

Article

# High Energy Channeling and Malleable Transition States: Molecular Dynamics Simulations and Free Energy Landscapes for the Thermal Unfolding of Protein U1A and 13 Mutants

Na Le Dang , Anne M. Baranger <sup>†</sup> and David L. Beveridge <sup>\* </sup>

Department of Chemistry and Molecular Biophysics Program, Wesleyan University, Middletown, CT 06459, USA; ldang@wesleyan.edu (N.L.D.); abaranger@berkeley.edu (A.M.B.)

\* Correspondence: dbeveridge@wesleyan.edu

† Present address: Department of Chemistry, University of California, Berkeley, CA 94720, USA.

**Abstract:** The spliceosome protein U1A is a prototype case of the RNA recognition motif (RRM) ubiquitous in biological systems. The in vitro kinetics of the chemical denaturation of U1A indicate that the unfolding of U1A is a two-state process but takes place via high energy channeling and a malleable transition state, an interesting variation of typical two-state behavior. Molecular dynamics (MD) simulations have been applied extensively to the study of two-state unfolding and folding of proteins and provide an opportunity to obtain a theoretical account of the experimental results and a molecular model for the transition state ensemble. We describe herein all-atom MD studies including explicit solvent of up to 100 ns on the thermal unfolding (UF) of U1A and 13 mutants. Multiple MD UF trajectories are carried out to ensure accuracy and reproducibility. A vector representation of the MD unfolding process in RMSD space is obtained and used to calculate a free energy landscape for U1A unfolding. A corresponding MD simulation and free energy landscape for the protein CI2, well known to follow a simple two state folding/unfolding model, is provided as a control. The results indicate that the unfolding pathway on the MD calculated free energy landscape of U1A shows a markedly extended transition state compared with that of CI2. The MD results support the interpretation of the observed chevron plots for U1A in terms of a high energy, channel-like transition state. Analysis of the MDUF structures shows that the transition state ensemble involves microstates with most of the RRM secondary structure intact but expanded by ~14% with respect to the radius of gyration. Comparison with results on a prototype system indicates that the transition state involves an ensemble of molten globule structures and extends over the region of ~1–35 ns in the trajectories. Additional MDUF simulations were carried out for 13 U1A mutants, and the calculated  $\phi$ -values show close accord with observed results and serve to validate our methodology.

**Keywords:** Molecular dynamics; protein unfolding; transition state; U1A



**Citation:** Dang, N.L.; Baranger, A.M.; Beveridge, D.L. High Energy Channeling and Malleable Transition States: Molecular Dynamics Simulations and Free Energy Landscapes for the Thermal Unfolding of Protein U1A and 13 Mutants. *Biomolecules* **2022**, *12*, 940. <https://doi.org/10.3390/biom12070940>

Academic Editors: Steven R. Van Doren, C. Martin Lawrence and Brigita Urbanc

Received: 15 April 2022

Accepted: 27 June 2022

Published: 4 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Protein folding and unfolding at the molecular level has been elucidated to a considerable extent by studies on simple two-state processes, which typically have relatively narrow and sharply peaked activation barriers between native (N) and denatured (D) states on the free energy landscape [1]. An interesting variation on the theme of two-state protein folding has been observed for the folding/unfolding of the spliceosome protein U1A, which shows that the folding/unfolding of U1A is essentially two-state but takes place by “high energy channeling” via a broad and flat activation barrier and a “malleable” transition state [2]. Further understanding of these phenomena at the molecular level is an important next step.

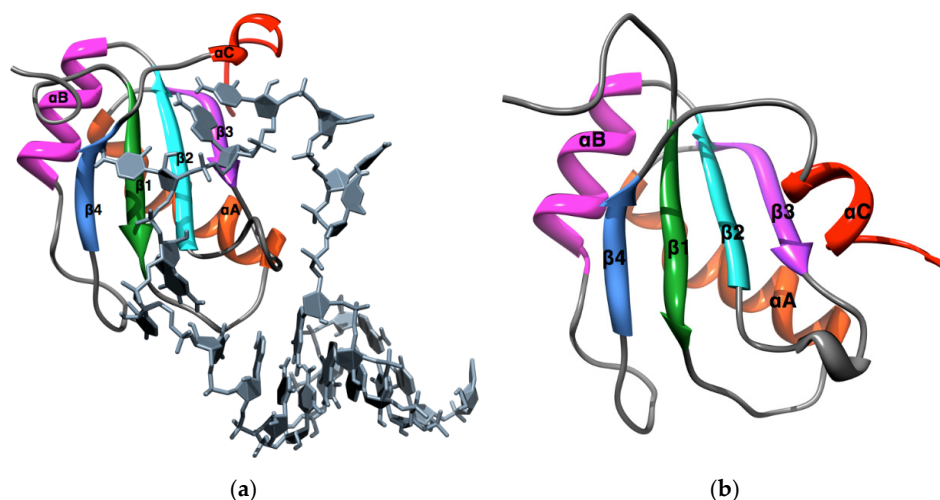
The synergy between molecular simulations and experiments has been evolving rapidly with the improvement of force fields and access to high-performance computing. Refs. [2–6] Molecular dynamics (MD) simulations have recently proved to be quite

successful in obtaining computational models for two-state protein folding [7,8] and unfolding [9,10]. Daggett and coworkers have published extensively on thermal unfolding (UF) of small proteins using MD [11], henceforth abbreviated MDUF, and elucidated transition state processes from a representation of MDUF trajectories in root-mean-square-deviation (RMSD) space [12,13]. This methodology provides an opportunity to investigate the extent to which MD simulations can account theoretically for the particular behavior observed for U1A and investigate the nature of the transition state ensemble at the molecular level. In the course of this study, we have combined previous procedures into a practical method for calculating a statistical free energy landscape in RMSD space from MD simulations.

We describe, herein, all-atom MDUF simulations including solvent on U1A and 13 mutants. Our simulations address the following specific questions: (a) Does the MDUF energy landscape for U1A support the interpretation of kinetic experiments in terms of a high energy channel and malleable transition state? (b) What is the molecular model obtained from MDUF for the transition state ensemble? (c) What is the mechanism of unfolding indicated by the MDUF simulations and is this relevant to folding pathways? (d) What elements of secondary structure constitute the nucleation site, and what residues are involved in the folding nucleus? (e) How do the MD results on U1A unfolding compare with mechanisms and structures derived in previous experimental and theoretical studies on U1A? This project builds on our previous studies of all-atom, explicit solvent MD on the equilibrium dynamics of U1A and F56 mutants [14–18].

## 2. Background

The protein U1A belongs to a family of RNA recognition motifs (RRM), one of the most abundant protein domains in eukaryotes [19]. Proteins containing RRM participate in all steps of gene expression and RNA processing and bind to single-stranded RNAs of diverse sequences in a variety of structural contexts [20,21]. The crystal structure of U1A complexed with a 21-mer stem-loop RNA [22] shows a four stranded, antiparallel  $\beta$ -sheet flanked by two  $\alpha$ -helices, characteristic of an RRM, and organized according to the pattern  $\beta_1\alpha_A\beta_2\beta_3\alpha_B\beta_4(\alpha_C)$  with respect to sequence and secondary structure (Figure 1a). In the native tertiary structure, the RRM is organized as an antiparallel  $\beta$ -sheet  $\beta_4\beta_1\beta_3\beta_2$  with  $\alpha_A$  and  $\alpha_B$  as supports. The structure of uncomplexed U1A in solution [23] (Figure 1b), is similar to that of the complex with RNA with the exception of the C-terminal helix  $\alpha_C$ , which folds over and obstructs the access of ligand RNA to the RRM. The intrinsic stability of the RRM in U1A and the dynamics of  $\alpha_C$  and also loop 3 contribute significantly to U1A-RNA recognition and are thus of particular interest.



**Figure 1.** (a) Crystal Structure of the protein U1A-Stem loop 2 RNA Complex [24], with secondary structure and other elements annotated; (b) NMR solution structure of uncomplexed U1A [23], structure #5 from the ensemble of pdb #1fht.

Silow and Oliveberg [25] measured experimentally the in vitro kinetics of U1A folding and unfolding as a function of denaturant concentration using stopped flow and analyzed the results with chevron plots [1]; linearity in the arms of the chevron is diagnostic of a two-state process [26]. In the case of U1A, a deviation from simple two-state behavior was indicated by a slight but significant curvature in the arms. This was interpreted based on  $\beta$ -values [1] as evidence for large movements and changes in solvent exposure in the transition state structures along the top of a very broad free energy barrier. Further experimental studies of the formation and growth of the folding nucleus of U1A based on  $\phi$ -value analysis were reported by Ternstrom et al. [27] and supported the delocalized nature of the transition state. A model was proposed, in which the folding nucleus involved  $\beta_2$ ,  $\beta_3$  and  $\alpha_A$  of the RRM. A previous theoretical study of wild-type U1A and various mutants at a coarse-grained level has been reported based on a variational free energy functional theory [28] and provided a description of the growth of the folding nucleus in terms of residue-based isodensity surfaces [29]. Lindberg and Oliveberg [30] discussed the folding of U1A and other proteins in terms of modular assemblies of competing two-strand-helix nucleation motifs and pointed out that pathway malleability may be a more general phenomenon in protein folding and induced in a variety of ways. This is supported by recent work on the homeodomain superfamily [31].

