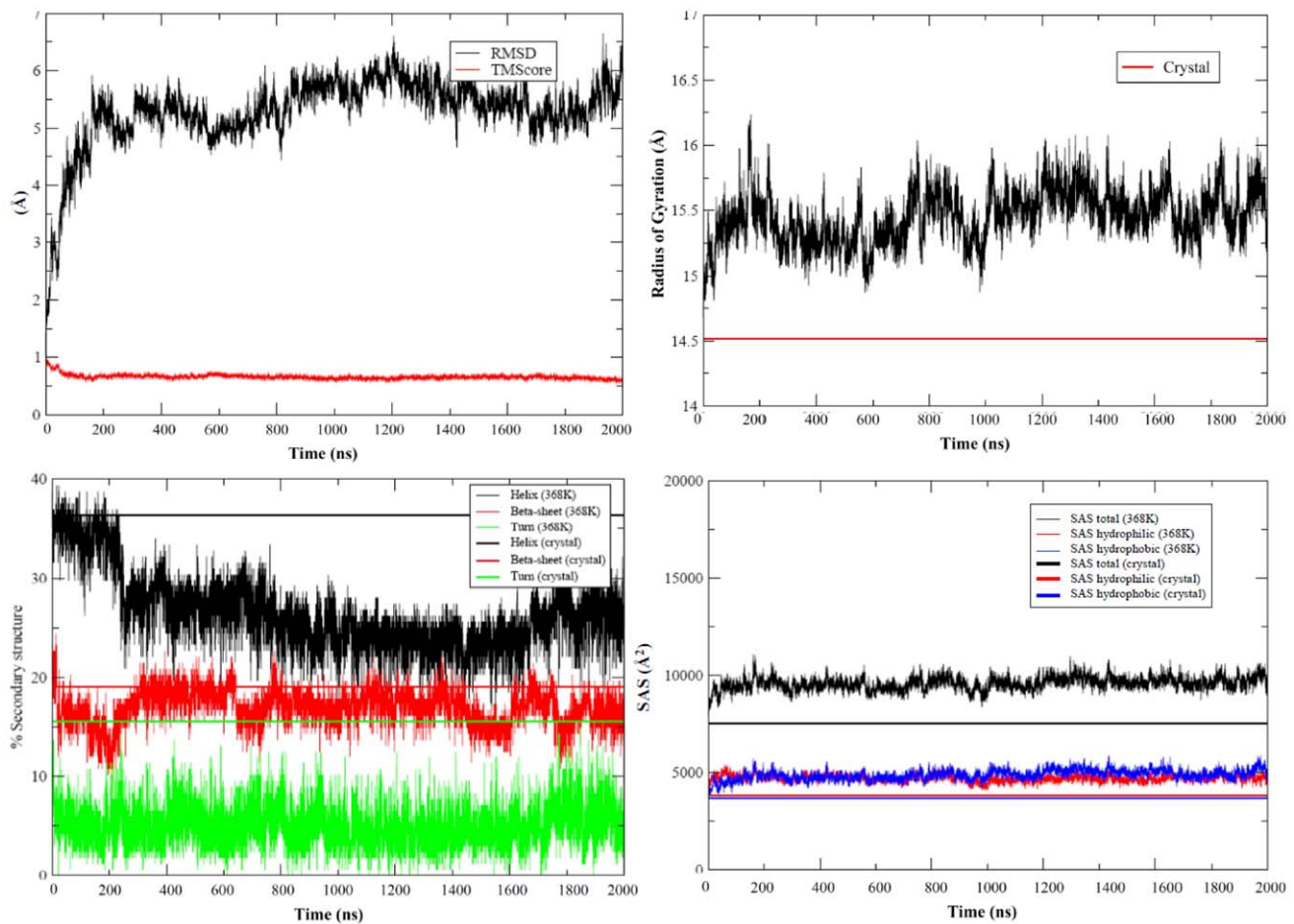


Figure 4. Evolution of different metrics along the 2 μ sec unfolding simulation of apoflavodoxin at 368 K. TOP/LEFT: All atoms root mean square deviation and TM-score from crystal structure (both in Å). TOP/RIGHT: Radii of gyration (in Å). BOTTOM/LEFT: content of secondary structure. BOTTOM/RIGHT Solvent accessible surface (total, hydrophobic and hydrophilic, all in Å²). Reference lines represent always the values found in the crystal structure (PDB code 1FTG).

doi:10.1371/journal.pcbi.1002647.g004

basin, which is very close to the crystal structure (RMSd to the crystal 1.5–2.5Å). The trajectory also populates two alternative basins (two clusters with population 4% each; RMSDs to crystal 2.0–2.5Å) that only differ in the conformation of the long loop characteristic of the long-chain flavodoxin family (including β_6 and β_7 ; positions 120–135) and, at a minor extent, in the 90–100 loop (that connecting β_4 and α_4). Conformational changes in the loops yield to a marginal loss of native contacts in the region (see Figures 2 and 3), without further changes in the global structure. In summary, extended MD simulations demonstrate that selected force-field and simulation conditions are able to represent the folded form of the protein, which seems to be quite rigid except for local movements in the aforementioned loops.

Extended unfolding at high temperature

It is never clear what is the effective temperature in a classical MD simulation, since it is force-field dependent [88–89]. It is then almost impossible to define a simulation temperature as to guarantee that a finite time simulation will populate the experimentally characterized thermal intermediate. Thus, as described in Methods, we decided to locate (by comparison with experimental data) the intermediate as a transient conformational ensemble populated during unfolding at high temperature (below water boiling point).

The increase in the temperature does not lead to complete protein unfolding in 2 μ s (see Figures 4–5), something that could be expected only in very fast-folder proteins, typically small proteins with simple kinetic folding mechanisms. The maintenance

