

# Analysis of Protein Folding Simulation with Moving Root Mean Square Deviation

Yutaka Maruyama,\* Ryo Igarashi, Yoshitaka Ushiku, and Ayori Mitsutake



Cite This: *J. Chem. Inf. Model.* 2023, 63, 1529–1541



Read Online

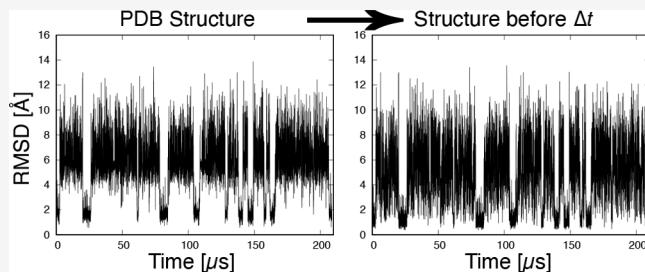
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** We apply moving root-mean-square deviation (mRMSD), which does not require a reference structure, as a method for analyzing protein dynamics. This method can be used to calculate the root-mean-square deviation (RMSD) of structure between two specified time points and to analyze protein dynamics behavior through time series analysis. We applied this method to the Trp-cage trajectory calculated by the Anton supercomputer and found that it shows regions of stable states as well as the conventional RMSD. In addition, we extracted a characteristic structure in which the side chains of Asp1 and Arg16 form hydrogen bonds near the most stable structure of the Trp-cage. We also determined that  $\geq 20$  ns is an appropriate time interval to investigate protein dynamics using mRMSD. Applying this method to NuG2 protein, we found that mRMSD can be used to detect regions of metastable states in addition to the stable state. This method can be applied to molecular dynamics simulations of proteins whose stable structures are unknown.



## INTRODUCTION

Molecular dynamics (MD) simulations are widely used to investigate phenomena in biomolecules such as proteins. This method is particularly suitable for investigating detailed motions at the atomic level on femto- or picosecond time scales, which are difficult to observe experimentally. In addition, application of a graphics processing unit (GPU) to scientific calculations has accelerated many MD programs, such as Amber,<sup>1,2</sup> NAMD,<sup>3,4</sup> OpenMM,<sup>5</sup> gromacs,<sup>6,7</sup> and ACEMD.<sup>8</sup> Using these programs, it has become possible to perform MD simulations of the microsecond order. With the use of special purpose supercomputers for MD simulation such as MD-GRAPE<sup>9,10</sup> and Anton,<sup>11–13</sup> it is possible to run MD simulations on the order of milliseconds. This has made it possible to perform protein folding processes in conventional MD simulations.<sup>14</sup>

Long-time conventional MD simulations of protein systems are especially important when using time-dependent analysis methods, e.g., relaxation mode analysis (RMA),<sup>15–24</sup> dynamic component analysis (DCA),<sup>25,26</sup> and time-lagged or time-structure based independent component analysis (tICA).<sup>27–31</sup> RMA was developed to investigate the dynamic properties of random spin systems<sup>32</sup> and homopolymer systems for Monte Carlo<sup>33</sup> and molecular dynamics.<sup>34</sup> It has also been applied to biomolecular systems<sup>15–24</sup> and can be used to extract slow motion that cannot be extracted by the principal component analysis because it uses time information.<sup>15</sup> In free-energy topography using RMA, the minima tend to be aligned along the axis, and transitions between stable and metastable states have been well described.<sup>16</sup> The differences among RMA,

DCA, and tICA have been previously explained in refs 20, 21, and 23. Generating time series data to use these methods requires long-time conventional MD simulations.

In conventional MD simulation, we use a time series of root-mean-square deviation (RMSD) for heavy or  $C_\alpha$  atom positions to investigate protein behavior.<sup>35,36</sup> In addition, plotting various energies against RMSD is a useful approach for investigating the structural stability of a protein.<sup>37–43</sup> However, the reference structure of RMSD, which, in such a case, is a native structure, must be known in advance. One possible approach would be to use an extended structure as a reference. However, in this case, the value indicating the native structure is unclear.

In the previous study, we plotted the various energies of four proteins as a function of the end-to-end distance between the N- and C-termini.<sup>44</sup> The distance does not require a reference structure, and it is experimentally measurable using techniques such as fluorescent resonance energy transfer (FRET). However, the correlations between RMSD and end-to-end distance depend on individual proteins. In other words, end-to-end distance is not an appropriate indicator for investigating stable or metastable states of proteins.

Received: November 17, 2022

Published: February 23, 2023



In this work, we attempted to analyze the MD trajectory using time series RMSD without a reference structure. We calculated RMSD using the structure before a specified time rather than a fixed reference structure. Therefore, it is useful for MD simulations of proteins whose native structures have not yet been experimentally determined. The method of calculating such RMSD values is a known technique for checking equilibration in short-time MD simulation. This is also a simplified version of moving RMSD (mRMSD) proposed by Harada et al. for sampling efficiency of the nontargeted parallel cascade selection molecular dynamics.<sup>45–47</sup> We used the time series of mRMSD to analyze Anton's long-time simulations of the Trp-cage<sup>48,49</sup> and NuG2<sup>50</sup> proteins.<sup>14</sup>

## METHODS

**Moving RMSD.** RMSD is a commonly used measure for comparing protein conformations.<sup>35,36</sup> We considered two different structures, v and w. These structures have  $n$  three-dimensional coordinates; RMSD was defined by

$$\text{RMSD}(v, w) = \sqrt{\frac{1}{n} \sum_{i=1}^n \|v_i - w_i\|^2} \quad (1)$$

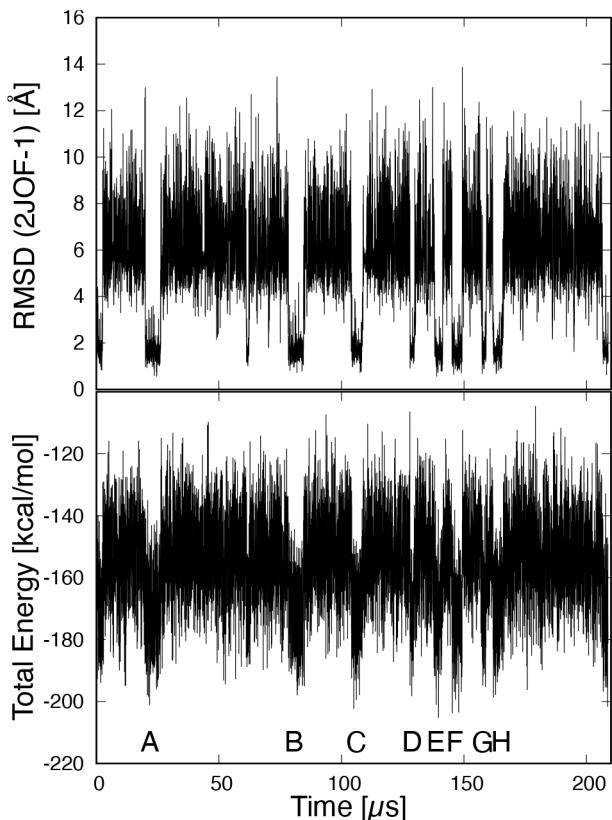
The two protein conformations were prealigned to minimize the distance between them. There are several ways to calculate RMSD: all-atom RMSD, heavy-atom RMSD, and  $C_\alpha$ -atom RMSD. Here we chose  $C_\alpha$ -atom RMSD to investigate protein structural stability.