Early studies of MD applied to protein unfolding were due to Daggett and Levitt [32,33], and the current state of the field has been recently reviewed [6,11,13,34]. Notably, an extensive database of protein dynamics has been created [35]. There have been a number of subsequent articles describing the use of MD to study the thermal unfolding of proteins and to characterize transition states [11]. Daggett and coworkers have defined a general protocol for MDUF simulations [36] and documented the applicability of MDUF to realistic protein unfolding, including the correspondence between thermal unfolding and chemical denaturation for small proteins [37], and that high temperature accelerates the unfolding of small proteins without essentially changing the pathway [38,39]. Results from 100 independent MDUF simulations on CI2 [40], showed that 5–10 MDUF trajectories are required to capture properly the average properties. Daggett and coworkers in a number of studies [11] have utilized the calculated pair-wise  $C^\alpha$  RMSD matrix (2DRMSD) reduced by classical multidimensional scaling (MDS) [12,13,41], to represent a trajectory by plotting the vector components of the RMSD difference between successive MD structures (MD step vectors) in the RMSD space. In this method, the transitional region for protein unfolding is defined as the first point of exit from the native state ensemble, i.e., the first precipitous change in the RMSD between successive MDUF structures. This method is a type of leader algorithm [42] and has appeared elsewhere in the protein folding literature [43]. Stereoscopic views of MD step vector plots in the 3D RMSD space have been presented [13,44], but typically a 2D projection has been sufficient to elucidate the transition [12].

There is considerable literature on the determination of reaction coordinates for protein folding and unfolding. One line of approach involves using linear or non-linear MDS to reduce the MD trajectory to a representation in terms of order parameters (i.e., native contacts, the radius of gyration, etc.) in two or three dimensions that can be readily interpreted and used to calculate free energy landscapes and locate transition states [45–48]. A concern with this type of approach is that depending on the order parameter, MDS reduction may lead to an oversimplified free energy landscape, especially in the region of highly unfolded structures [49,50], but the specifics of this issue are unsettled [11] and await more comparative studies. More rigorous methods have been proposed [49,51], and enable analyses that are even finer-grained than the level of resolution of most experiments. Theoretically, structures that comprise the transition state ensemble should have an equal probability  $P_{fold}$  to fold and unfold [52,53]. A critical consideration and further methodological developments based on  $P_{fold}$  type approaches have been provided by Muff and Caflisch [43]. Several recent studies have reported transition state ensembles for proteins comprised of structures with  $P_{fold} = 0.5$  [13,54]. Beck and Daggett [13] demonstrated that the transition state ensemble obtained by the RMSD space method is quite similar to that of

the computationally much more intensive  $P_{fold}$  method, and showed that the transition state ensemble obtained by their method is quite similar to that based on structures which have an equivalent propensity to fold and unfold. Transition disconnectivity graphs [55] and complex network analysis [50] provide additional tools for detailed analysis of the energy and free energy landscapes, oftentimes used together with the above methods [49,51]. Daggett and coworkers have also proposed a comprehensive embedded 1D reaction coordinate for protein unfolding and folding [56].

### 3. Materials and Methods

#### 3.1. Molecular Dynamics Simulations

The MD simulations in this study are carried out on an all atom, explicit solvent model of the system using the *AMBER* suite of programs [57]. Explicit consideration of solvent water and ions is necessary to describe the molecularity of water–protein hydrogen bonding and the hydrophobic effect [11]. The simulations are based on the protein force field *parm99SB* [58], which performed among the best in a recent study of the various options for MD on proteins [8]. The *TIP3P* force field [59] was used for water molecules and *ions08* [60] was used for monovalent ions. The system for simulation studies was configured under minimal neutralizing salt conditions, and comprised of protein, ~5000 water molecules, and eight  $\text{Cl}^-$  ions. All MDs were performed under constant volume (T, V, N) ensemble conditions with particle-mesh Ewald periodic boundary conditions. For the integration of the equations of motion, an MD time step of 1 fs was adopted. Our simulations follow closely the MDUF protocol of Beck and Daggett [36], with a water density of 0.829 gm/cc and a temperature of 498 K. All MD simulations were carried out from an energy minimized form of the NMR solution structure, of U1A, structure #5 of PDB ID #1fht (Figure 1b). The structures of the NMR ensemble are all very similar. To ensure accuracy and reproducibility, five independent thermal unfolding simulations with different initial velocity distributions were carried out.

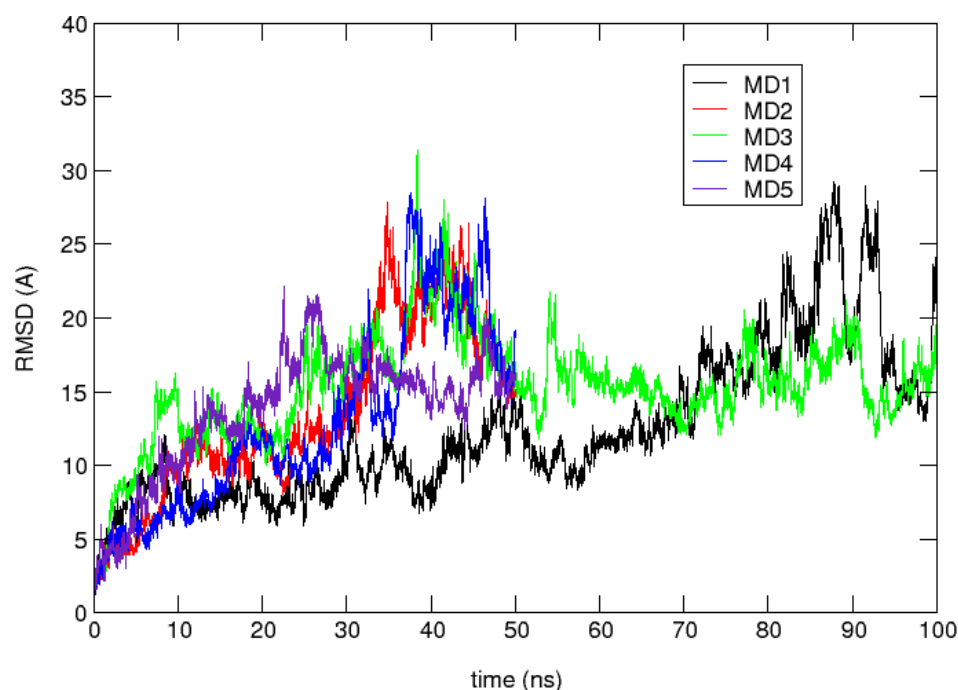
#### 3.2. Analysis

MD calculated free energy landscapes were obtained using a particular combination of two existing methodologies. The first method, as discussed above, is the representation of a simulation in terms of a plot of MD step vectors in RMSD space [12,13]. The native state is represented as a relatively tight cluster of step vectors. In a system with potentially extended activation barriers, the first point of exit from the native state ensemble defines the onset of the UF transition region, but a demarcation of the outer bound of the transition is also necessary (see below). The method is quite general and has been applied extensively to calculate free energy surfaces for proteins with respect to the first two eigenvectors of a principal component analysis (PCA) of a protein MD trajectory [61–64]. Here, a 2D grid is defined on a representation of a trajectory in PCA eigenvector space and the population of structures  $P_{ij}$  associated with each grid element is determined. From this, normalized probabilities  $P_{ij}/P_0$  can be obtained, where  $P_0$  is the maximum population. Statistical free energy is then calculated as  $\Delta F_{ij} = -RT \ln (P_{ij}/P_0)$  and displayed as a contour map. In PCA, one typically considers a subspace of the motions defined by the two most important principal components of the motions (the essential dynamics [61]) which typically accounts for >2/3 of the total motion and sometimes more, depending on the system. Similar applications of this approach have been described in terms of collective variables, such as native contacts [65,66] and radius of gyration  $R_g$  [13]. The free energy landscapes for this study are obtained in an analogous manner but are based on population maps of the terminals of MD step vectors in the 3D of RMSD space. An advantage of this approach is that 100% of the unfolding trajectory is represented, albeit in a space of reduced dimensionality.