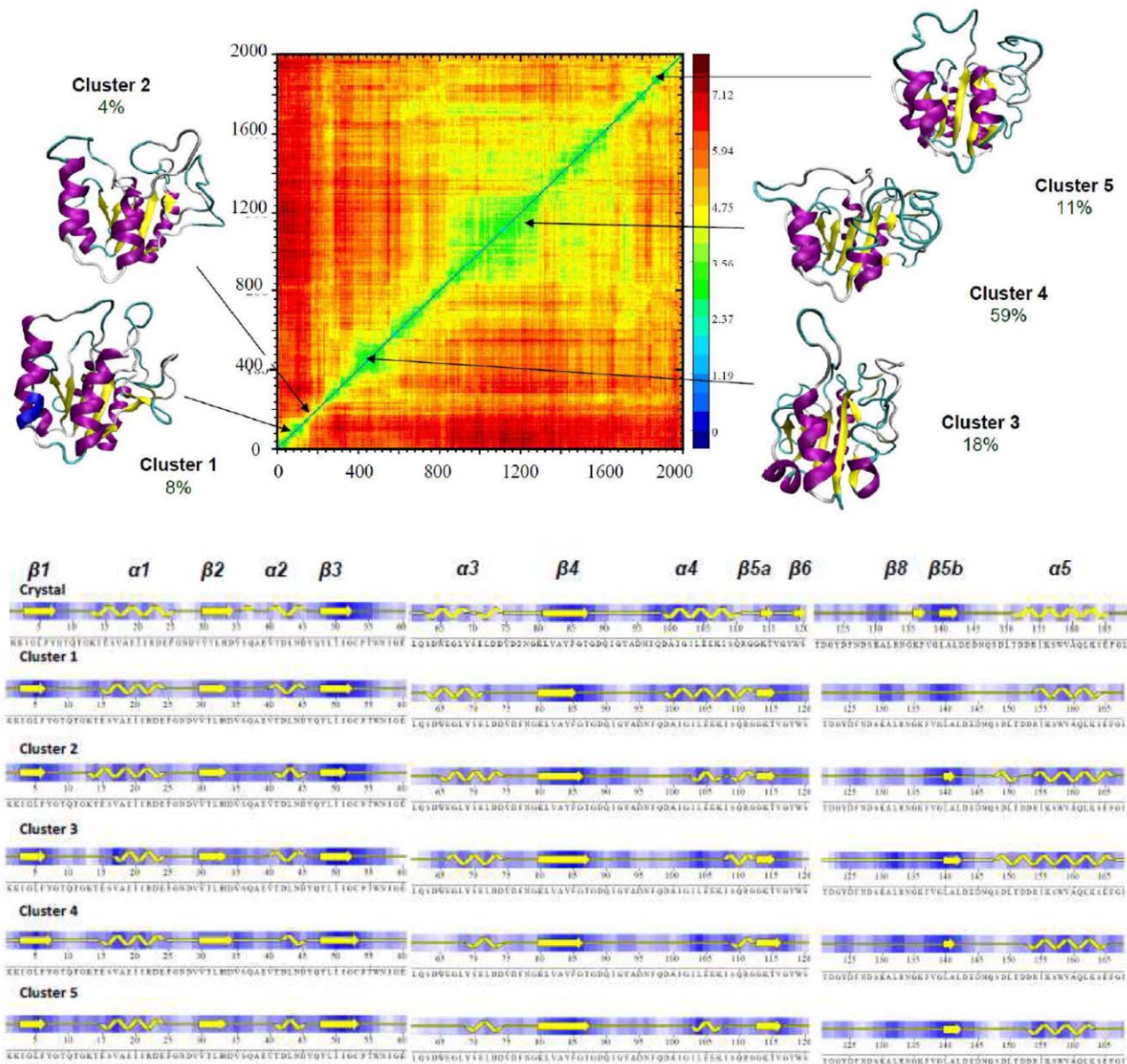


Figure 5. Cross-root mean square deviation plot for the 2 μ sec unfolding simulation of apoflavodoxin at 368 K (RMSd color code is in Å). The central structure of the different clusters (cluster radii 4.6 Å) sampled during the trajectories are projected into the cross RMSd plot. The typical secondary structure content of the structures in the different clusters is shown with a reference to X-Ray structure (PDB code 1FTG). doi:10.1371/journal.pcbi.1002647.g005

of TM-score and the hydrophobic solvent accessible surface demonstrate that the protein core is preserved even until 2 μ s of trajectory at high temperature. However, although the general fold is maintained, structural distortions from native structure are

significant at the end of the simulation (as noted in the large RMSd) and affect key elements of α and β secondary structure (Figures 4–5). Major distortions are first located at the loops, as expected from native simulations (see above), but are later

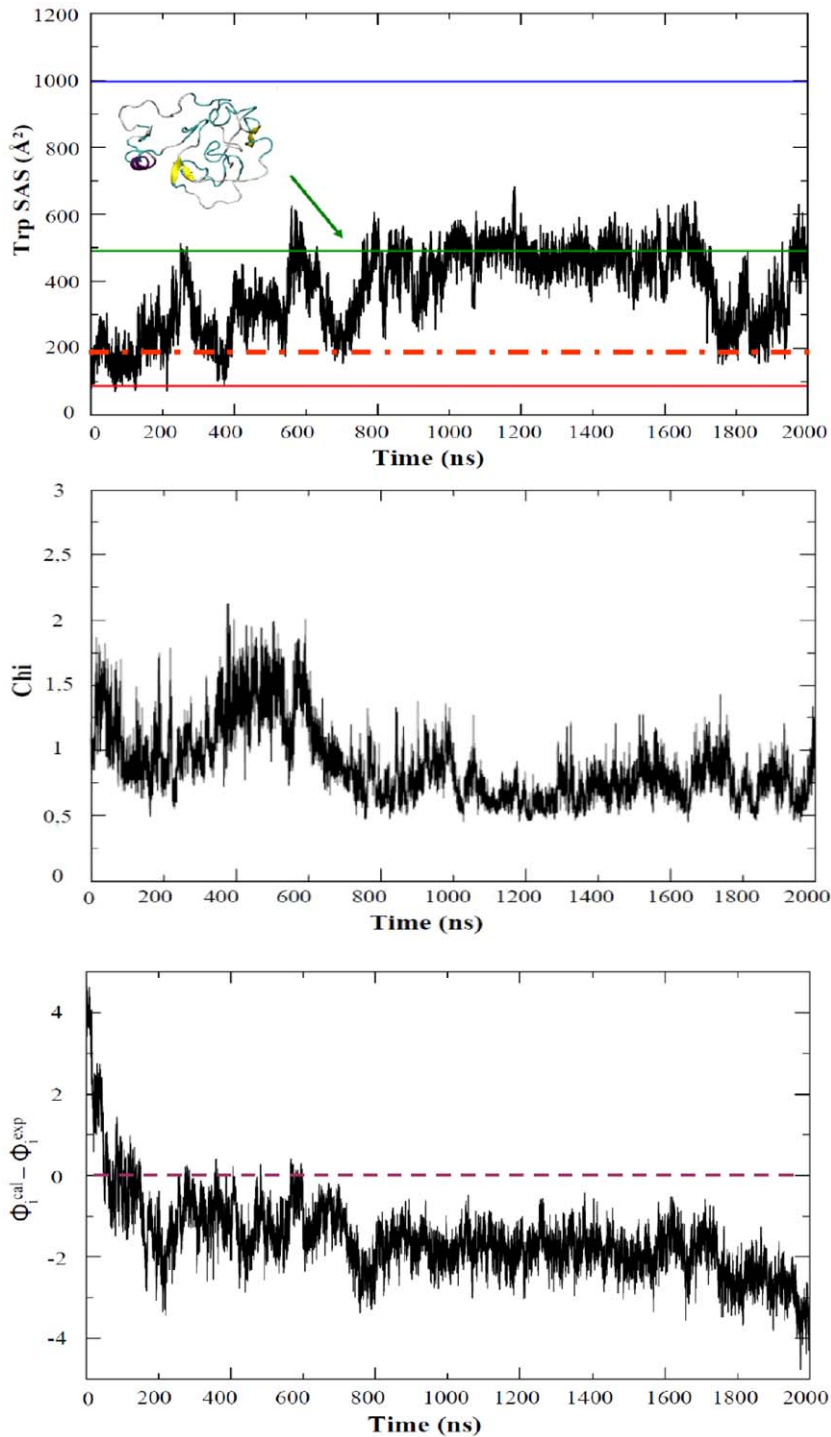


Figure 6. Evolution of different “experimental observables” that can be derived from the 2 μ sec unfolding simulation of apoflavodoxin at 368 K. TOP: Solvent accessible surface of the 4 Trp of the protein; reference lines correspond to crystal structure (solid red), low temperature MD meta-trajectory (dashed red), a highly unfolded structure obtained by extreme heating of the protein (in green) and four fully exposed Trp (blue). MIDDLE: Evolution of the merit function for fitting experimental and MD-simulated small angle scattering profile (see text). BOTTOM: total difference between experimental and simulated Φ values, where positive values indicate structure too close to the native state and negative values signals too unfolded conformations. doi:10.1371/journal.pcbi.1002647.g006

propagated to the neighbouring elements of secondary structure (see Figures 4–5). Thus, the large movements of loop 90–100 lead to distortions in neighbouring helix α_4 , which is shortened in 0.2–0.5 μs part of the trajectory and is almost completely lost at the end of it. Similarly, distortions in the long loop 120–135 produce early in the unfolding trajectory the disruption of small β -sheet elements β_6 , β_7 and β_{5b} and the shortening of terminal helix α_5 . Large movements of other smaller loops like 53–62 and 75–80 lead also to distortions of neighbouring secondary elements (for example helix α_3), but this happens late in the trajectory and is less dramatic than those noted above. Clearly, our long simulation has not statistical power to describe the intermediate, but suggests a general picture where the perturbation in the loops corrupts in a first step short elements of secondary structure, which has no impact in global structure, but later the α -helices segments are compromised which should eventually yield to the complete unfolding of the protein in longer time scales.