Conventionally, RMSD is calculated by fixing a reference. In the case of protein folding simulations, native structures obtained experimentally are used as references. Now consider a simulation of a protein whose the native structure is unknown. In the absence of a suitable reference structure, we calculate the RMSD as follows. v is a structure at time  $t$ , and w is one at time  $t - \Delta t$ , where  $\Delta t$  is the time interval for RMSD calculation. This description is a simplified version of the moving RMSD (mRMSD).<sup>45,46</sup> In the original mRMSD, a reference is an average of several structures before and after the time  $t$  of interest. (See Figure 1 in ref 45.) In this work, a single structure before a certain time is used as a reference, and the RMSD between  $v(t)$  and  $w(t - \Delta t)$  is referred to as the mRMSD. We calculated the time series of  $\text{RMSD}(v(i\Delta s), w(i\Delta s - \Delta t))$  with the sampling time interval  $\Delta s$  at time step  $i$  to investigate the protein dynamics behavior. The advantage of this measure is that it does not require a reference structure. We mention that mRMSD is not suitable as an axis for energy plots because the value of mRMSD varies with time interval  $\Delta t$ .

**Total Energy.** We introduce the total energy,  $E_{\text{tot}}$ , which is the sum of the conformational energy,  $E_{\text{conf}}$ , and the solvation free energy (SFE),  $E_{\text{sol}}$ :

$$E_{\text{tot}} = E_{\text{conf}} + E_{\text{sol}} \quad (2)$$

The conformational energy is determined from the protein structure and depends on the force field used in the calculation. This energy can be calculated using molecular mechanics (MM) or MD programs. The SFE is defined as the amount of change in chemical energy when one solute molecule is moved from one position in the gas phase to another in the solvent. The SFE can be calculated using the three-dimensional reference interaction site model (3D-RISM)



**Figure 1.** (top) Time series of the root-mean-square deviation (RMSD) with respect to the native structure of Trp-cage (PDB 2JOF, model 1). (bottom) Time series of the total energy, which is the sum of the conformational energy and the solvation free energy, of Trp-cage. A–H indicate the regions of the stable state.

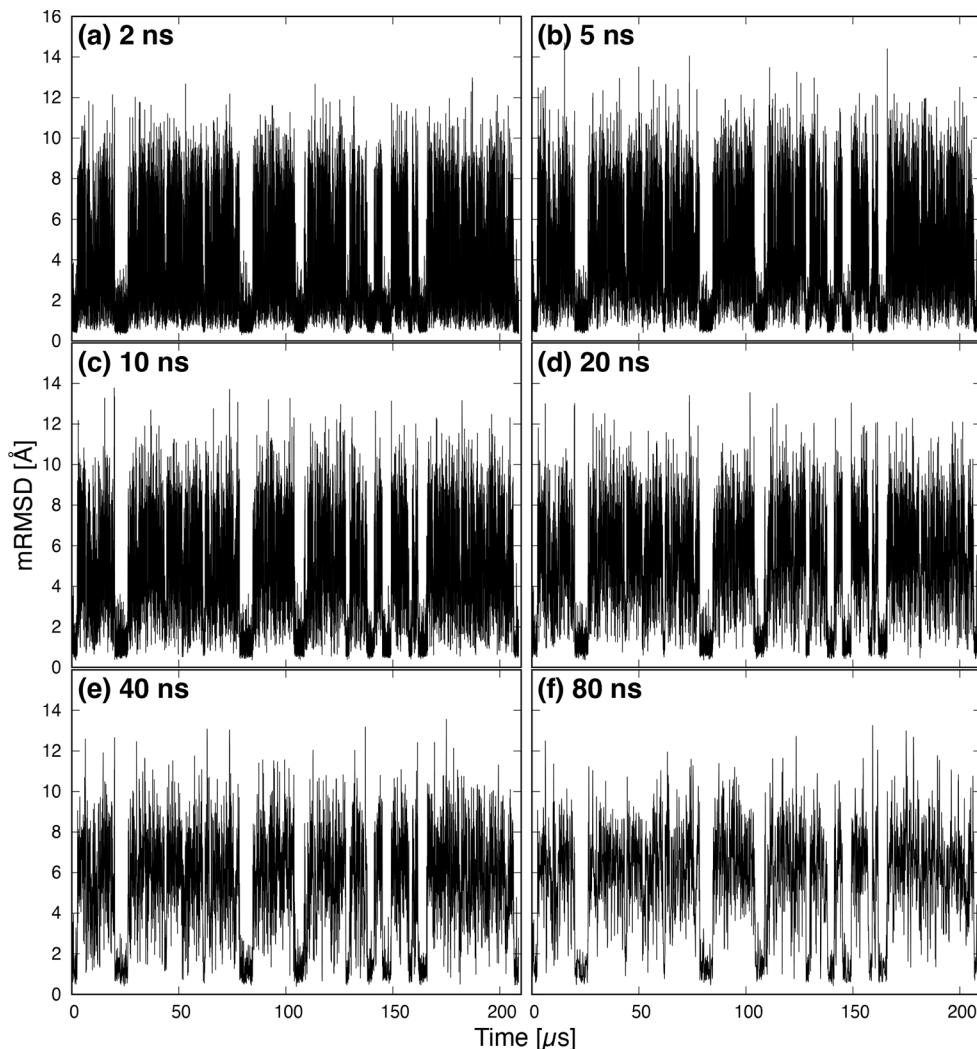
theory,<sup>51,52</sup> the energy representation (ER) theory,<sup>53–55</sup> and the grid inhomogeneous solvation theory (GIST).<sup>56,57</sup> This total energy is a reasonable indicator of protein structural stability.<sup>41–44,58</sup>

## COMPUTATIONAL DETAILS

We used the following two proteins: the Trp-cage and the triple mutant of the redesigned protein G variant NuG2 (an  $\alpha + \beta$  protein). The details of Anton's MD simulations for these proteins can be found in the Supporting Information in ref 14.

We used GROMACS<sup>59</sup> for calculating the conformational energy with CHARMM22\* force field.<sup>60–62</sup> The SFE of the proteins was calculated using the 3D-RISM theory with the reference-modified density functional theory.<sup>63–65</sup> The number density of water was  $0.033329 \text{ \AA}^{-3}$  and the optimal hard sphere diameter was  $2.88 \text{ \AA}$  for the thermodynamic states at  $298.15 \text{ K}$ . We performed the SFE calculation using 3D-RISM theory with the original code for GPU.<sup>66,67</sup> An SFE calculation of a NuG2 protein takes 10 s on NVIDIA V100 GPU ( $256^3$  grids, water solvent).