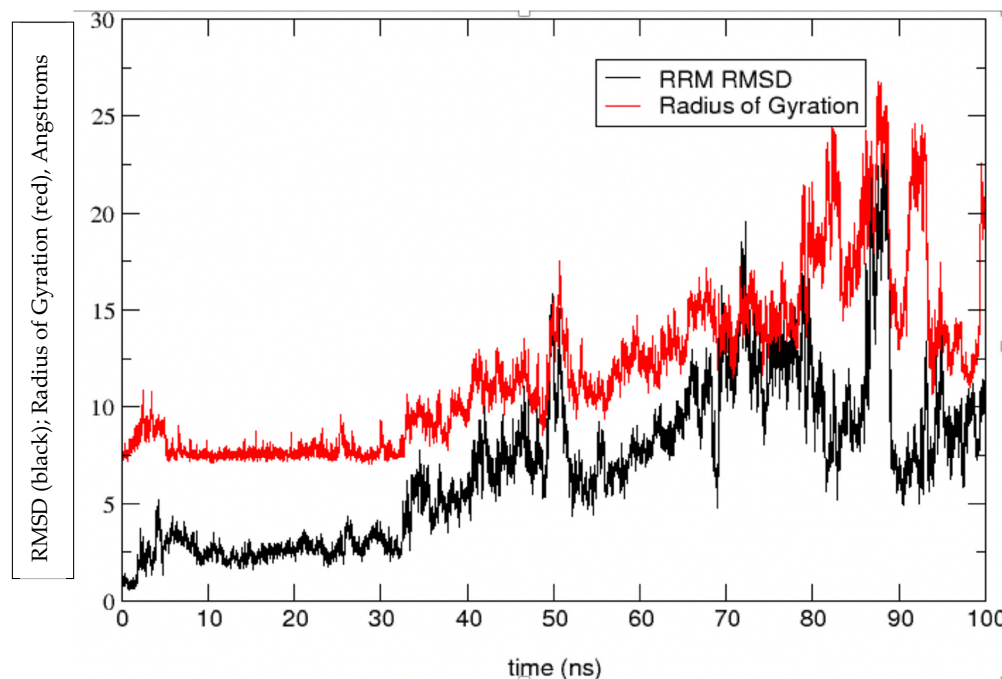
#### 4. Results and Discussion

Validation of the MDUF was first considered by plotting the C $^{\alpha}$  RMSD with respect to the initial MD structure and the corresponding radius of gyration  $R_g$  as a function of time. The results are shown in Figures 2 and 3, respectively. The MD calculated C $^{\alpha}$  RMSD and  $R_g$  closely parallel one another, showing an initial rapid departure from the initial configuration, equilibration after ~10 ns to a meta-stable region of ~30 ns in which the MD structures cluster in the vicinity of ~7.5 Å rmsd. After ~30 ns, the protein unfolds more or less monotonically until 80 ns. From there on until the MD was terminated at 100 ns, the protein exhibits high amplitude oscillations in the range of 10 Å to 27 Å RMSD from the initial structure.

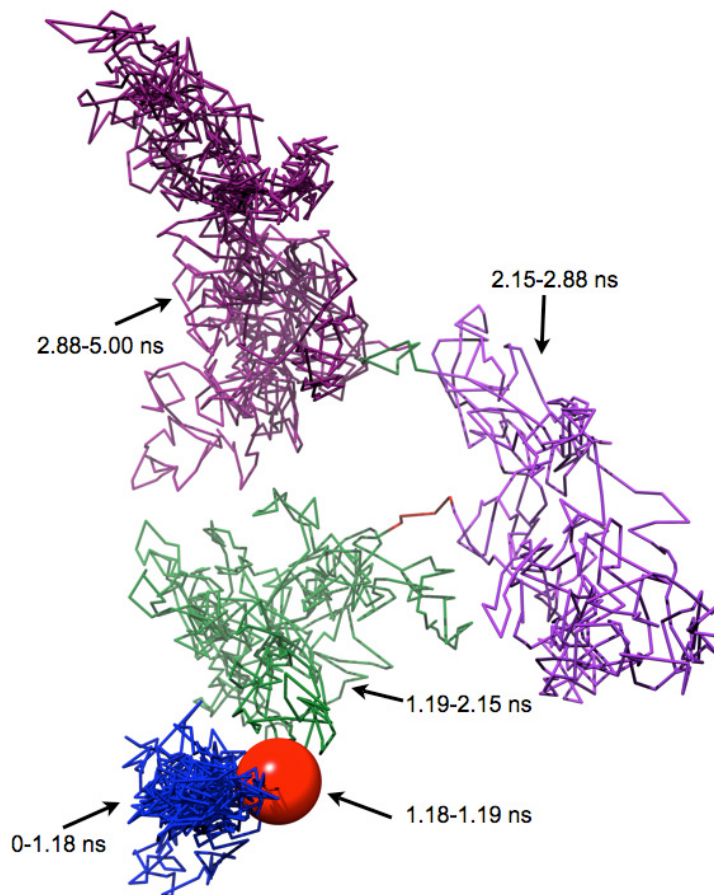
An RMSD vector representation of the MDUF trajectory for U1A in a 2D projection of the 3D RMSD space is shown in Figure 4. The onset of UF transition occurs in the region of ~1.18–1.19 ns (indicated as a red sphere). Using the methodology described in the preceding section, a population map was obtained and converted to a statistical free energy landscape. The result is presented as a contour map in Figure 5. This describes the MD UF transition as proceeding from the native state (lower left corner) over a free energy barrier and continuing along a relatively extended valley on the surface. A UF reaction coordinate is constructed from perpendiculars to the isoenergy contour lines. The free energy for the first 5 ns of MD simulated unfolding of U1A as a function of this coordinate is plotted in Figure 6. The plot indicates an activation free energy of 3.8 kcal/mol. Using the transition state theory for proteins [1] the  $k_u$  is approximately  $10^6 \exp(-\Delta F/RT)$  which computes to  $2.7 \times 10^{-3} \text{ s}^{-1}$ . The free energy profile in Figure 6 for the first 5 ns of MDUF shows that the activation process is to be followed by a slight free energy minimum before unfolding further. However, the well depth of this minimum is only ~0.2 kcal/mol, so this does not correspond to a thermally stable intermediate. Thus, the MD UF results on U1A describe a two-state process.



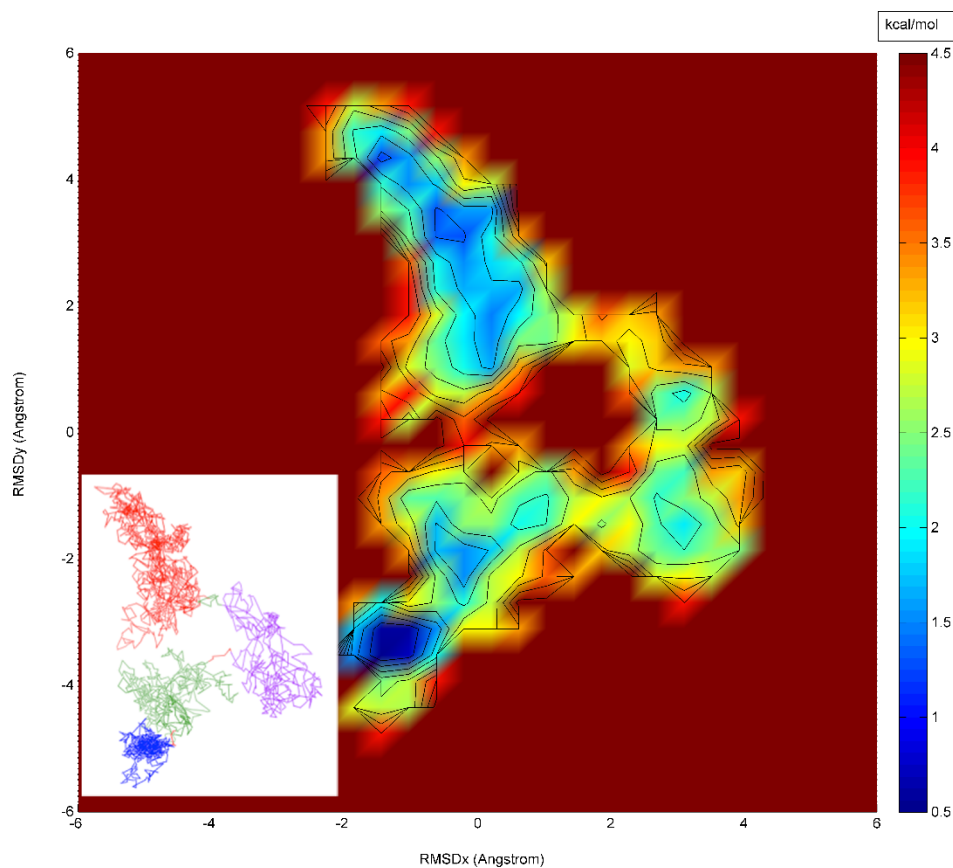
**Figure 2.** Root mean square deviation (RMSD) with respect to the initial structure as a function of time for 5 independent MDUF trajectories for U1A. Only MD3 and MD5 were extended; all were showing essentially similar behavior after 50 ns.