Cartesian cluster analysis reveals significant population (more than 100 ns) of 5 structural families along the 2 μs trajectories (Figure 5), which illustrates the increasing level of deformation gained along the simulation. It is tempting to assign the most populated family (cluster 4) as the putative intermediate, but as discussed above there is no guarantee that effective microscopic

simulation temperature matches the experimental macroscopic temperature at which the intermediate is detected. Accordingly, we cannot be sure which family represents better the intermediate ensemble and we do not know at which time frame intermediate is populated during our MD unfolding simulations. Clearly, comparison with experimental observables can help to locate the intermediate in our ensemble.

The UV spectra determined experimentally for the intermediate (see Methods) is very similar to that of the native state, without the blue shift in the spectra which is clear in the unfolded state (see Suppl. Figure S5). Thus, we can be quite sure that the exposure of Trp side chains has not changed much from native to intermediate state. Based on this criteria the intermediate is detected during the beginning of the simulation (around 0.2 μs ; Figure 6), while structures sampled at the second half of the trajectory yield too exposed Trp to justify experimental spectra. The SAXS spectra of the intermediate is well reproduced in the region 0.1–0.3 μs and later in the second half of the trajectory (as noted in χ values in Figure 6). Finally, the Φ -profile (see Methods) computed experimentally is well reproduced in the 0.1–0.2 μs region, while structures collected before are too “native-like” and those collected later have advanced too much in the unfolding pathway. In summary, comparison with experimental data strongly suggests

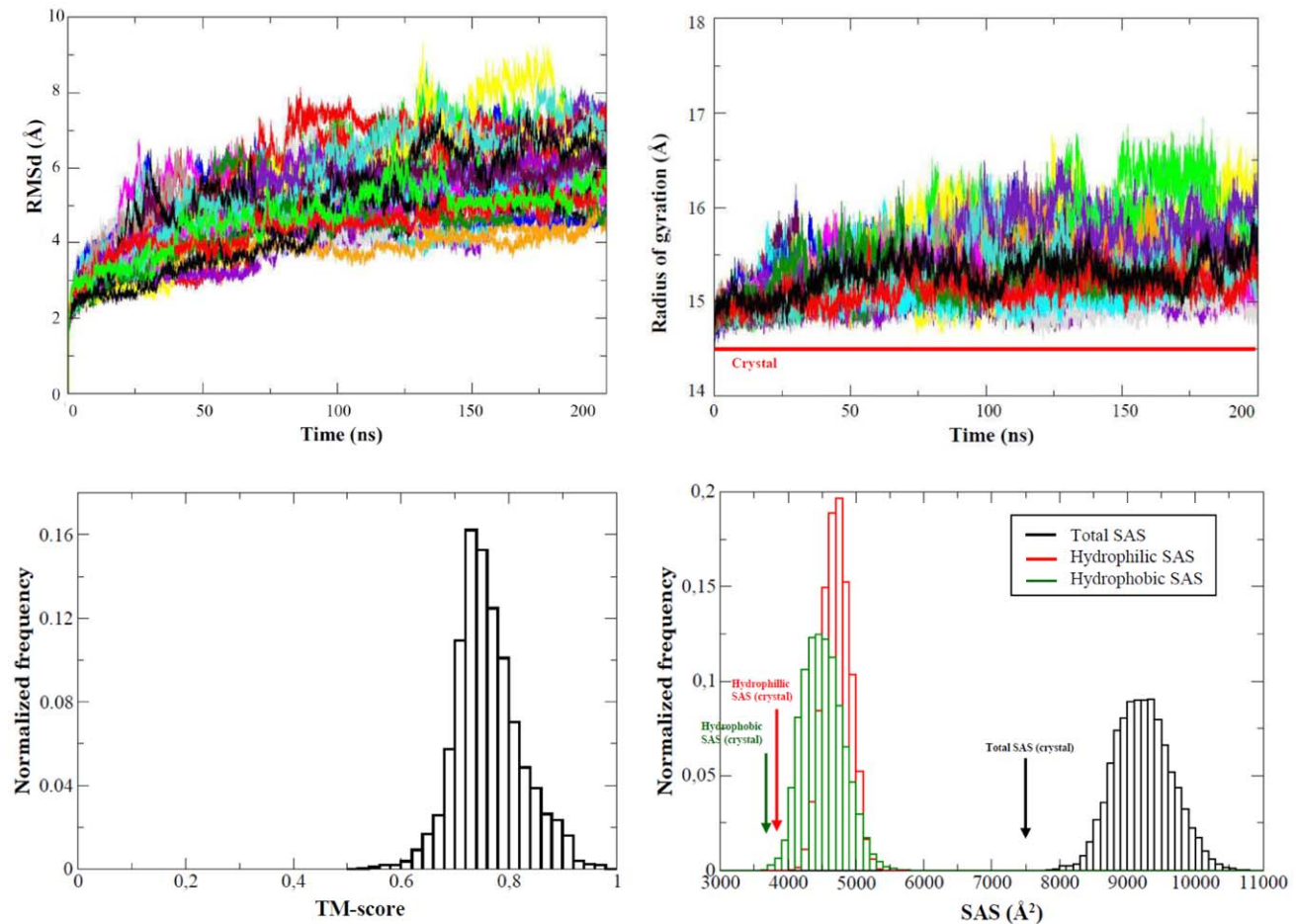


Figure 7. Evolution of different metrics along the 50 0.2 μs trajectories of apoflavodoxin at $T = 368$ K. TOP/LEFT: All atoms root mean square deviation from crystal structure (in \AA). TOP/RIGHT: Radii of gyration (in \AA). BOTTOM/LEFT TM-score (from the crystal) distribution (in \AA). BOTTOM/RIGHT: histogram of solvent accessible surface (total, hydrophobic and hydrophilic all in \AA^2). When reference lines appear they correspond to crystal structure.

doi:10.1371/journal.pcbi.1002647.g007

that the intermediate is going to be closer to clusters 1–3 than to the most populated cluster 4 (see Figure 6), and that it is reached quite fast (around 0.2 μ s) during our unfolding simulation.

Ensemble simulations at high temperature

Following the findings obtained from the analysis of the 2 μ s trajectory, which suggested that native \rightarrow intermediate transition happens early in the simulation, we performed 50 independent 0.2 μ s trajectories, which combined provide us a 10 μ s ensemble enriched in the intermediate state. All the different trajectories advance towards protein denaturation (see Figure 7), with a range of velocities that show a normal distribution with unfolding velocities ranging from 0.4 to 0.8 nm RMSd/0.2 μ s. The lack of unusually slow or fast unfolding pathways [90] suggests the existence of a unique mechanism for the transition from folded to intermediate state under the selected simulation conditions, which is characterized by first a focalization of structural deformations in loops (Figure 7) and later a transfer of such perturbation to the surrounding elements of secondary structure (see Figure 8), matching the general unfolding trends found in the 2 μ s trajectory.

Cartesian clustering of the 10 μ s meta-trajectory allowed us to detect six major “states (clusters)”, four of them with populations

above 5%. Not surprisingly, the most populated one (69% of meta-trajectory) is that describing a near-native conformation, which appears populated in the beginning of all the individual trajectories. As the unfolding progresses, partially unfolded conformations, characterized by distorted loop conformations and partial losses of neighbouring secondary structure become populated (Figure 8). Thus, in structures assigned to cluster 2 (10% meta-trajectory, populated in 55% trajectories) the large movements of the long loop (120–135) have led to the loss of short β strand elements β_6 and β_7 . Ensembles represented by clusters 3 and 4 (12% and 7% meta-trajectory, populated in 65% and 45% individual trajectories) are characterized by an advance in the distortion produced by loop oscillations, either to the helix α_4 (cluster 3) or the helix α_3 (cluster 4). Finally, the minor clusters 5 and 6 represent much more distorted conformations, where a significant amount of secondary structure is lost and the departure from native basin is quite evident (Figure 8). Clusters 5 and 6 account for less than 1% of the entire meta-trajectory and are sampled only in two of the individual trajectories (one for each), which suggest that they do not fit the experimental requirements of the intermediate.