Structures were extracted from Anton's trajectory in the calculations and used every 2 ns (every 10 samples). For NuG2 protein, we used the fourth trajectory (NuG2-3 Trajectory). The total number of sampling structures was 104 400 and 86 377 for Trp-cage and NuG2, respectively. For the  $C_\alpha$ -RMSD calculations with respect to the native structure, the references were taken from 2JOF.pdb (model 1) and 1MIO.pdb for Trp-cage and NuG2, respectively. We used 2, 5, 10, 20, 40, and 80 ns time intervals for the mRMSD



**Figure 2.** Time series of the moving root-mean-square deviation (mRMSD) of Trp-cage. Panels a–f have time intervals  $\Delta t$  of 2, 5, 10, 20, and 80 ns, respectively.

calculations. We set  $\Delta s = \Delta t$  in the mRMSD calculations. The end-to-end distance is defined as the distance between  $C_\alpha$  atoms of the N- and C-termini. Energy calculations for Trp-cage were performed every 20 ns. In addition, calculations were performed every 2 ns during the 20–26.2  $\mu$ s period.

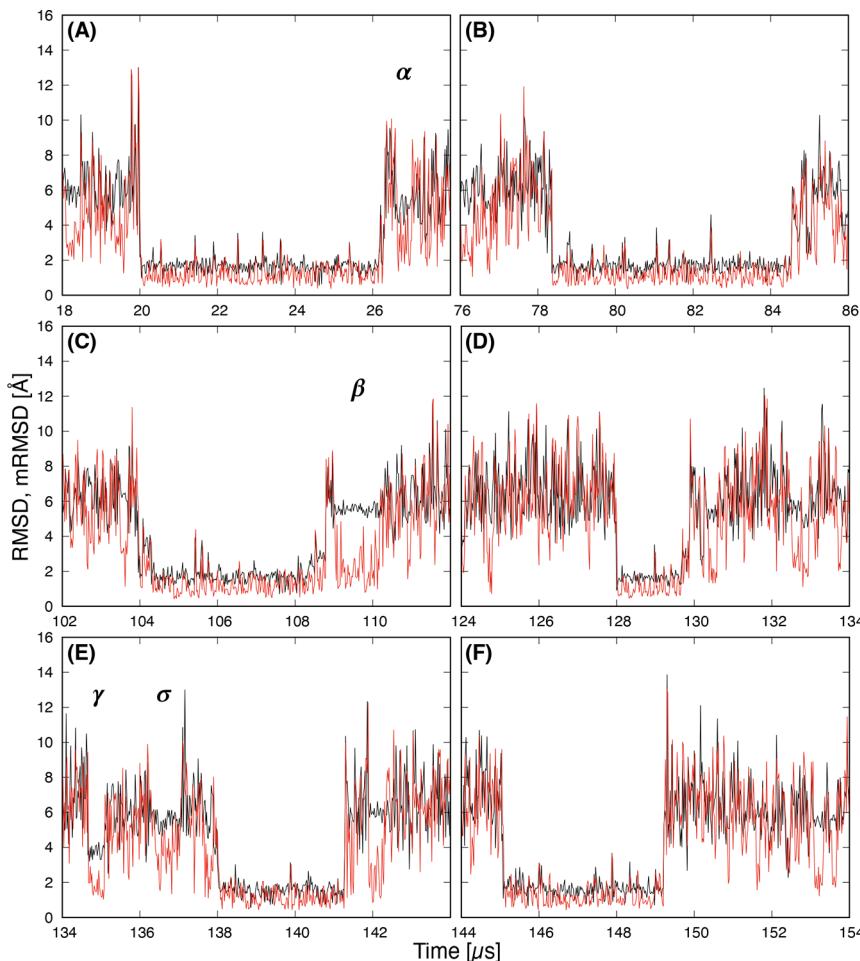
## RESULTS AND DISCUSSION

**Trp-cage.** Trp-cage is an artificial protein consisting of 20 residues that forms a tryptophan-centered hydrophobic core. It is used as a test system for various new methods.<sup>68–75</sup> In addition, Trp-cage is often used to study the folding process of proteins.<sup>76–80</sup> We used the N1D/L2A/I4A triple mutant of Trp-cage<sup>49</sup> in this study.

The time series of the conventional RMSD of  $C_\alpha$  atom from a native structure and the total energy are shown in Figure 1. Here, the native structure is the first coordinate of 2JOF.pdb. The stable state has an RMSD value of  $\sim 2$  Å, with  $\sim 10$  distinct regions. The total energy in these regions is lower than in regions with larger RMSD values. In other words, the total energy is also an indicator of the protein's stable state. (The most stable structures in the stable state regions are shown in Figure S1 of the Supporting Information.)

Figure 2 shows the mRMSD time series of the  $C_\alpha$  atom with time intervals as follows: (a) 2, (b) 5, (c) 10, (d) 20, (e) 40, and (f) 80 ns. At all time intervals, we were able to identify the regions of the stable state indicated by the conventional RMSD. In the stable region, the mRMSD shows values of  $\sim 1$  Å while the conventional RMSD shows values of  $\sim 2$  Å. However, in (a) 2 and (b) 5 ns, where time intervals were short, the mRMSD values were less than 2 Å even in the nonstable state. This is because the conformational change of the protein is smaller when the time interval is short. As the time interval increased to (c) 10 or (d) 20 ns, the mRMSD values in the nonstable state region increased while the values in the stable state did not change. As the time interval is further lengthened, the separation between stable and nonstable states becomes clearer (Figure 2e,f).

Next, we show the comparison of RMSD (black) and the mRMSD with a 20 ns time interval (red) near the stable state regions in Figure 3. Labels A–F correspond to A–F of the stable state regions shown in Figure 1. The mRMSD values are low and stable in all stable state regions indicated by RMSD. This suggests that the mRMSD functions properly as an indicator of the stable state. The regions indicated by  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\sigma$  in Figure 3 are metastable state regions. In these regions, discrepancies between RMSD and mRMSD values were

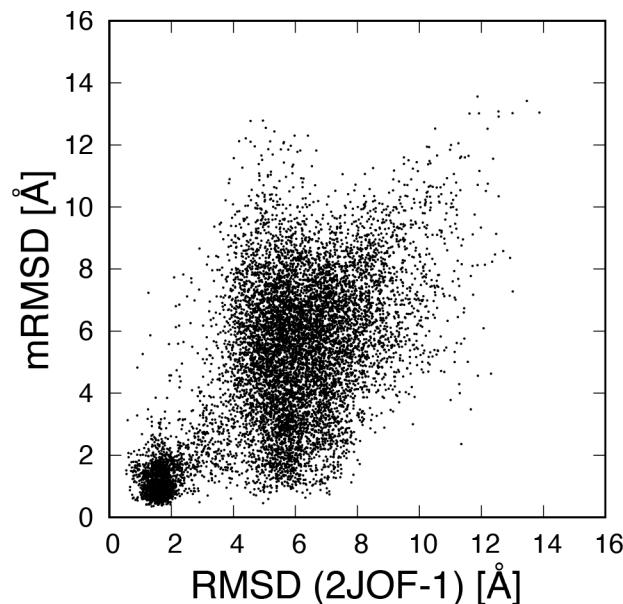


**Figure 3.** Time series of the RMSD with respect to the native structure (black) and the mRMSD with 20 ns time interval (red). Labels A–F correspond to the region of the stable state in Figure 1.  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\sigma$  indicate the regions of metastable states.

observed. In addition, the fluctuations in values were small for RMSD, but large for mRMSD. This can be explained as follows. In the conventional RMSD, which uses a reference, the changes of the RMSD value become smaller when the difference between the structure and the reference is large. On the other hand, the mRMSD is more sensitive to structural changes. In the metastable state, the value of mRMSD fluctuates significantly because the structural fluctuation is larger than in the stable state.