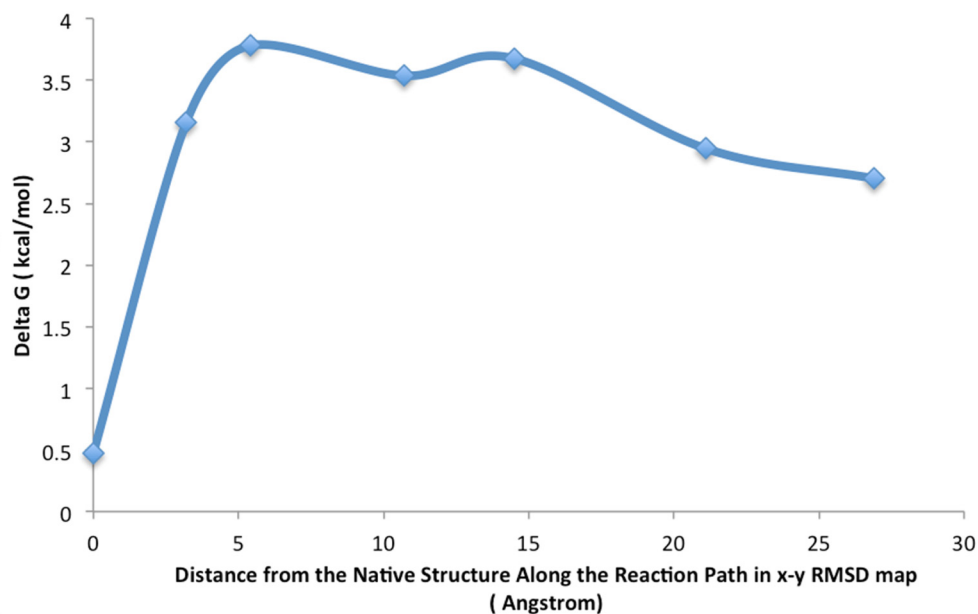
**Figure 3.** Radius of Gyration (red line) and RMSD (black line) of the RNA recognition motif of U1A with respect to the initial structure of MDUF simulation #1 on U1A.



**Figure 4.** Vectorial representation of the first 5 ns of a 100 ns MDUF trajectory #1 in RMSD space obtained using the RMSD method of Daggett and coworkers [12,13]. The coloring distinguishes the various transitional intermediates.

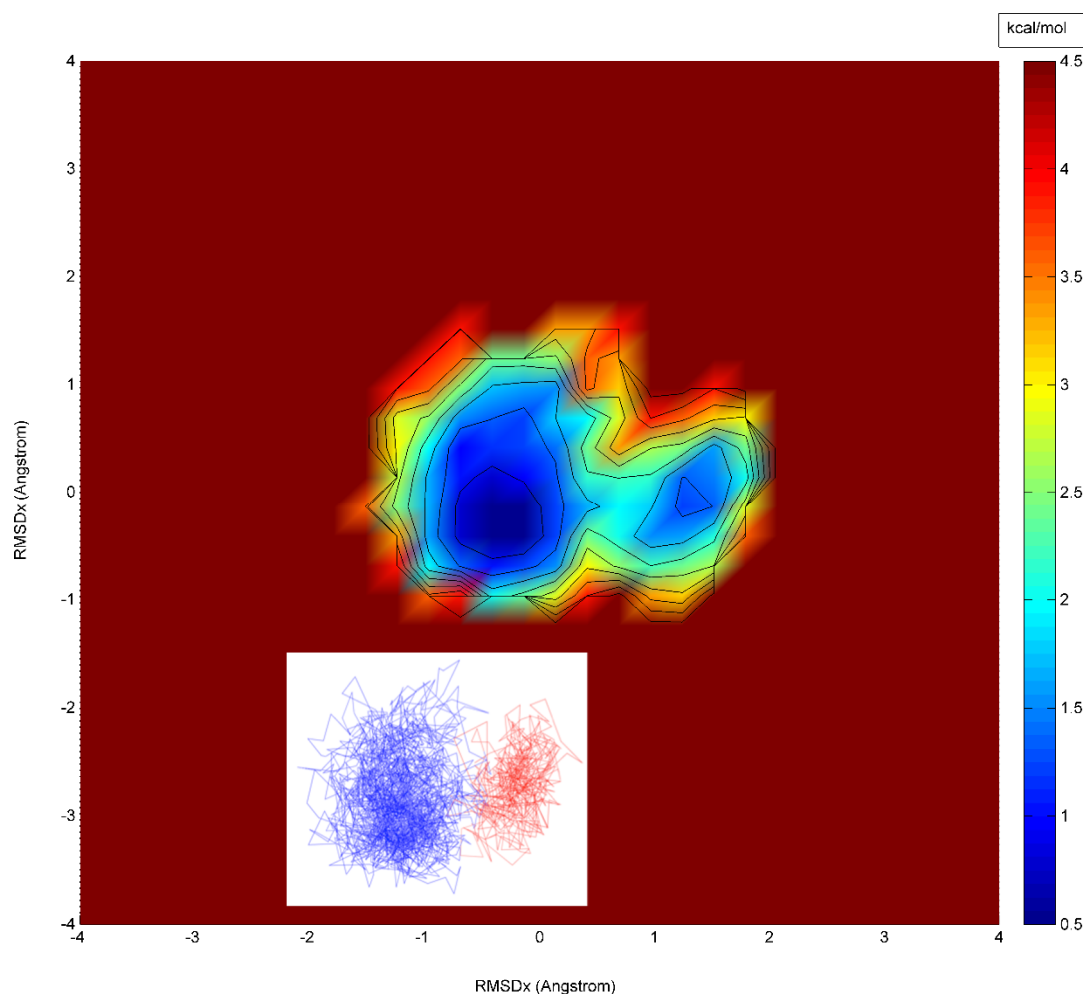


**Figure 5.** Calculated PCA free energy landscape for the first 5 ns of a 100 ns MDUF trajectory #1 on U1A. The (non-equilibrium) unfolding reaction coordinate based on the gradient of isoenergy contour lines is indicated as a dotted line. Color scheme: yellow to green to blue: high to medium to low free energies, respectively.



**Figure 6.** Calculated free energy profile for U1A unfolding with respect to the reaction coordinate indicated in Figure 4. The depth of the slight minimum after the activation barrier is ~0.2 kcal/mol and does not correspond to a thermally stable intermediate.

To confirm that our calculated unfolding pathway involves a high energy channel and extended transition state, we have taken as a control, the well-characterized [12] two-state protein CI2, and performed an analogous MDUF with a similar protocol to that used for U1A and calculated a corresponding free energy landscape (Figure 7). The result is clearly consistent with that of a simple two-state process, with a narrow and sharply peaked barrier to activation and the transition state covered in an interval of <1 ns. Comparing the free energy landscape calculated for U1A in Figure 5 with that of CI2 in Figure 7, the results indicate that the MDUF on U1A does indeed show a relatively extended transition state. Similar results were obtained from the analysis of our other four MDUF simulations on U1A.

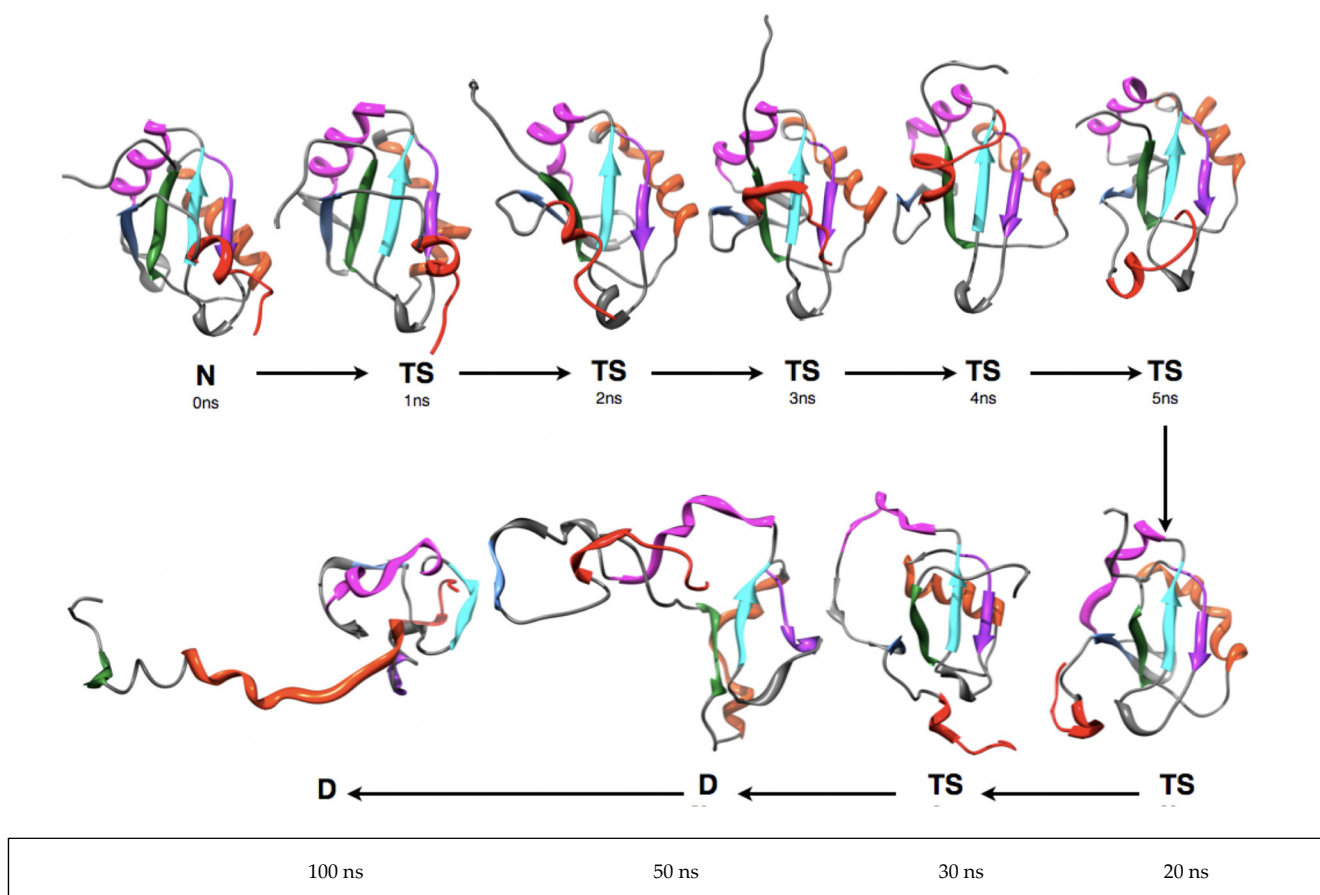


**Figure 7.** Calculated PCA free energy landscape for an MDUF trajectory on CI2. Color scheme: yellow to green to blue: high to medium to low free energies, respectively.