It is very tempting to try to identify one of the above mentioned clusters with the thermal intermediate, but analysis of the

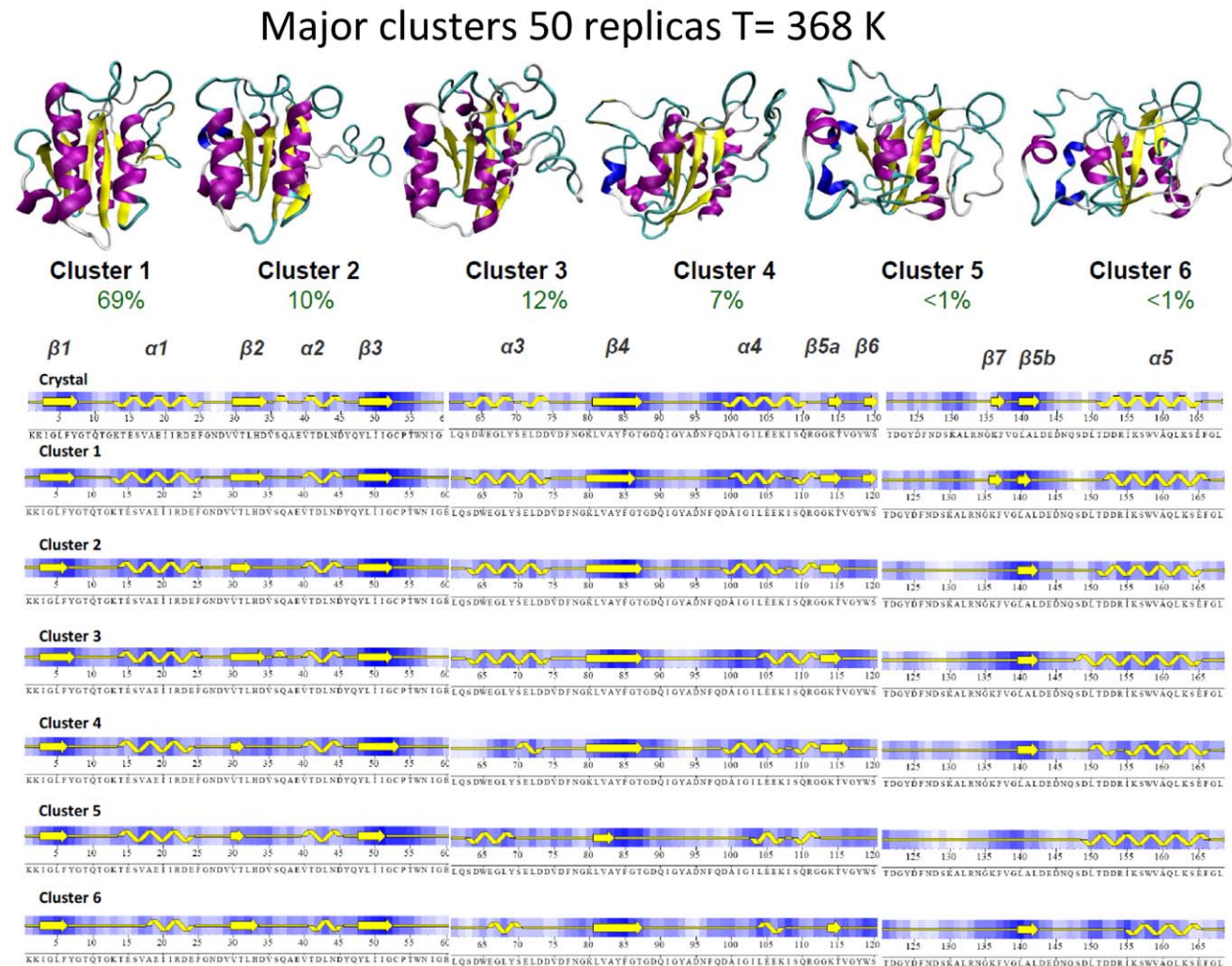


Figure 8. Major clusters (radius 5.5 Å) obtained in the meta-trajectory at 368 K. The percentage of population of these clusters and their typical secondary structure content are displayed. doi:10.1371/journal.pcbi.1002647.g008

individual trajectories show that in reality clusters 2–4 and part of structures assigned to cluster 1 interchange in a fast way and share many common characteristics, with a well conserved central core and largely distorted loop regions. The fast and large movement of such loops (and neighbouring secondary elements) generates a large dispersion in the structures when projected into the Cartesian space, which is reflected in the different assignment of

structures to different clusters, when they share many key structural characteristics. It is also worth to note that structures which are within the same cluster can yield very different values of some experimental observables (see Figure 9), while structures very distant in terms of RMSd, and accordingly assigned to the different clusters can be indistinguishable in terms of experimental observables (see Figure 9). In summary, it seems that the

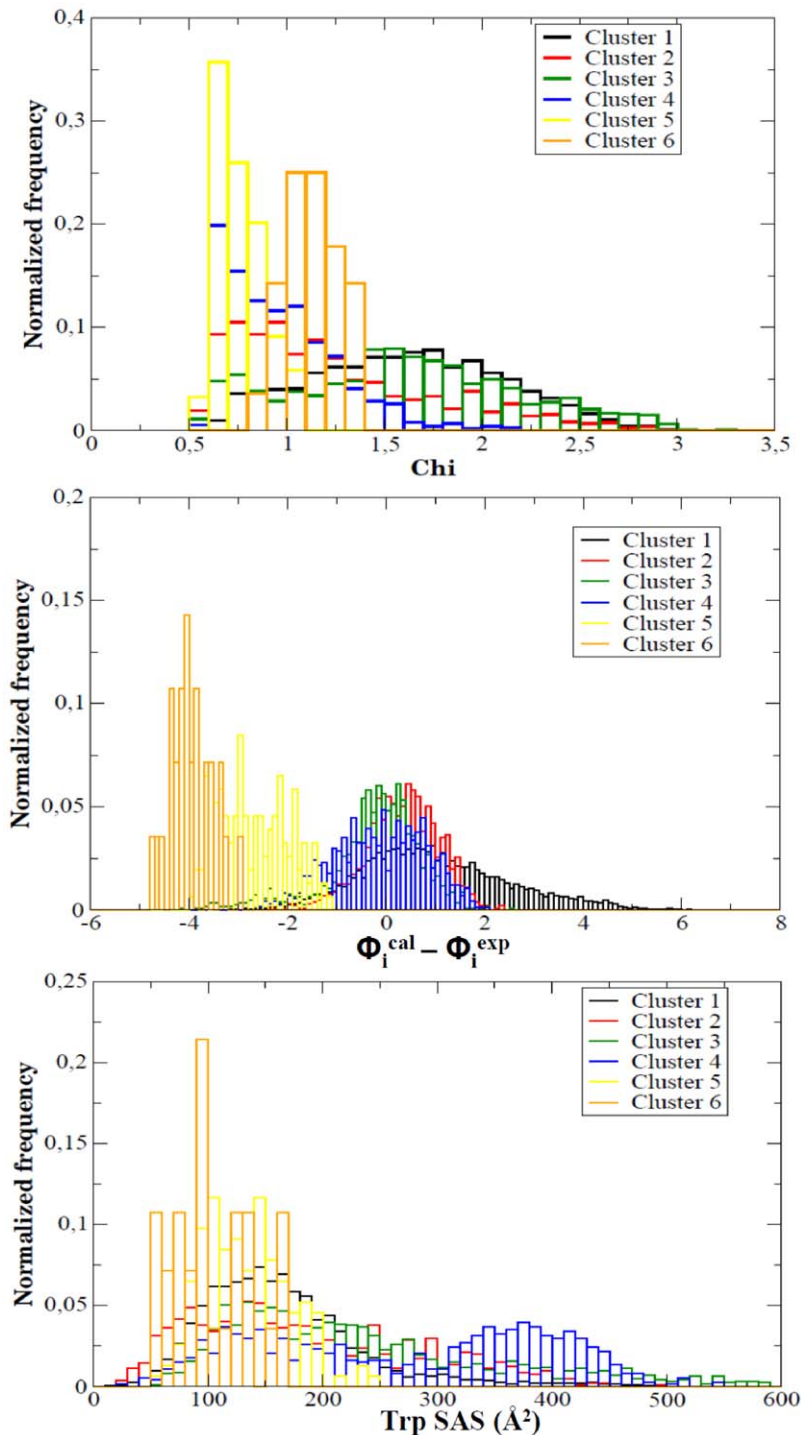


Figure 9. Histogram representation of the distributions of “experimental observables” obtained from the different clusters. From TOP to BOTTOM: fitting to SAXs curve, Φ -value error and Trp solvent accessible surface (see text and Figure 6 for details). doi:10.1371/journal.pcbi.1002647.g009

intermediate cannot be represented as a small ensemble defined as a narrow basin centered in a well-defined structure, but as a wide ensemble of conformations that cover a wide range of Cartesian space, but that share a common conformational core.