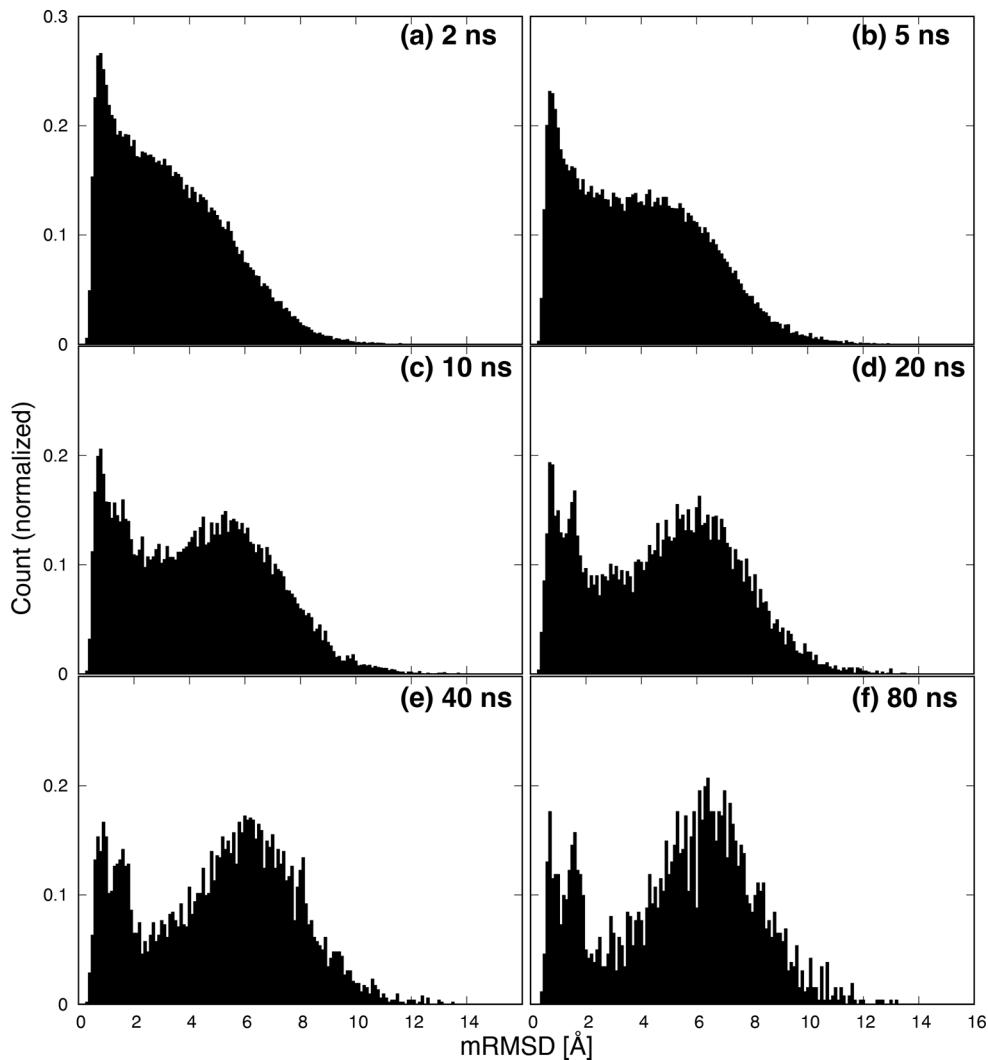
Figure 4 shows the correlation between the RMSD and the mRMSD with a 20 ns time interval. A circular region centered at point (1.5, 1.5) is observed. This indicates that there was a clear correlation in the stable state region. On the other hand, around RMSD 6 Å, the distribution extends vertically. Around RMSD 6 Å, there were various structures that deviated from the stable state. Since these structures undergo various fluctuations and structural changes, the mRMSD values fluctuate. The pattern of the correlation distribution depends on the time interval of mRMSD, but the correlation in the stable state remains independent of the time interval.

Figure 5 shows histograms of mRMSD values. The counts of histograms are normalized. At (a) 2 ns, there is a strong peak near 1 Å that falls gently to 10 Å. However, as the time interval increases from (b) 5 ns to (c) 10 ns, a new peak becomes visible around 6 Å. This indicates that when the mRMSD time interval is short, the conformational change of the protein is too small to detect the movement of several tens of



**Figure 4.** Correlation between the RMSD with respect to the native structure and the mRMSD with 20 ns time interval.

nanoseconds. Furthermore, at (d) 20 ns, the first peak splits into two peaks at 0.7 and 1.6 Å. We will discuss the peak at 1.6 Å later. Further increasing the time interval does not



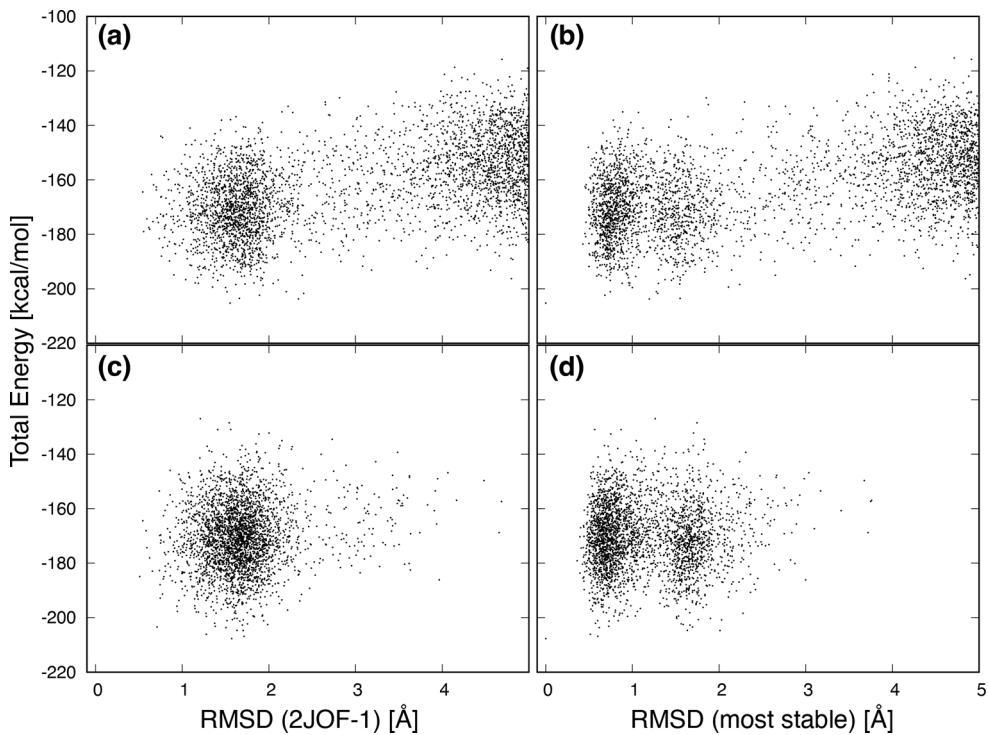
**Figure 5.** Histogram of mRMSD values. Panels a–f have time intervals  $\Delta t$  of 2, 5, 10, 20, 40, and 80 ns, respectively. Counts are normalized.

significantly change the position of the peak or the shape of the histogram (Figure 5e,f). Therefore, we consider that a mRMSD time interval of  $\geq 20$  ns is appropriate for analysis of protein dynamics.