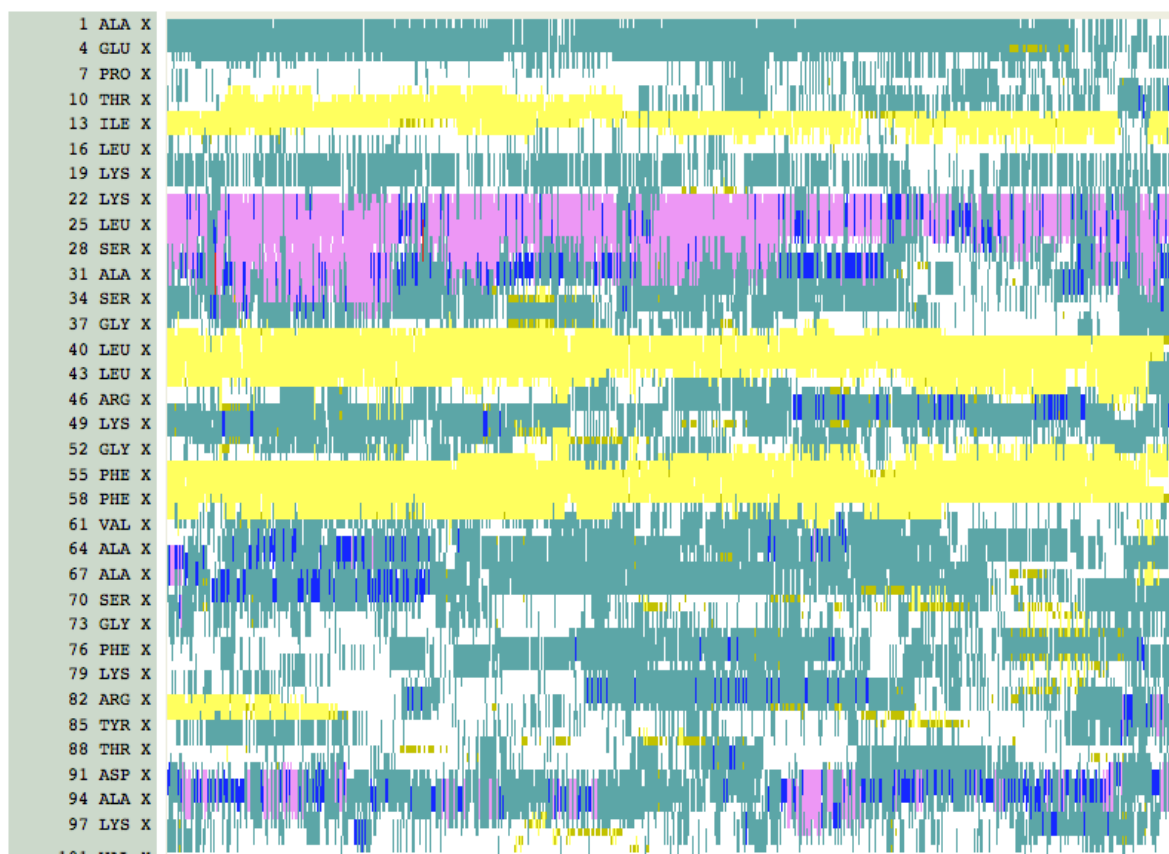
A time series of structures from MDUF #1 on U1A is shown in Figure 8. Here, again similar results were obtained from our other four independent MDs on U1A. By 1 ns into the trajectory, helix  $\alpha_C$  shows dynamical reorientation which continues while the helix is still intact. This feature is expected since  $\alpha_C$  dynamics are directly implicated in the U1A-RNA binding mechanism. From there on, the overall structure expands. Between 2 and 3 ns, the helix  $\alpha_A$  exhibits an oscillation between its position in the native state and a considerable displacement, but the antiparallel  $\beta$ -sheet remains assembled. Beyond 3 ns,  $\beta_4$  separates and at 4 ns is perpendicular to the  $\beta_2\beta_3\beta_4$  motif. At this point,  $\alpha_B$  separates more from the remaining core of the structure. At 5 ns,  $\beta_4$  has returned and the four-stranded  $\beta$  sheet has reformed but remains expanded relative to the native state. In the next interval of time,  $\beta_4$  oscillates between an antiparallel form with  $\beta_2\beta_3\beta_4$  and a form perpendicular



to it. These structures represent only the early stages of the unfolding transition state (see below). Overall, the RRM is qualitatively intact but shows a 14% increase in  $R_g$ , i.e., slightly expanded compared to the native state structure. The antiparallel  $\beta_1\beta_3\beta_2$  tertiary motif is still intact after 20 ns. Over the next 10 ns,  $\beta_4$  along with  $\alpha_B$  becomes detached from the  $\beta_1\beta_3\beta_2$  motif and the secondary structure of both dissipates. At 50 ns, a large portion of the structure has unfolded, but interestingly the  $\beta_1\beta_3\beta_2$  remains somewhat intact and  $\alpha_A$  has returned back into closer proximity to the truncated  $\beta$ -sheet. It is interesting that 2/3 of the RRM motif  $\beta_1\alpha_A\beta_2\beta_3\alpha_B\beta_4$  are relatively intact in an albeit expanded, molten globule-like form up to 35 ns of MDUF. By 100 ns the tertiary structure has unfolded further, although even at this point vestiges of the partially formed secondary structure remain, and a small but non-native element of tertiary assembly is found. A plot of the time evolution of secondary structure is shown in Figure 9, and clearly shows the persistence of the  $\beta_1\alpha_A\beta_2\beta_3$  motif. Our results indicate that the folding nucleus of U1A is comprised of a substantial amount of the RRM. The MDUF trajectory for U1A read in the folding direction follows a nucleation-condensation mechanism [67].