Discussion

The thermal intermediate

We interrogate our 10 μ s ensemble to determine how many of these structures fulfill all the experimental requirements of the ensemble known experimentally for the thermal intermediate. Considering a loosely criteria ($SAS_{T_{ip}}$ between 100 and 300 \AA^2 , fitting the SAXs curve with a χ below 1.5 and fitting the Φ -value profile with absolute accumulated error below 2) almost 30% of the ensemble is annotated as intermediate. If we assume that experimental measurements for the intermediate are very accurate and use a much more restrictive criterion ($SAS_{T_{ip}}$ between 100 and 300 \AA^2 , $\chi < 1.0$ and $\Phi\text{-error} < 1.0$) the intermediate ensemble is reduced to around 10% of the meta-trajectory. Such an ensemble is contributed by all individual trajectories and is proportionally enriched with structures assigned to clusters 2–4, with no contribution of clusters 5 and 6.

When analyzed, the intermediate sampling shows a quite interesting picture of the structure that is transiently populated during thermal unfolding of the protein (Figures 10–11). The

structure has enlarged with respect to the solution ensemble and hydrophobic solvent accessible surface has increased significantly, a fingerprint of a partially unfolded structure. A significant number of native contacts (defined as those present in the solution ensemble) are lost, especially those involving the protein loops, which have disappeared completely (Suppl. Figure S6). However, the structure maintains still many native inter-residue contacts, mostly located in the central core, where the amount of secondary structure has decreased, but is still quite significant (Figures 10–11). Clearly, analysis of the results demonstrates that the intermediate is not an alternative structure of the protein, but has to be represented as a wide ensemble (average RMSd between structures in the ensemble is around 0.6 nm; Figure 11). Two broad regions can be easily recognized in the protein: the central core, where the native fold is well preserved and the loops (including the long loop hosting a small β -sheet encompassing strands 6 and 7), which adopt a canonical random coil confirmation (Figure 11). It is very interesting to realize that the large flexibility movements governing the essential dynamics in the intermediate ensemble are already a maximization of the intrinsic deformation pattern of the native state of the protein (absolute similarity (γ) = 0.52; relative similarity (κ) = 0.76, see eqs. 1 and 2), as it was already suggested by B-factor distributions (see Figure 2). Altogether, the intermediate fits perfectly in the definition of a partially disordered protein with a solid-like core and a liquid-like

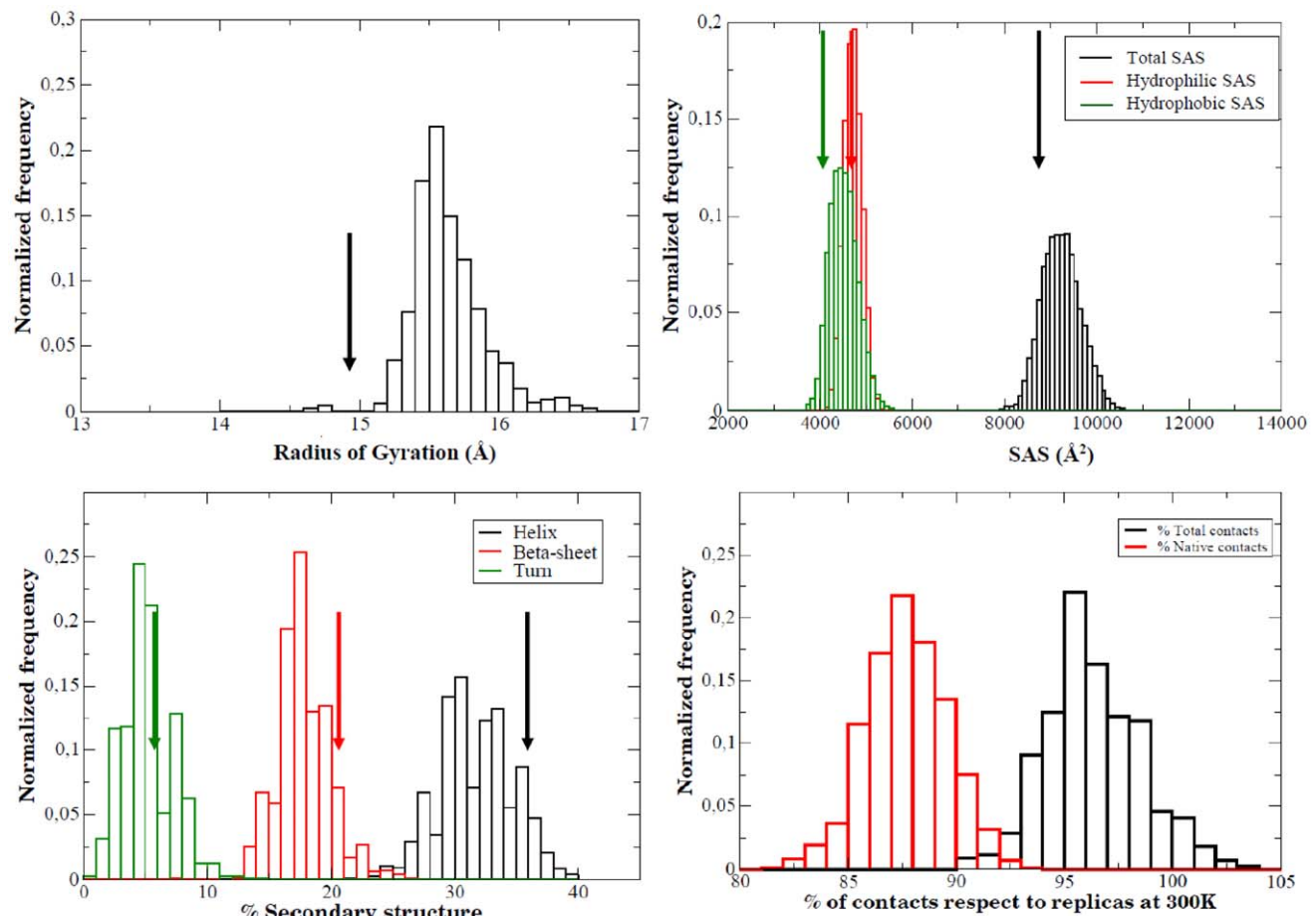


Figure 10. Histogram representation of different structural descriptors for the intermediate ensemble. TOP/LEFT: radii of gyration (in \AA); TOP/RIGHT: solvent accessible surface (total, hydrophobic and hydrophilic, all in \AA^2); BOTTOM/LEFT: secondary structure content; BOTTOM/RIGHT: native and total inter-residue contacts. All reference arrows correspond to crystal values.
doi:10.1371/journal.pcbi.1002647.g010