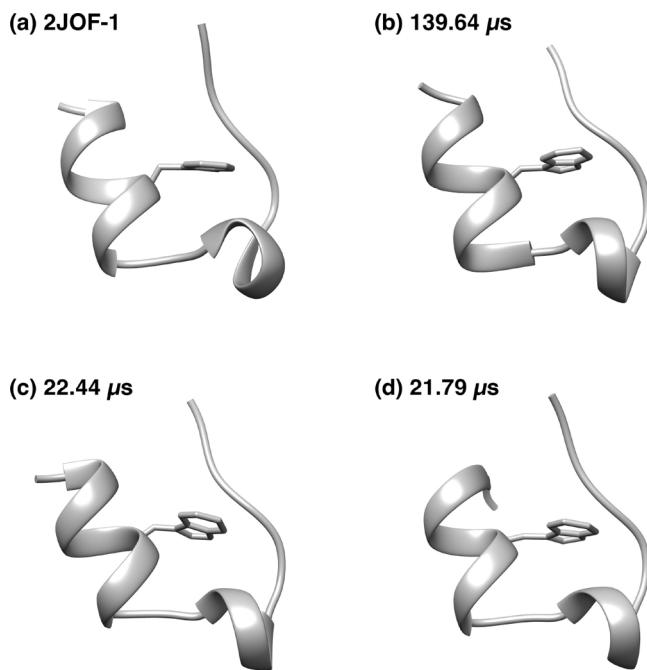
We show the values of the total energy of Trp-cage as functions of  $C_{\alpha}$ -RMSD in Figure 6. Panels a and b were sampled every 20 ns from the entire trajectory, while panels c and d were sampled every 2 ns in region A (20–26.2  $\mu$ s) of Figure 1. In Figure 6a,c, the first structure of 2JOF.pdb was used as a reference for  $C_{\alpha}$ -RMSD, and in Figure 6b,d, the structure with the lowest total energy in each sample was used as a reference. Panels c and d assume that the MD calculation was able to sample sufficient stable structures in region A. Comparing panels a and c, the lowest total energy in Figure 6a is  $-205.3$  kcal/mol, while it is  $-206.9$  kcal/mol in Figure 6c. In addition, there are seven structures below  $-200$  kcal/mol in Figure 6a, while there are 19 in Figure 6c. Naturally, it is possible to obtain lower energy structures by intensively sampling stable state regions. In other words, if the goal is to find a stable structure of a protein, stopping the MD simulation at the end of region A and sampling will achieve this. Next, comparing panel a with panel b and panel c with panel d, the separated distributions are observed around 0.8 and 1.6  $\text{\AA}$  in Figure 6b,d. These separations of distributions correspond to

the separations of the first peak observed in Figure 5d,e,f. This suggests that Trp-cage has a structure with a similar shape to the stable structure. On the other hand, it is difficult to distinguish these two in Figure 6a,c. Thus, depending on the structure used as a reference, the value of RMSD will change, and the information obtained from it will vary.

Figure 7 shows Trp-cage backbone structures with the side chain of the sixth amino acid tryptophan. Panel a is the first structure of 2JOF.pdb, panels b and c are the most stable structures in Figure 6a,c, respectively, and panel d is the most stable structure, with an RMSD value  $\sim 1.6$   $\text{\AA}$  in Figure 6d. Structures a, b, and c are very similar in shape, but structure a differs in the shape of the central helix. Except for this, structures a and c are very similar in terms of the shape of the first helix. We can say that structure c is an intermediate structure between structures a and b. Structure d is close to the most stable structure, but the N-terminus is closer to the C-terminus. The N-terminal Asp1 side chain forms a hydrogen bond with the side chain of Arg16. It is well-known that the side chain of Arg16 forms a hydrogen bond with the side chain of Asp9 in the stable structure of Trp-cage.<sup>49,81–87</sup> In Figure 7d, Arg16 forms hydrogen bonds with the side chains of both Asp1 and Asp9.

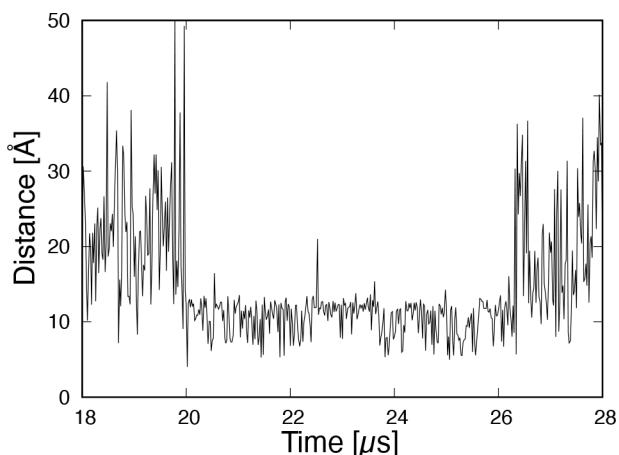


**Figure 6.** Total energy as a function of the RMSD with respect to (a, c) the native structure (PDB 2JOF, model 1) and (b, d) the most stable structure in the simulation. Panels a and b were sampled at 20 ns intervals from the entire region, while panels c and d were sampled at 2 ns intervals from the stable state region A (20.0–26.2  $\mu$ s).



**Figure 7.** Backbone structures with the side chain of the sixth amino acid tryptophan. (a) PDB 2JOF, model 1. (b) The most stable structure when sampling the entire region at 20 ns intervals. (c) The most stable structure when sampling the stable state region A (20.0–26.2  $\mu$ s) at 2 ns intervals. (d) The most stable structure with RMSD value  $\sim$ 1.6  $\text{\AA}$  in Figure 6d.