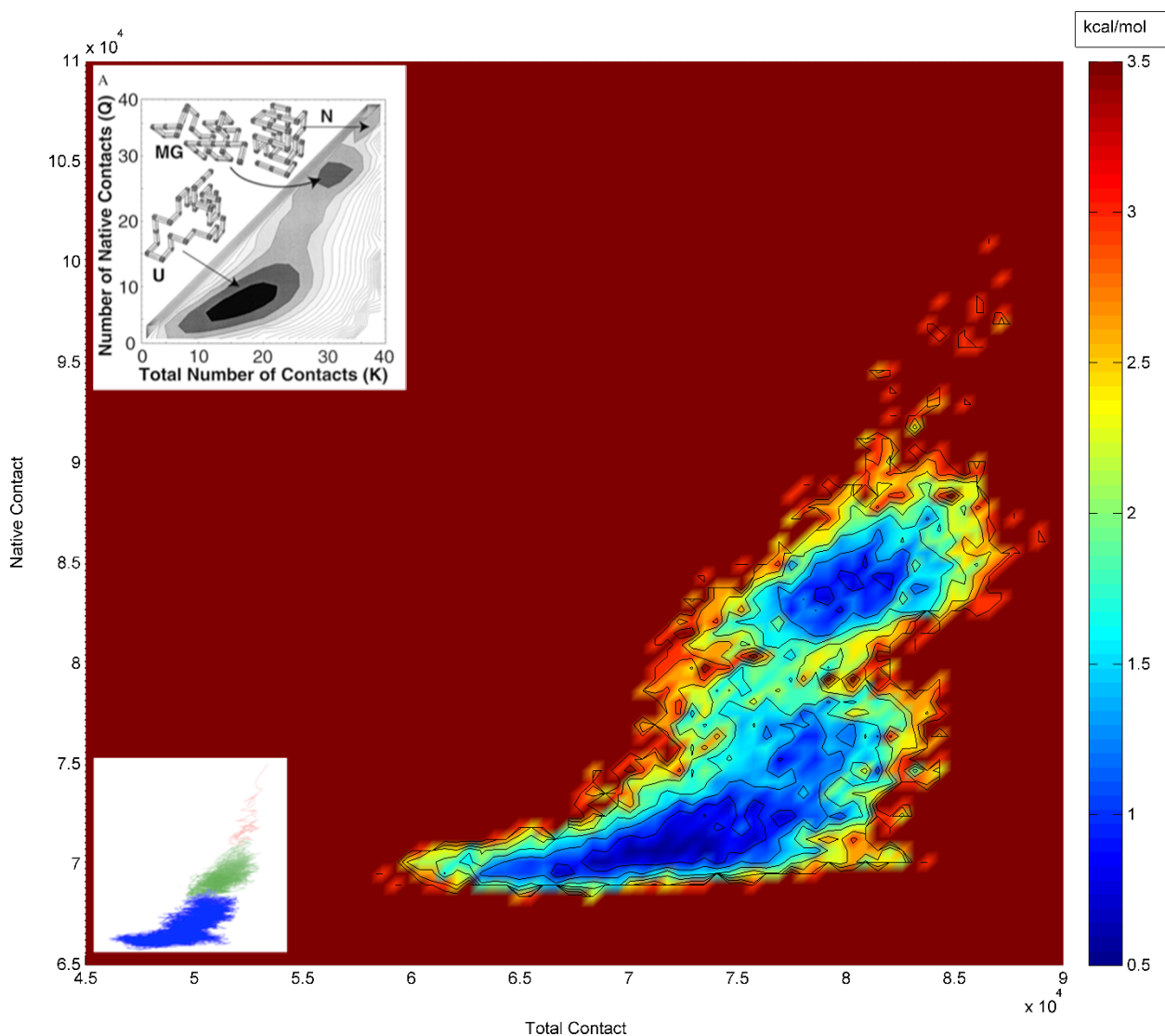


**Figure 8.** A series of MD snapshots from 100 ns of the MDUF trajectory #1 for U1A.



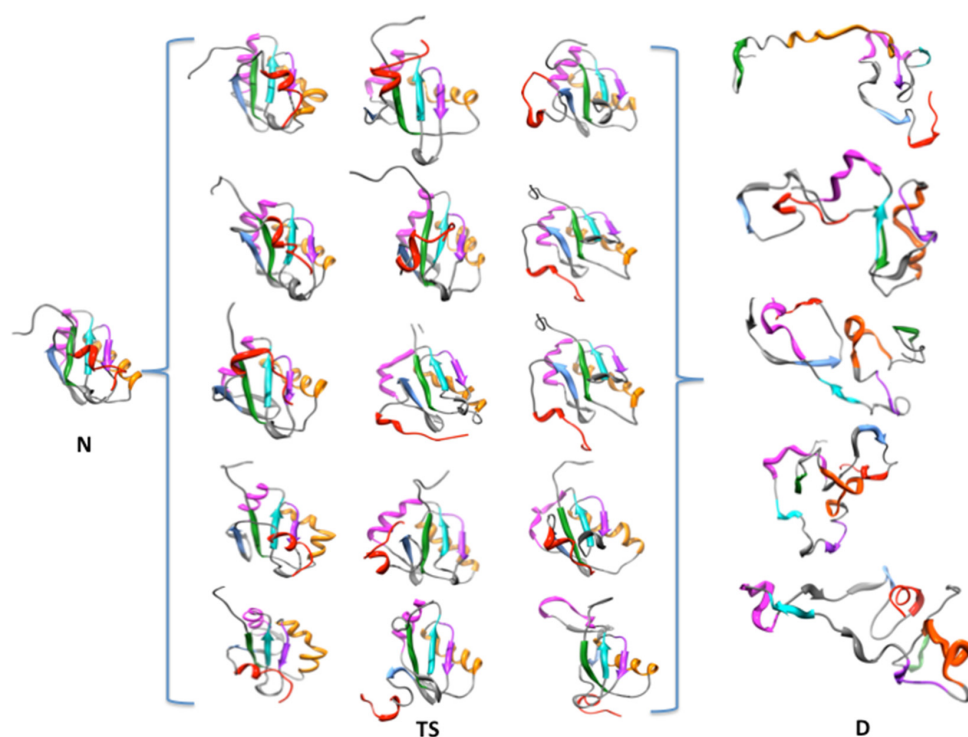
**Figure 9.** Evolution of secondary structure in the MDUF trajectory #1 of U1A. Key: Yellow:  $\beta$ -sheets. Red:  $\alpha$ -helices. This plot was obtained using VMD [68].

Molten globules are protein structures that retain native-like secondary structures but lack the close packing of the native state ensemble [69]. To explore further if this applies to the transition state ensemble of U1A, we refer to the study of Pande and Rokhsar [65,70], who simulated the folding/unfolding pathway of a lattice model protein using Monte Carlo methods. In their analysis, they plotted the number of contacts in the native state vs. the total number of contacts and found that a molten globule state was well-defined. In Figure 10, we show the corresponding plot from the entire 100 ns MDUF simulation on U1A; Figure 1a of Pande et al. [70] is reproduced in an inset. Their features of the prototype unfolding are found to be well reproduced by our all-atom MDUF simulation on U1A, consistent with the behavior of a molten globular intermediate. The regions indicated to be molten globules by this analysis occur between 1.2 and 35 ns of the 100 ns trajectory, which we take as the demarcation of the upper bound of the extended transition region. Molecular structures from the transition state ensemble including results from all five MDUF simulations are shown in Figure 11. The transitional structures all show by inspection, variations on the theme of protein molten globules. Thus, MDUF simulations predict that the malleable transition state of U1A is an ensemble of molten globule microstates.

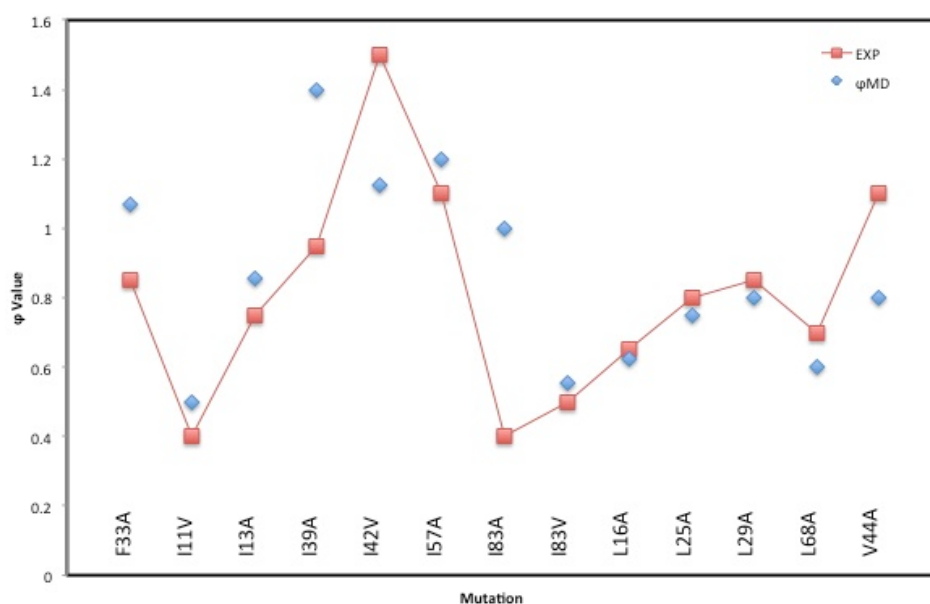


**Figure 10.** Evolution of secondary structure in the MDUF trajectory #1 of U1A. Key: Yellow:  $\beta$ -sheets. Red:  $\alpha$ -helices. This plot was obtained using VMD [68]. Top left insert: Corresponding lattice model results from Pande and Rokhsar [70].

To further validate our MD simulations, we present the results of 100 ns MDUF simulations on the 13 mutant forms of U1A for which the experimentally observed  $\phi$ -values are measures of how the energetics of protein folding are changed by single-site mutations [27]. Each of these simulations was performed using a protocol analogous to that described for our MD on wild-type U1A. The MD  $\phi$ -values were calculated from the ratio of the difference between average native contacts in the mutant and wild-type transition states divided by the difference between average native contacts in the folded mutant and wild-type states of U1A [12]. A comparison of the observed and MD calculated results on  $\phi$ -values is shown in Figure 12. The trends observed are well reproduced by the MD calculated values, especially since the observed results are from chemical denaturation and the calculated results are from thermal denaturation. This supports the idea that the mechanisms of chemical and thermal denaturing processes for small proteins are similar, but this is likely to be system-specific.



**Figure 11.** Results from 5 MDUF simulations: the brackets contain an ensemble of transition state structures.



**Figure 12.** A comparison of experimentally observed [27], and MD calculated  $\phi$ -values for 13 mutants forms of U1A.