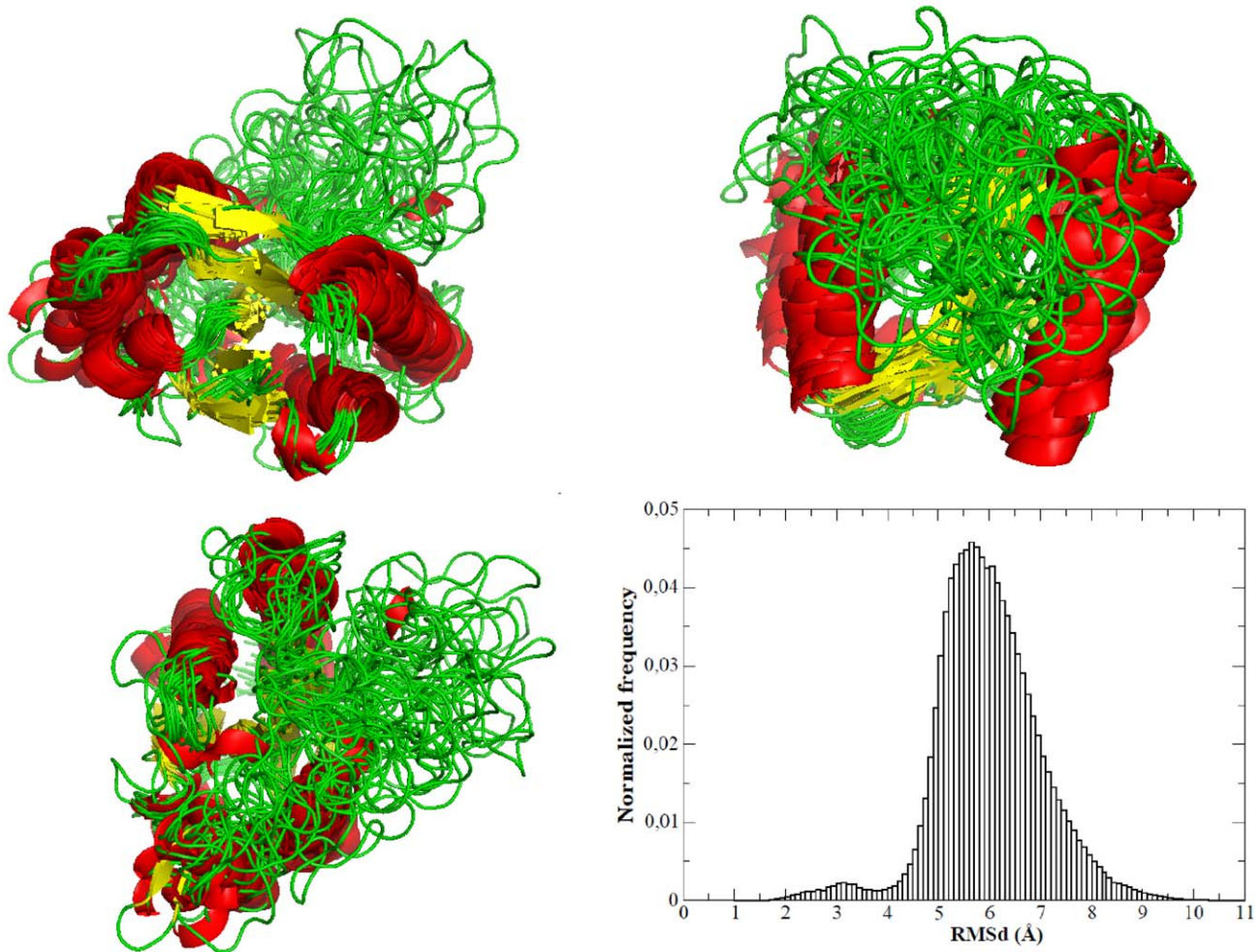


Figure 11. Three different views of the intermediate ensemble (10 randomly selected structures are superimposed) and the distribution of the root mean square distribution between different structures in the ensemble.
doi:10.1371/journal.pcbi.1002647.g011

external loop core. It is very encouraging that such a representation of the intermediate fits well with the picture derived from the analysis of the NMR spectra of a mutant, which is believed to adopt intermediate-like conformation under native conditions [91].

Our MD simulations suggest a quite complete picture of the initial stages of the thermal unfolding of apo flavodoxin, which might be common to other proteins having long loops stabilized by weak contacts. Thus, under native conditions the protein has an intrinsic tendency to become a partially disordered protein, but several loop-loop contact keep the potentially flexible part of the protein reasonably organized. When the temperature increases these loops gain kinetic energy and in a quite short period of time become random coils (see Figure 2). The anchoring points of the loops, with the exception of short β -sheet elements, are very stable and held together the core of the structure defining the experimentally detected intermediate. Additional thermal energy will be then concentrated in the anchoring points of the loops, particularly in the helices 3 and 4, which are the Achilles' heel of the apo flavodoxin core. The distortion of these helices opens the structure and should lead to the final disruption of the three dimensional structure of the protein in longer time scales. Under this general picture, the lack of intermediate when denaturing agent is urea [92] can be easily rationalized, since urea will attack

directly the core of the protein [39], eliminating the resistance points that stop the thermal unfolding pathways in a partially disordered conformation.

Under native conditions the thermal intermediate acts as an “in-path” stationary state, since the essential deformations of the intermediate implicitly code the intermediate→native transition, as noted in the high overlap ($Ov=0.63$; see eq. 3) between the intermediate essential deformation subspace and the intermediate→native transition vector. This finding strongly suggests that the intermediate should be considered as an “activated-high entropy” form of the native state, (see RMSd oscillations in Figure 11), with properties of partially disordered protein, which acts as an attractor of folding routes toward a state that in the absence of an excess of kinetic energy will converge in a down-hill manner to the native form. We can hypothesize that a non-negligible number of partially disordered proteins, which adopt a well-defined three dimensional structure only in the presence of partner, can be considered as generalized examples of three-state folder proteins, which in native conditions populates conformations containing well-structured cores and very mobile regions. The flexibility pattern of such intermediates should favour a down-hill transition to a well-defined three dimensional structure in the presence of interactions stabilizing the disordered region (in these case binding partners).

Supporting Information

Figure S1 Distribution of different structural descriptors in the meta-trajectory of apoflavodoxin obtained at room temperature. TOP/LEFT: native and total contacts (referred to crystal contacts); TOP/RIGHT: solvent accessible surface (total, hydrophobic and hydrophilic, all in \AA^2); BOTTOM/LEFT: native contacts of structures in the three clusters; BOTTOM/RIGHT: secondary structure content. All reference arrows correspond to crystal values. (TIF)

Figure S2 Distribution of the Trp solvent accessible (in \AA^2) surface obtained descriptors in the meta-trajectory of apoflavodoxin obtained at room temperature. Reference values correspond to crystal, a highly distorted protein and four fully exposed Trp (see Figure 6 for details). (TIF)

Figure S3 Distribution of the χ values (the fitting merit function) obtained when MD ensembles of apoflavodoxin at room temperature were used to fit SAXs experimental spectra. (TIF)