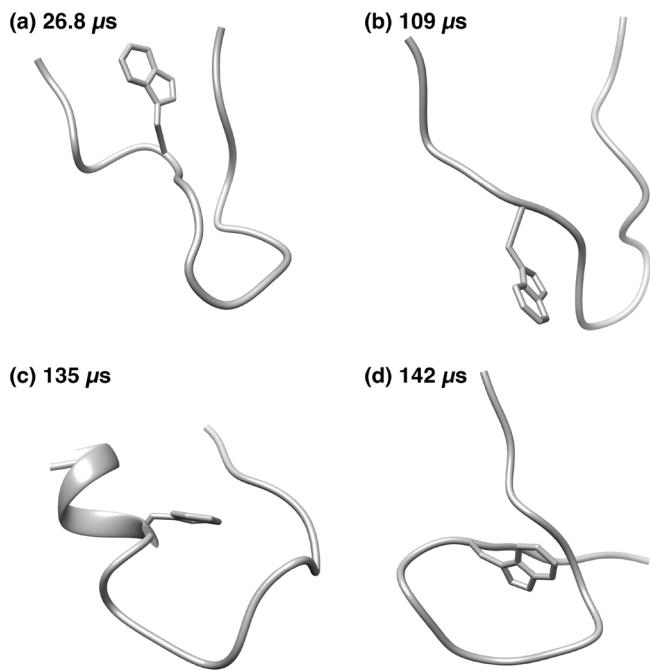
We calculated the end-to-end distance in region A (Figure 8). Between 20 and 26.2  $\mu$ s, the end-to-end distance moves back and forth between values of 8 and 12  $\text{\AA}$ . The number of oscillations is 107, giving an average period of  $\sim$ 58 ns. It was



**Figure 8.** Time series of the end-to-end distance between the N- and C-termini in the stable state region A (20.0–26.2  $\mu$ s).

found that a mRMSD time interval of  $\geq$ 20 ns was required to detect this mode of hydrogen bond formation and breakage. Since this phenomenon lasts for  $\geq$ 6  $\mu$ s, it is not expected to have a significant effect on the structural stability of the Trp-cage.

At the end of this subsection, we show the structures of the metastable regions in Figure 9. Panels a, b, c, and d are the structures in the metastable regions of  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\sigma$  in Figure 3, respectively. These structures are not exactly in a metastable state because they do not form secondary structures. They are similar to those analyzed for clustering in previous studies.<sup>88–90</sup> These structures form loops and are further maintained by contact between hydrophobic side chains. By examining the time series of mRMSD, we can find states that are different from the native structure but less fragile.

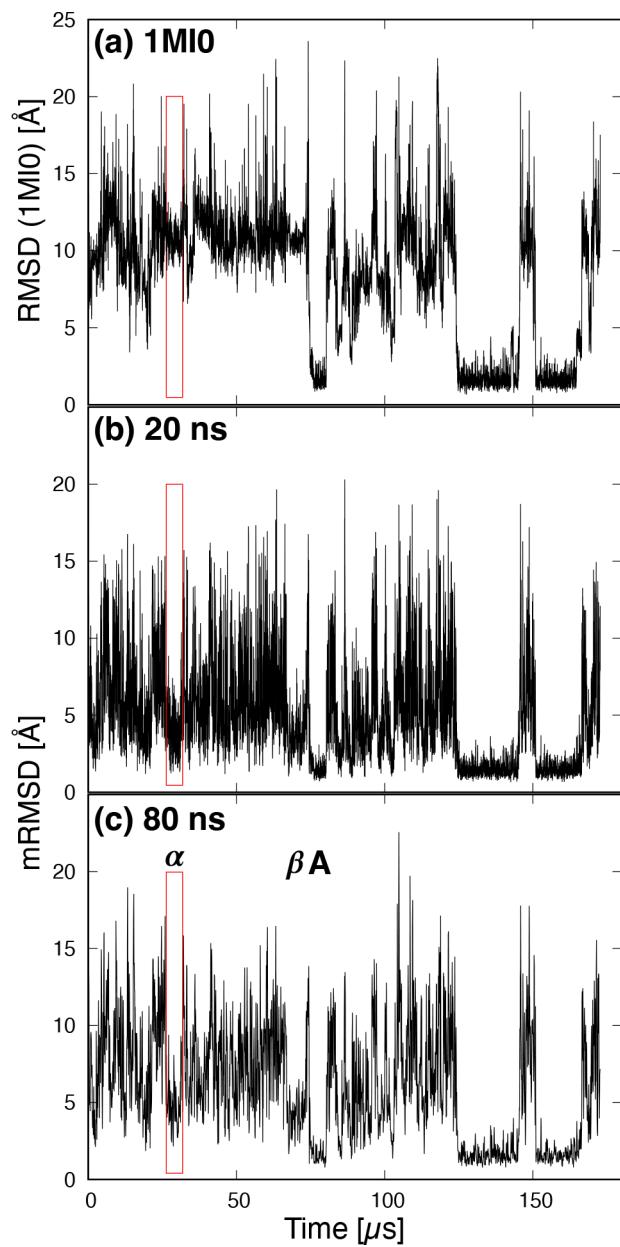


**Figure 9.** Backbone structures with the side chain of the sixth amino acid tryptophan for the metastable regions (a)  $\alpha$ , (b)  $\beta$ , (c)  $\gamma$ , and (d)  $\delta$  in Figure 3.

**NuG2.** NuG2, which contains 56 amino acids, is the N37A/A46D/D47A triple mutant of protein G.<sup>50</sup> NuG2 protein consists of a single  $\alpha$ -helix packed against two  $\beta$  sheets. In other words, the first  $\beta$  ( $\beta_1$ ) strand, the second  $\beta$  ( $\beta_2$ ) strand, the  $\alpha$ -helix, the third  $\beta$  ( $\beta_3$ ) strand, and the fourth  $\beta$  ( $\beta_4$ ) strand are arranged in order from the N-terminus to the C-terminus. NuG2 is redesigned to switch folding pathways and to fold through a similar pathway to protein L in the early folding process. In the early folding process, the first and second  $\beta$  strands formed an N-terminal antiparallel  $\beta$ -sheet ( $\beta$ -hairpin). NuG2 has been analyzed by several groups using Anton's MD trajectories.<sup>41,91–98</sup> In addition, we have suggested the existence of many metastable states in a previous study using RMA.<sup>22</sup>

Figure 10 shows the time series of (a) the RMSD of the native structure (PDB 1M10) and the mRMSD with time intervals of (b) 20 ns and (c) 80 ns. (See Supporting Information for the time series of mRMSD with time intervals 2, 5, 10, and 40 ns.) The three stable state regions with lower RMSD  $\sim$ 2 Å are shown in Figure 10a. The first one is designated as stable state region A in Figure 10c. As in the case of Trp-cage, the stable state regions can be seen in the same location in Figure 10b,c. In Figure 10c,  $\alpha$  and  $\beta$  indicate the metastable regions, which have slower structural changes. In both RMSD and mRMSD, the beginning and end of the region  $\beta$  can be clearly identified. On the other hand, it is difficult to determine the starting point of region  $\alpha$  from the time series of RMSD. This shows the superiority of mRMSD, which can be used to calculate the difference from the structure before a certain time interval. The mRMSD is a very useful value because it can extract regions of the stable structure and metastable structures in which structural changes are slow.