## 5. Summary and Conclusions

All-atom, explicit solvent MD thermal unfolding studies on U1A result in a free energy landscape which is consistent with observed kinetics of denaturation in terms of a high energy folding channel. The MD structures corresponding to this region provide a detailed all-atom molecular model for the malleable transition state, with the variation due to sampling different configurations of dynamical elements of the structure, while the RRM core remains intact but slightly expanded. This conforms to the definition of molten globules, and thus our MDUF on U1A indicates the transition state for unfolding

is an ensemble of molten globule states. MD calculated results are generally consistent interpretations of experimental data from chevron plots and  $\phi$ -value analysis. This also provides support for the relevance of free energy landscapes in RMSD space to the protein unfolding problem. The observed progress of the folding is consistent with the expansion of residue-based isodensity surfaces [29] and provides a molecular model in support of the mechanism proposed from  $\phi$ -values studies that the nucleation motif likely involves a two-sheet helix motif [27,30]. In this project, we compare the results of MD on thermal unfolding with experimental results on chemical denaturation, and the general agreement obtained, supports the idea that the folding pathway is similar in the two cases [5,39,71,72]. To further investigate this point, a corresponding MD study of the chemical denaturation of U1A and a corresponding comparison with experimental data is currently in progress. A more detailed analysis of the kinetics is being carried out using the program Wordom [73,74].

*Note added in proof:* A referee has asked that we comment on the sensitivity of our results to our choice of the AMBER force field. Rueda, Ferrer-Costa, Meyer and Orozco [75] have reported an extensive longitudinal comparison of results from the four most popular MD force fields including AMBER, which indicates that MD results at the level of resolution of our study are expected to be force field independent.

**Author Contributions:** Conceptualization: D.L.B., A.M.B., N.L.D.; methodology: D.L.B., N.L.D.; software and calculations: N.L.D.; formal analysis: N.L.D., D.L.B.; writing—original draft preparation: D.L.B., N.L.D.; writing—review and editing, D.L.B.; funding acquisition: D.L.B., A.M.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded in part by NSF Grant #A2950 to AMB and NIH Grant #1-R15-GM129715 to DLB. General support for this project was provided by NIGMS Training grant GM-008271. This work was partially supported by the National Science Foundation through the TeraGrid resources provided by grant number CHE10004 for use of the Ranger Cluster at the Texas Advanced Computing Center.

**Acknowledgments:** The authors would like to acknowledge Valerie Daggett and Saraswathi Vishveswara for helpful discussions and sharing of analysis routines.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Fersht, A.R. *Structure and Mechanism in Protein Science*; W.H. Freeman & Company: New York, NY, USA, 1998.
2. Oliveberg, M.; Wolynes, P.G. The experimental survey of protein-folding energy landscapes. *Q. Rev. Biophys.* **2005**, *38*, 245–288. [[CrossRef](#)]
3. Ladurner, A.G.; Itzhaki, L.S.; Daggett, V.; Fersht, A.R. Synergy between simulation and experiment in describing the energy landscape of protein folding. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 8473–8478. [[CrossRef](#)]
4. Daggett, V. Long timescale simulations. *Curr. Opin. Struct. Biol.* **2000**, *10*, 160–164. [[CrossRef](#)]
5. Mayor, U.; Johnson, C.M.; Daggett, V.; Fersht, A.R. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 13518–13522. [[CrossRef](#)] [[PubMed](#)]
6. Dill, K.A.; Ozkan, S.B.; Shell, M.S.; Weikl, T.R. The protein folding problem. *Annu. Rev. Biophys.* **2008**, *37*, 289–316. [[CrossRef](#)] [[PubMed](#)]
7. Simmerling, C.; Strockbine, B.; Roitberg, A.E. All-Atom Structure Prediction and Folding Simulations of a Stable Protein. *J. Am. Chem. Soc.* **2002**, *124*, 11258–11259. [[CrossRef](#)]
8. Piana, S.; Lindorff-Larsen, K.; Shaw, D.E. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **2011**, *100*, L47–L49. [[CrossRef](#)]
9. Li, A.; Daggett, V. Identification and Characterization of the Unfolding Transition State of Chymotrypsin Inhibitor 2 by Molecular Dynamics Simulations. *J. Mol. Biol.* **1996**, *257*, 412–429. [[CrossRef](#)]
10. Fersht, A.R.; Daggett, V. Protein Folding and Unfolding at Atomic Resolution. *Cell* **2002**, *108*, 573–582. [[CrossRef](#)]
11. Daggett, V. Protein folding-simulation. *Chem. Rev.* **2006**, *106*, 1898–1916. [[CrossRef](#)]
12. Li, A.; Daggett, V. Characterization of the transition state of protein unfolding by use of molecular dynamics: Chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 10430–10434. [[CrossRef](#)] [[PubMed](#)]
13. Beck, D.A.; Daggett, V. A one-dimensional reaction coordinate for identification of transition states from explicit solvent P(fold)-like calculations. *Biophys. J.* **2007**, *93*, 3382–3391. [[CrossRef](#)] [[PubMed](#)]
14. Kormos, B.L.; Baranger, A.M.; Beveridge, D.L. Do Collective Atomic Fluctuations Account for Cooperative Effects? Molecular Dynamics Studies of the U1A–RNA Complex. *J. Am. Chem. Soc.* **2006**, *128*, 8992–8993. [[CrossRef](#)]

15. Zhao, Y.; Kormos, B.L.; Beveridge, D.L.; Baranger, A.M. Molecular dynamics simulation studies of a protein–RNA complex with a selectively modified binding interface. *Biopolymers* **2005**, *81*, 256–269. [[CrossRef](#)] [[PubMed](#)]
16. Kormos, B.L.; Benitex, Y.; Baranger, A.M.; Beveridge, D.L. Affinity and Specificity of Protein U1A-RNA Complex Formation Based on an Additive Component Free Energy Model. *J. Mol. Biol.* **2007**, *371*, 1405–1419. [[CrossRef](#)]
17. Anunciado, D.; Agumeh, M.; Kormos, B.L.; Beveridge, D.L.; Knee, J.L.; Baranger, A.M. Characterization of the Dynamics of an Essential Helix in the U1A Protein by Time-Resolved Fluorescence Measurements. *J. Phys. Chem. B* **2008**, *112*, 6122–6130. [[CrossRef](#)]
18. Kormos, B.L.; Pieniazek, S.N.; Beveridge, D.L.; Baranger, A.M. U1A protein-stem loop 2 RNA recognition: Prediction of structural differences from protein mutations. *Biopolymers* **2011**, *95*, 591–606. [[CrossRef](#)]
19. Maris, C.; Dominguez, C.; Allain, F.H.-T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **2005**, *272*, 2118–2131. [[CrossRef](#)]
20. Varani, G.; Nagai, K. Rna Recognition By Rnp Proteins During Rna Processing. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 407–445. [[CrossRef](#)]
21. Cléry, A.; Blatter, M.; Allain, F.H.-T. RNA recognition motifs: Boring? Not quite. *Curr. Opin. Struct. Biol.* **2008**, *18*, 290–298. [[CrossRef](#)]
22. Nagai, K.; Oubridge, C.; Ito, N.; Jessen, T.H.; Avis, J.; Evans, P. Crystal structure of the U1A spliceosomal protein complexed with its cognate RNA hairpin. *Nucleic Acids Symp. Ser.* **1995**, *34*, 1–2.
23. Avis, J.M.; Allain, F.H.-T.; Howe, P.W.; Varani, G.; Nagai, K.; Neuhaus, D. Solution Structure of the N-terminal RNP Domain of U1A Protein: The Role of C-terminal Residues in Structure Stability and RNA Binding. *J. Mol. Biol.* **1996**, *257*, 398–411. [[CrossRef](#)] [[PubMed](#)]
24. Oubridge, C.; Ito, N.; Evans, P.R.; Teo, C.H.; Nagai, K. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **1994**, *372*, 432–438. [[CrossRef](#)] [[PubMed](#)]
25. Silow, M.; Oliveberg, M. High-Energy Channeling in Protein Folding. *Biochemistry* **1997**, *36*, 7633–7637. [[CrossRef](#)]
26. Jackson, S.E.; Fersht, A.R. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **1991**, *30*, 10428–10435. [[CrossRef](#)]
27. Ternstrom, T.; Mayor, U.; Akke, M.; Oliveberg, M. From snapshot to movie: Phi analysis of protein folding transition states taken one step further. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14854–14859. [[CrossRef](#)]
28. Plotkin, S.S.; Onuchic, J.N. Investigation of routes and funnels in protein folding by free energy functional methods. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6509–6514. [[CrossRef](#)]
29. Shen, T.; Hofmann, C.P.; Oliveberg, M.; Wolynes, P.G. Scanning Malleable Transition State Ensembles: Comparing Theory and Experiment for Folding Protein U1A. *Biochemistry* **2005**, *44*, 6433–6439. [[CrossRef](#)]
30. Lindberg, M.O.; Oliveberg, M. Malleability of protein folding pathways: A simple reason for complex behaviour. *Curr. Opin. Struct. Biol.* **2007**, *17*, 21–29. [[CrossRef](#)]
31. Banachewicz, W.; Religa, T.L.; Schaeffer, R.D.; Daggett, V.; Fersht, A.R. Malleability of folding intermediates in the homeodomain superfamily. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 5596–5601. [[CrossRef](#)]
32. Daggett, V.; Levitt, M. Molecular dynamics simulations of helix denaturation. *J. Mol. Biol.* **1992**, *223*, 1121–1138. [[CrossRef](#)]
33. Daggett, V.; Levitt, M. Protein Unfolding Pathways Explored Through Molecular Dynamics Simulations. *J. Mol. Biol.* **1993**, *232*, 600–619. [[CrossRef](#)] [[PubMed](#)]
34. Scheraga, H.A.; Khalili, M.; Liwo, A. Protein-Folding Dynamics: Overview of Molecular Simulation Techniques. *Annu. Rev. Phys. Chem.* **2007**, *58*, 57–83. [[CrossRef](#)] [[PubMed](#)]
35. van der Kamp, M.W.; Schaeffer, R.D.; Jonsson, A.L.; Scouras, A.D.; Simms, A.M.; Toofanny, R.D.; Benson, N.C.; Anderson, P.C.; Merkley, E.D.; Rysavy, S.; et al. Dymeomics: A comprehensive database of protein dynamics. *Structure* **2010**, *18*, 423–435. [[CrossRef](#)] [[PubMed](#)]
36. Beck, D.A.; Daggett, V. Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods* **2004**, *34*, 112–120. [[CrossRef](#)]
37. Bennion, B.J.; Daggett, V. The molecular basis for the chemical denaturation of proteins by urea. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5142–5147. [[CrossRef](#)]
38. Day, R.; Bennion, B.J.; Ham, S.; Daggett, V. Increasing Temperature Accelerates Protein Unfolding Without Changing the Pathway of Unfolding. *J. Mol. Biol.* **2002**, *322*, 189–203. [[CrossRef](#)]
39. Gianni, S.; Guydosh, N.R.; Khan, F.; Caldas, T.D.; Mayor, U.; White, G.W.N.; DeMarco, M.L.; Daggett, V.; Fersht, A.R. Unifying features in protein-folding mechanisms. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13286–13291. [[CrossRef](#)]
40. Day, R.; Daggett, V. Ensemble versus single-molecule protein unfolding. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13445–13450. [[CrossRef](#)]
41. Daggett, V.; Li, A.; Itzhaki, L.S.; Otzen, D.; Fersht, A.R. Structure of the Transition State for Folding of a Protein Derived from Experiment and Simulation. *J. Mol. Biol.* **1996**, *257*, 430–440. [[CrossRef](#)]
42. Hartigan, J. *Clustering Algorithms*; John Wiley and Sons: New York, NY, USA, 1975.
43. Muff, S.; Caflisch, A. Identification of the protein folding transition state from molecular dynamics trajectories. *J. Chem. Phys.* **2009**, *130*, 125104. [[CrossRef](#)] [[PubMed](#)]