References

- Baum J, Dobson CM, Evans PA, Hanley C (1989) Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig alpha-lactalbumin. *Biochemistry* 28: 7–13.
- Hamada D, Hoshino M, Kataoka M, Fink AL, Goto Y (1993) Intermediate conformational states of apocytochrome c. *Biochemistry* 32: 10351–10358.
- Uversky VN, Pitsyn OB (1994) “Partly folded” state, a new equilibrium state of protein molecules: four-state guanidinium chloride-induced unfolding of beta-lactamase at low temperature. *Biochemistry* 33: 2782–2791.
- Zhu H, Celinski SA, Scholtz JM, Hu JC (2001) An engineered leucine zipper a position mutant with an unusual three-state unfolding pathway. *Protein Sci* 10: 24–33.
- Pedroso I, Irun MP, Machicado C, Sancho J (2002) Four-state equilibrium unfolding of an scFv antibody fragment. *Biochemistry* 41: 9873–9884.
- Hobart SA, Meinhold DW, Osuna R, Colon W (2002) From two-state to three-state: the effect of the P61A mutation on the dynamics and stability of the factor for inversion stimulation results in an altered equilibrium denaturation mechanism. *Biochemistry* 41: 13744–13754.
- Rami BR, Krishnamoorthy G, Udgaonkar JB (2003) Dynamics of the core tryptophan during the formation of a productive molten globule intermediate of barstar. *Biochemistry* 42: 7986–8000.
- Campos LA, Sancho J (2003) The active site of pepsin is formed in the intermediate conformation dominant at mildly acidic pH. *FEBS Lett* 538: 89–95.
- Martins SM, Chapeaurouge A, Ferreira ST (2003) Folding intermediates of the prion protein stabilized by hydrostatic pressure and low temperature. *J Biol Chem* 278: 50449–50455.
- Jackson SE (1998) How do small single-domain proteins fold? *Folding Des* 3: R81–R91.
- Privalov PL (1996) Intermediate states in protein folding. *J Mol Biol* 258: 707–725.
- States DJ, Creighton TE, Dobson CM, Karplus M (1987) Conformations of intermediates in the folding of the pancreatic trypsin inhibitor. *J Mol Biol* 195: 731–739.
- Thomas PJ, Qu BH, Pedersen PL (1995) Defective protein folding as a basis of human disease. *Trends Biochem Sci* 20: 456–459.
- Prusiner SB (1998) Prions. *Proc Natl Acad Sci U S A* 95: 13363–13383.
- Ellis RJ, van der Vies SM (1991) Molecular chaperones. *Annu Rev Biochem* 60: 321–347.
- Plesofsky-Vig N, Brambl R (1995) Disruption of the gene for hsp30, an alpha-crystallin-related heat shock protein of *Neurospora crassa*, causes defects in thermotolerance. *Proc Natl Acad Sci U S A* 92: 5032–5036.
- Parsell DA, Kowal AS, Singer MA, Lindquist S (2002) Protein disaggregation mediated by heat shock protein Hsp104. *Nature* 372: 475–478.
- van der Goot FG, Gonzalez-Manas JM, Lakey JH, Pattus F (1991) A “molten-globule” membrane insertion intermediate of the pore-forming domain of colicin A. *Nature* 354: 408–410.
- Bychkova VE, Berni R, Rossi GL, Kutysenko P, Pitsyn OB (1992) Retinol-binding protein is in the molten globule state at low pH. *Biochemistry* 31: 7566–7571.
- Uversky VN, Narizhneva NV, Ivanova TV, Kirkitadze MD, Tomashevski AY (1997) Ligand-free form of human-fetoprotein: evidence for the molten globule state. *FEBS Lett* 410: 280–284.

Figure S4 Detail of the loop packing in the crystal lattice of two crystal structures of apoflavodoxin (1FTG, that used as reference here) and 1DX9 (which displays a different crystal symmetry). (TIF)

Figure S5 Experimental ultraviolet spectra of native, unfolded and intermediate states of apoflavodoxin (see text for details). (TIF)

Figure S6 Contact plots corresponding to the native and intermediate ensembles. Regions where loss of contacts are especially remarkable are marked. (TIF)

Acknowledgments

We thank Tim Meyer for his help with some of the analysis.

Author Contributions

Conceived and designed the experiments: JS MO. Performed the experiments: RGF. Analyzed the data: RGF MO. Contributed reagents/materials/analysis tools: PB SAT. Wrote the paper: MO.

- Cai S, Singh BR (2001) Role of the disulfide cleavage induced molten globule state of type A botulinum neurotoxin in its endopeptidase activity. *Biochemistry* 40: 15327–15333.
- Fersht A (1999) *Structure and Mechanism in Protein Science*. New York: W. H. Freeman.
- Alexandrescu AT, Evans PA, Pitkeathly M, Baum J, Dobson CM (1993) Structure and dynamics of the acid-denatured molten globule state of alpha-lactalbumin: a two-dimensional NMR study. *Biochemistry* 32: 1707–1718.
- Schulman BA, Kim PS, Dobson CM, Redfield C (1997) A residue-specific NMR view of the noncooperative unfolding of a molten globule. *Nat Struct Biol* 4: 630–634.
- Zhang O, Forman-Kay JD, Shortle D, Kay LE (1997) Triple-resonance NOESY-based experiments with improved spectral resolution: applications to structural characterization of unfolded, partially folded and folded proteins. *J Biomol NMR* 9: 181–200.
- Chamberlain AK, Marqusee S (1998) Molten globule unfolding monitored by hydrogen exchange in urea. *Biochemistry* 37: 1736–1742.
- Eliezer D, Yao J, Dyson JH, Wright PE (1998) Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat Struct Biol* 5: 148–155.
- Ramboarina S, Redfield C (2003) Structural characterisation of the human alpha-lactalbumin molten globule at high temperature. *J Mol Biol* 330: 1177–1188.
- Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 224: 771–782.
- Matouschek A, Kellis JT Jr, Serrano L, Fersht AR (1989) Mapping the transition state and pathway of protein folding by protein engineering. *Nature* 340: 122–126.
- Matouschek A, Kellis JT Jr, Serrano L, Bycroft M, Fersht AR (1990) Transient folding intermediates characterized by protein engineering. *Nature* 346: 440–445.
- Daggett V, Levitt M (1992) A model of the molten globule state from molecular dynamics simulations. *Proc Natl Acad Sci U S A* 89: 5142–5146.
- Mark AE, van Gunsteren WF (1992) Simulation of the thermal denaturation of hen egg white lysozyme: trapping the molten globule state. *Biochemistry* 31: 7745–7748.
- Daggett V, Levitt M (1993) Protein unfolding pathways explored through molecular dynamics simulations. *J Mol Biol* 232: 600–619.
- Tirado-Rives J, Jorgensen WL (1993) Molecular dynamics simulations of the unfolding of apomyoglobin in water. *Biochemistry* 32: 4175–4184.
- Callisch A, Karplus M (1994) Molecular dynamics simulation of protein denaturation: solvation of the hydrophobic cores and secondary structure of barnase. *Proc Natl Acad Sci U S A* 91: 1746–1750.
- Alonso D, Daggett V (1995) Molecular dynamics simulations of protein unfolding and limited refolding: Characterization of partially unfolded states of ubiquitin in 60% methanol and in water. *J Mol Biol* 247: 501–520.
- Lazaridis T, Karplus M (1997) “New view” of protein folding reconciled with the old through multiple unfolding simulations. *Science* 278: 1928–1931.
- Tirado-Rives J, Orozco M, Jorgensen WL (1997) Molecular dynamics simulations of the unfolding of barnase in water and 8 M aqueous urea. *Biochemistry* 36: 7313–7329.