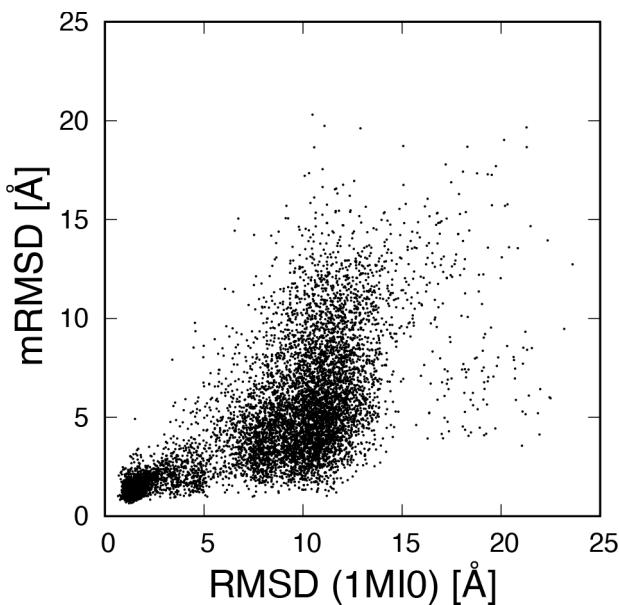
The correlation between the RMSD and the mRMSD with a 20 ns time interval is shown in Figure 11. Similar to the Trp-cage, there is a strong correlation near the region centered at (2, 2), which indicates the stable state. Compared with the



**Figure 10.** Time series of (a) the RMSD of NuG2 protein with respect to native structure (PDB 1M10) and the mRMSD with time interval (b) 20 ns and (c) 80 ns.  $\alpha$  and  $\beta$  indicate the metastable state regions, and A is the stable state region. The red rectangle indicates the extent of region  $\alpha$ .

Trp-cage correlation, longitudinally extending distribution is shifted around RMSD 10 Å. This is because the value of RMSD for unstable structures is protein dependent. Structures with mRMSD values smaller  $<5$  Å are considered to be metastable states because of small structural changes. In other words, we can obtain information on metastable structures using mRMSD even if the RMSD value is large.

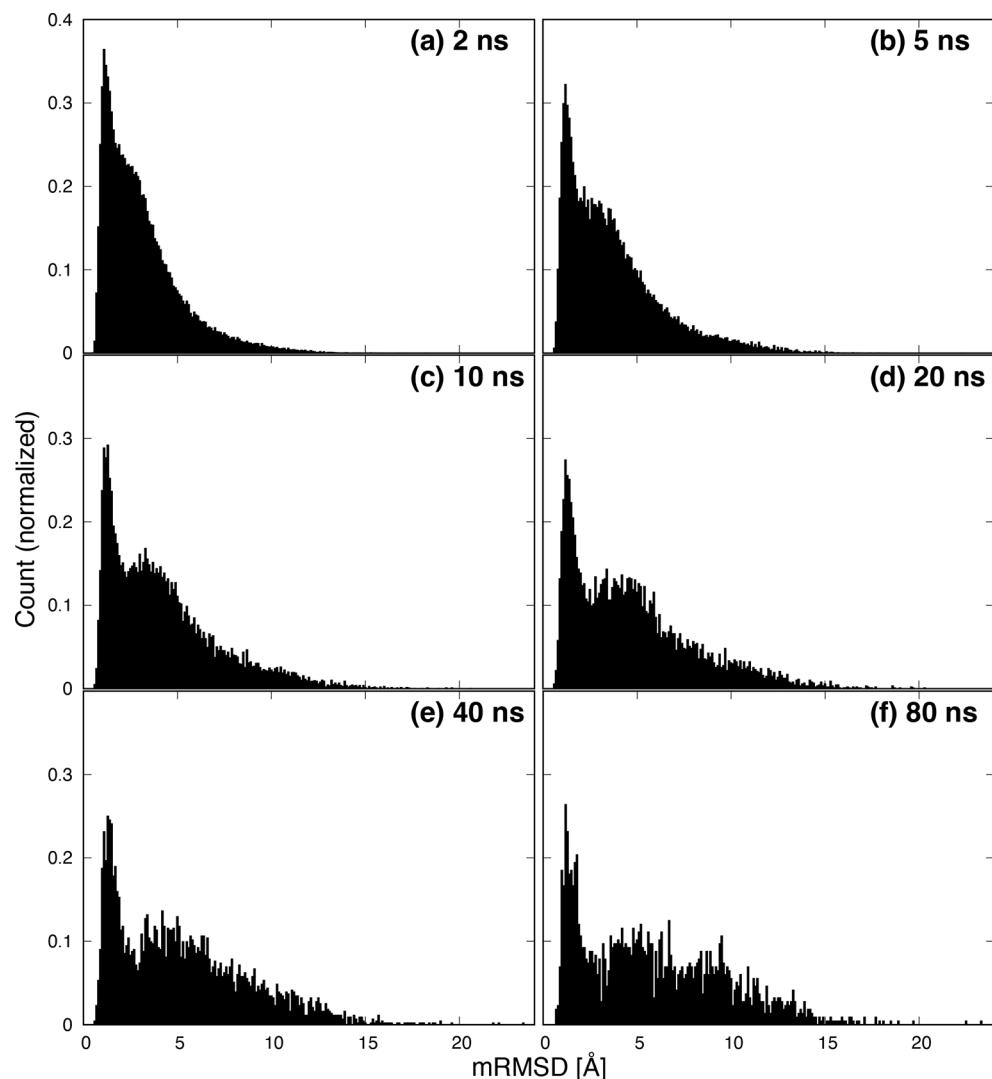
Figure 12 shows the histograms of mRMSD values for each time interval. No separation is observed in the peaks at (a) 2 ns or (b) 5 ns. At time intervals of  $\geq$ 10 ns, separation of the peaks is observed. The splitting of the first peak observed in Trp-cage is not observed at (d) 20 ns. The separation of the first peak is a Trp-cage specific phenomenon. This is because NuG2 is larger and thus mRMSD values are less sensitive to terminal



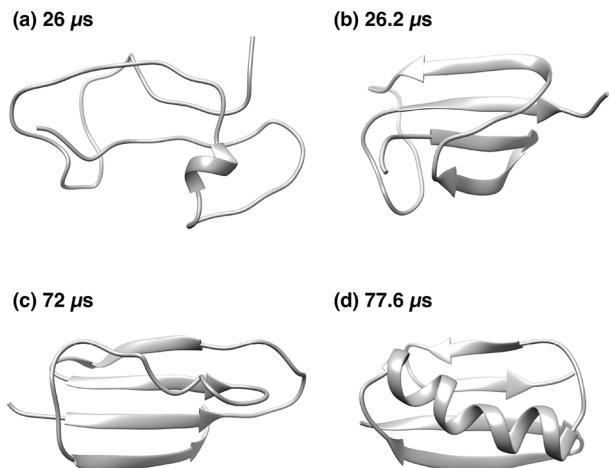
**Figure 11.** Correlation between the RMSD with respect to the native structure and the mRMSD with 20 ns time interval.

fluctuations. At (f) 80 ns, three peaks related to NuG2 dynamics were observed.