44. DeMarco, M.L.; Alonso, D.O.; Daggett, V. Diffusing and Colliding: The Atomic Level Folding/Unfolding Pathway of a Small Helical Protein. *J. Mol. Biol.* **2004**, *341*, 1109–1124. [[CrossRef](#)] [[PubMed](#)]
45. Klimov, D.; Thirumalai, D. Progressing from folding trajectories to transition state ensemble in proteins. *Chem. Phys.* **2004**, *307*, 251–258. [[CrossRef](#)]
46. Das, P.; Moll, M.; Stamati, H.; Kaviraki, L.E.; Clementi, C. Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 9885–9890. [[CrossRef](#)]
47. Rohrdanz, M.A.; Zheng, W.; Maggioni, M.; Clementi, C. Determination of reaction coordinates via locally scaled diffusion map. *J. Chem. Phys.* **2011**, *134*, 124116. [[CrossRef](#)] [[PubMed](#)]
48. Ceriotti, M.; Tribello, G.A.; Parrinello, M. Simplifying the representation of complex free-energy landscapes using sketch-map. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 13023–13028. [[CrossRef](#)]
49. Krivov, S.V.; Karplus, M. Hidden complexity of free energy surfaces for peptide (protein) folding. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 14766–14770. [[CrossRef](#)]
50. Caflisch, A. Network and graph analyses of folding free energy surfaces. *Curr. Opin. Struct. Biol.* **2006**, *16*, 71–78. [[CrossRef](#)]
51. 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. [[CrossRef](#)]
52. Du, R.; Pande, V.S.; Grosberg, A.Y.; Tanaka, T.; Shakhnovich, E.S. On the transition coordinate for protein folding. *J. Chem. Phys.* **1998**, *108*, 334–350. [[CrossRef](#)]
53. Hummer, G. From transition paths to transition states and rate coefficients. *J. Chem. Phys.* **2004**, *120*, 516–523. [[CrossRef](#)] [[PubMed](#)]
54. Best, R.B.; Hummer, G. Reaction coordinates and rates from transition paths. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6732–6737. [[CrossRef](#)] [[PubMed](#)]
55. Becker, O.M.; Karplus, M. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *J. Chem. Phys.* **1997**, *106*, 1495–1517. [[CrossRef](#)]
56. Toofanny, R.D.; Jonsson, A.L.; Daggett, V. A Comprehensive Multidimensional-Embedded, One-Dimensional Reaction Coordinate for Protein Unfolding/Folding. *Biophys. J.* **2010**, *98*, 2671–2681. [[CrossRef](#)]
57. Case, D.A.; Darden, T.; Cheatham, T.E., III; Simmerling, C.; Wang, J.; Duke, R.E.; Luo, R.; Merz, K.M.; Pearlman, D.A.; Crowley, M.; et al. *AMBER 9*; The University of California: San Francisco, CA, USA, 2006.
58. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinform.* **2006**, *65*, 712–725. [[CrossRef](#)]
59. Jorgensen, W.L. Revised TIPS for simulations of liquid water and aqueous solutions. *J. Chem. Phys.* **1982**, *77*, 4156–4163. [[CrossRef](#)]
60. Joung, I.S.; Cheatham, T.E., III. Determination of Alkali and Halide Monovalent Ion Parameters for use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041. [[CrossRef](#)]
61. Amadei, A.; Linssen, A.B.M.; Berendsen, H.J.C. Essential dynamics of proteins. *Proteins* **1993**, *17*, 412–425. [[CrossRef](#)]
62. Daidone, I.; Amadei, A.; Di Nola, A. Thermodynamic and kinetic characterization of a beta-hairpin peptide in solution: An extended phase space sampling by molecular dynamics simulations in explicit water. *Proteins* **2005**, *59*, 510–518. [[CrossRef](#)]
63. Prada-Gracia, D.; Gomez-Gardenes, J.; Echenique-Robba, P.; Faló, F. Exploring the Free Energy Landscape: From Dynamics to Networks and Back. *PLOS Comput. Biol.* **2009**, *5*, e1000415. [[CrossRef](#)]
64. Bahattacharyya, M.; Ghosh, A.; Hansia, P.; Visveshwara, S. Allostery and conformational free energy changes in human tryptophanyl-t-RNA synthetase from Essential Dynamics and Structure Networks. *Proteins* **2009**, *78*, 506–517. [[CrossRef](#)] [[PubMed](#)]
65. Pande, V.S.; Rokhsar, D.S. Is the molten globule a third phase of proteins? *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 1490–1494. [[CrossRef](#)] [[PubMed](#)]
66. Das, P.; Matysiak, S.; Clementi, C. Balancing energy and entropy: A minimalist model for the characterization of protein folding landscapes. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10141–10146. [[CrossRef](#)] [[PubMed](#)]
67. Fersht, A.R. Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10869–10873. [[CrossRef](#)] [[PubMed](#)]
68. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]
69. Ogushi, M.; Wada, A. ‘Molten-globule state’: A compact form of globular proteins with mobile side-chains. *FEBS Lett.* **1983**, *164*, 21–24. [[CrossRef](#)]
70. Pande, V.S.; Rokhsar, D.S. Folding pathway of a lattice model for proteins. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 1273–1278. [[CrossRef](#)]
71. Day, R.; Daggett, V. All-Atom Simulations of Protein Folding and Unfolding. *Adv. Protein. Chem.* **2003**, *66*, 373–403. [[CrossRef](#)]
72. Mayor, U.; Guydosh, N.R.; Johnson, C.M.; Grossmann, J.G.; Sato, S.; Jas, G.S.; Freund, S.M.V.; Alonso, D.O.V.; Daggett, V.; Fersht, A.R. The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **2003**, *421*, 863–867. [[CrossRef](#)]
73. Seeber, M.; Felling, A.; Raimondi, F.; Muff, S.; Friedman, R.; Rao, F.; Caflisch, A.; Fanelli, F. Wordom: A user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J. Comput. Chem.* **2010**, *32*, 1183–1194. [[CrossRef](#)]
74. Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. Wordom: A program for efficient analysis of molecular dynamics simulations. *Bioinformatics* **2007**, *23*, 2625–2627. [[CrossRef](#)] [[PubMed](#)]
75. Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Orozco, M. A Consensus View of Protein Dynamics. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 796–801. [[CrossRef](#)] [[PubMed](#)]