40. Tsai J, Levitt M, Baker D (1999) Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. *J Mol Biol* 291: 215–225.
41. Smith IJ, Dobson CM, van Gunsteren WF (1999) Side-chain conformational disorder in a molten globule: molecular dynamics simulations of the A-state of human alpha-lactalbumin. *J Mol Biol* 286: 1567–1580.
42. Mayor U, Johnson CM, Daggett V, Fersht AR (2000) Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc Natl Acad Sci U S A* 97: 13518–13522.
43. Prompers JJ, Scheurer C, Brüschweiler R (2001) Characterization of NMR relaxation-active motions of a partially folded A-state analogue of ubiquitin. *J Mol Biol* 305: 1085–1097.
44. Daggett V (2002) Molecular dynamics simulations of the protein unfolding/folding reaction. *Acc Chem Res* 35: 422–429.
45. Sancho J (2006) Flavodoxins: sequence, folding, binding, function and beyond. *Cell Mol Life Sci* 63: 855–864.
46. Genzor CG, Beldarrain A, Gomez-Moreno C, Lopez-Lacomba JL, Cortijo M, et al. (1996) Conformational stability of apoflavodoxin. *Protein Sci* 5: 1376–1388.
47. Irun MP, Garcia-Mira MM, Sanchez-Ruiz JM, Sancho J (2001) Native hydrogen bonds in a molten globule: the apoflavodoxin thermal intermediate. *J Mol Biol* 306: 877–888.
48. Apiyo D, Guidry J, Wittung-Stafshede P (2000) No cofactor effect on equilibrium unfolding of *Desulfovibrio desulfuricans* flavodoxin. *Biochim Biophys Acta* 1479: 214–224.
49. Muralidhara BK, Wittung-Stafshede P (2005) FMN binding and unfolding of *Desulfovibrio desulfuricans* flavodoxin: 'hidden' intermediates at low denaturant concentrations. *Biochim Biophys Acta* 1747: 239–250.
50. Nuallain BO, Mayhew SG (2002) A comparison of the urea-induced unfolding of apoflavodoxin and flavodoxin from *Desulfovibrio vulgaris*. *Eur J Biochem* 269: 212–223.
51. Genzor CG, Perales-Alcon A, Sancho J, Romero A (1996) Closure of a tyrosine/tryptophan aromatic gate leads to a compact fold in apo flavodoxin. *Nat Struct Biol* 3: 329–332.
52. Langdon GM, Jimenez MA, Genzor CG, Maldonado S, Sancho J, et al. (2001) Anabaena apoflavodoxin hydrogen exchange: on the stable exchange core of the alpha/beta(21345) flavodoxinlike family. *Proteins* 43:476–488.
53. Campos LA, Bueno M, Lopez-Llano J, Jimenez MA, Sancho J (2004) Structure of stable protein folding intermediates by equilibrium phi-analysis: the apoflavodoxin thermal intermediate. *J Mol Biol* 344: 239–255.
54. Ayuso-Tejedor S, Garcia-Fandiño R, Orozco M, Sancho J, Bernadó P (2011) Structural analysis of an equilibrium holding intermediate in the apoflavodoxin native ensemble by Small-angle X-ray scattering. *J Mol Biol* 406: 604–619.
55. Gelpi JL, Kalko SG, Barril X, Cirera J, de la Cruz X, et al. (2001) Classical molecular interaction potentials: Improved setup procedure in molecular dynamics simulations of proteins. *Proteins* 45: 428–437.
56. Jorgensen WL, Chandrasekhar J, Madura JD, Impey R, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79: 926–935.
57. Shields G, Laughton CA, Orozco M (1997) Molecular dynamics simulations of the d (T•A•T) triple helix. *J Am Chem Soc* 119: 7463–7469.
58. Darden T, York D, Pedersen L (1993) Particle mesh Ewald - an N.Log(N) method for Ewald sums in large systems. *J Chem Phys* 98: 10089–10092.
59. Tuckerman ME, Berne BJ (1991) Molecular dynamics algorithm for multiple time scales: Systems with disparate masses. *J Chem Phys* 94: 1465–1469.
60. Andersen HC (1983) Rattle - a velocity version of the shake algorithm for molecular-dynamics calculations. *J Comput Phys* 52: 24–34.
61. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, et al. (1999) NAMD2: Greater scalability for parallel molecular dynamics. *J Comput Phys* 151: 283–312.
62. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26: 1781–1802.
63. Foloppe N, MacKerell AD Jr (2000) All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J Comp Chem* 21: 86–104.
64. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, et al. (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B* 102: 3586–3616.
65. Case DA, Darden TA, Cheatham III TE, Simmerling CL, Wang J, et al. (2006) AMBER 9, University of California, San Francisco.
66. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham III TE, et al. (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp Phys Commun* 91: 1–41.
67. Case DA, Cheatham T, Darden T, Gohlke H, Luo R, et al. (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26: 1668–1688.
68. Feig M, Karanicolas J, Brooks CL 3rd. (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J Mol Graph Model* 22: 377–395.
69. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8: 477–486.
70. Hubbard SJ, Thornton JM (1993) 'NACCESS', Computer Program, Department of Biochemistry and Molecular Biology, University College London.
71. Amadei A, Linssen ABM, Berendsen HJC (1993) Essential dynamics of proteins. *Proteins* 17: 412–415.
72. Hess B (2000) Similarities between principal components of protein dynamics and random diffusion. *Phys Rev E Stat Nonlin Soft Matter Phys* 62: 8438–8448.
73. Pérez A, Blas JR, Rueda M, López-Bes JM, de la Cruz X, et al. (2005) Exploring the essential dynamics of B-DNA. *J Chem Theory Comput* 1: 790–800.
74. Rueda M, Chacón P, Orozco M (2007) Thorough Validation of Protein Normal Mode Analysis- A Comparative Study with Essential Dynamics. *Structure* 15: 565–575.
75. Bueno M, Campos LA, Estrada J, Sancho J (2006). Energetics of aliphatic deletions in protein cores. *Protein Sci* 15: 1858–1872.
76. Paci E, Karplus M (1999) Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *J Mol Biol* 288: 441–459.
77. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* 36: 1277–1282.
78. Guinier A (1939) La Diffraction des Rayons X aux Très Faibles Angles: Applications à l'Etude des Phénomènes Ultra-microscopiques. *Annales de physique (Paris)* 12: 161–237.
79. Tauler R, Smilde A, Kowalski B (1995) Selectivity, Local Rank, 3-Way Data-Analysis and Ambiguity in Multivariate Curve Resolution. *J Chemom* 9: 31–58.
80. Tauler R (1995) Multivariate curve resolution applied to second order data. *Chemom Intell Lab Syst* 30: 133–146.
81. de Juan A, Tauler R (2006) Multivariate curve resolution (MCR) from 2000: Progress in concepts and applications. *Crit Rev Anal Chem* 36: 163–176.
82. Blobel J, Bernadó P, Svergun DI, Tauler R, Pons M (2009) Low-Resolution Structures of Transient Protein-Protein Complexes Using Small-Angle X-ray Scattering. *J Am Chem Soc* 131: 4378–4386.
83. Svergun D, Barberato C, Koch MHJ (1995) CRYSOLE - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28: 768–773.
84. Cremades N, Velázquez-Campoy A, Freire E, Sancho J (2008) The flavodoxin from *Helicobacter pylori*: Structural determinants of thermostability and FMN cofactor binding. *Biochemistry* 47: 627–639.
85. Rueda M, Ferrer C, Meyer T, Pérez A, Camps J, et al. (2007) A consensus view to Protein Dynamics. *Proc Natl Acad Sci U S A* 104: 796–801.
86. Meyer T, D'Abramo M, Hospital A, Rueda M, Ferrer-Costa C, et al. (2010) MoDEL (Molecular Dynamics Extended Library). A database of atomistic molecular dynamics simulations. *Structure* 18: 1399–1409.
87. Schmid FX (1989) Spectral methods of characterizing protein conformation and conformational changes. In: Creighton TE, editor. *Protein Structure: a Practical Approach*. Oxford: IRL Press. pp 251–285.
88. Best RB, Mittal J (2011) Free-energy landscape of the GB1 hairpin in all-atom explicit solvent simulations with different force fields: Similarities and differences. *Proteins* 79: 1318–1328.
89. Naganathan AN, Orozco M (2011) The Native ensemble and folding of a protein molten-globule: functional consequence of downhill folding. *J Am Chem Soc* 133: 12154–61.
90. Portella G, Orozco M (2010) Multiple routes characterize the folding of a small DNA hairpin. *Angew Chem Int Ed Eng* 49: 7673–7676.
91. Ayuso-Tejedor S, Espinosa Angarica V, Bueno M, Campos LA, Abián O, et al. (2010) Design and Structure of an Equilibrium Protein Folding Intermediate: A Hint into Dynamical Regions of Proteins. *J Mol Biol* 400: 922–934.
92. Bueno M, Ayuso-Tejedor S, Sancho J (2006) Do Proteins with Similar Folds Have Similar Transition State Structures? A Diffuse Transition State of the 169 Residue Apoflavodoxin. *J Mol Biol* 359: 813–824.