**Figure 13** shows the characteristic structures of (a, b) the metastable state region  $\alpha$ , (c) the metastable region  $\beta$ , and (d) the stable state region A. (a) Just before the start of region  $\alpha$ , NuG2 forms a random coil structure. Immediately after, a parallel  $\beta$ -sheet is formed at the N- and C-termini. (b) This parallel  $\beta$ -sheet is stable, and then antiparallel  $\beta$ -sheets are formed between  $\beta_1$  and  $\beta_2$  and between  $\beta_3$  and  $\beta_4$ . This structure resembles the native structure without the  $\alpha$ -helix. This structure fits into B1 in the ref 22 classification. However, there is no direct transition from this metastable structure to the native structure. At the end of region  $\alpha$ , the parallel  $\beta$ -sheet of  $\beta_1$  and  $\beta_2$  breaks down and NuG2 becomes a random coil structure. (c) In the  $\beta$  region, the structure is formed as shown in **Figure 13c**. This is classified as the R4 in ref 22. The structure which forms the parallel  $\beta$ -sheets between  $\beta_1$  and  $\beta_4$  strands and between  $\beta_2$  and  $\beta_3$  strands cannot be the expected structure. This structure also does not directly change to the native structure. Between regions  $\beta$  and A, the parallel  $\beta$ -sheets of  $\beta_1$  and  $\beta_4$  are broken and NuG2 protein takes on an extended structure. Between regions  $\beta$  and A, the end-to-end distance exceeds 80 Å, as observed in **Figure 14**. (d) Then,  $\beta_1$

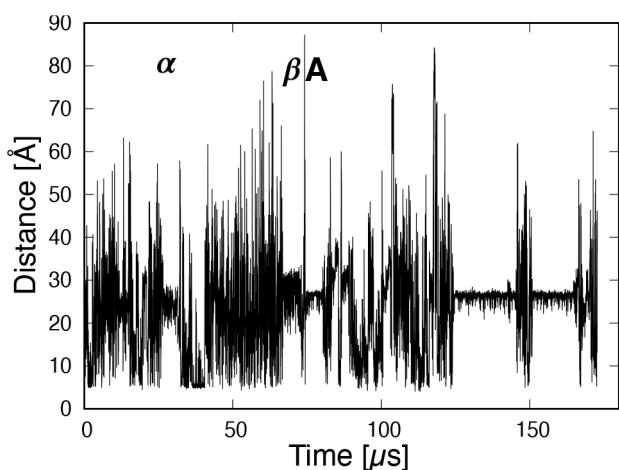


**Figure 12.** Histogram of mRMSD values. Panels a–f have time intervals  $\Delta t$  of 2, 5, 10, 20, and 80 ns, respectively. Counts are normalized.



**Figure 13.** Backbone structures of NuG2 protein. (a) Random coil structure at beginning of the metastable region  $\alpha$  in Figure 10c. (b) The metastable structure in the metastable region  $\alpha$ . (c) The metastable structure in the metastable region  $\beta$ . (d) The most stable structure in the stable region A.

and  $\beta_2$  strands form the antiparallel  $\beta$ -sheet and NuG2 folds rapidly to reach the native structure, as shown in Figure 13d.



**Figure 14.** Time series of the end-to-end distance between the N- and C-termini.  $\alpha$ ,  $\beta$  and A are the same as in Figure 10c.

The time series of the end-to-end distance between the N- and C-termini is shown in Figure 14.  $\alpha$ ,  $\beta$ , and A indicate the same regions as in Figure 10. Similar to RMSD and mRMSD, in addition to region A, regions of stable states starting from 125 and 155  $\mu$ s can be observed. However, the value of the end-to-end distance of the native structure and the width of the fluctuation of the end-to-end distance depends on the type of protein.<sup>44</sup> Therefore, it is not easy to search for stable states using only this measure. On the other hand, we can obtain information from the end-to-end distance in a different way from RMSD or mRMSD. For example, large values are observed at the end of the stable or metastable states. These indicate that the parallel  $\beta$ -sheet between  $\beta_1$  and  $\beta_4$  was broken. RMSD and mRMSD can only show that the structure changed significantly.

## CONCLUSIONS

We investigated protein folding processes by calculating the mRMSD, which does not require a reference structure. We have discussed here how the mRMSD can be incorporated into the stable and metastable states of proteins. In MD simulation trajectories of Trp-cage and NuG2 proteins, the time series of mRMSD can be used to correctly detect native structures. In addition, it is also able to detect the metastable states of both proteins. We also showed the difference in RMSD values when using each of the most stable structures within the MD simulation and the experimental structure as a reference. The misalignment between the reference structure and the most stable structure in the MD simulation affects the resolution of the energy analysis.

The mRMSD can be used to analyze MD simulations of proteins whose native structure is unknown. The behavior of the time series of the mRMSD can be examined to determine whether the simulation should be stopped. It is possible to determine with equal or higher accuracy than generating references with AlphaFold2,<sup>99</sup> RoseTTAFold,<sup>100</sup> etc. The flow of analyzing an MD simulation of a protein of unknown structure is as follows: 1. Perform the MD simulation, calculate mRMSD on the fly, and stop the simulation at the end of the stable state region after the stable state one appears several times. 2. Sample the structures of the stable state regions and determine the most stable structure by calculating the total energy or the frequency distribution. 3. Calculate conventional RMSD using the most stable structure as a reference, and analyze the various energies of the protein.

In general, it is difficult to extract native or metastable states from trajectories of MD simulations with large structural changes. The regions extracted by mRMSD calculation can be analyzed by other analytical methods to clarify the mechanisms of protein stability, dynamics, and folding processes. For example, the stability of the most stable and metastable states can be investigated by calculating total energy. The folding process can also be studied by analyzing the structures before and after the regions using clustering methods.

If a target is an intrinsically disordered protein (IDP) with partially flexible regions, this method may not work well. However, by splitting the structure into two or three parts, we can examine the behavior of RMSD of an IDP. An IDP undergoes transitions to more ordered states upon binding to its targets.<sup>101</sup> Calculating mRMSD is also useful to investigate such binding processes between an IDP and its diverse partners.

The appropriate time interval is  $\geq 20$  ns in the case of Trp-cage and NuG2. The appropriate time interval depends on the system. For larger proteins, the appropriate time interval for mRMSD calculations would be longer. If the sampling interval  $\Delta s$  and the mRMSD time interval  $\Delta t$  are the same, the number of samples is smaller. However, the sampling interval and mRMSD time interval need not be the same. We can increase the number of samples by adjusting the time interval of the number of samples and the mRMSD time interval separately. We intend to make such improvement in future work.

## ASSOCIATED CONTENT

### Data Availability Statement

MD trajectory data was provided by D. E. Shaw RESEARCH (e-mail: [inquiries@deshaw.com](mailto:inquiries@deshaw.com)). 3D-RISM input/output, mRMSD code, and parameter files obtained from the present

calculations can be downloaded from the following repository: <https://github.com/omron-sinicx/mRMSD-paper-data>. The code for mRMSD was written using MDAnalysis (ver. 2.3.0).<sup>102</sup> The 3D-RISM with RMDFT functional code for the single GPU system is available from the following repository: <https://github.com/drmaruyama/3D-RISM-CUDA/tree/rmdft>.

## Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.2c01444>.

The stable structures of Trp-cage and the time series of mRMSD of NuG2 ([PDF](#))

## AUTHOR INFORMATION

### Corresponding Author

Yutaka Maruyama – OMRON SINIC X Corporation, Tokyo 113-0033, Japan; Department of Physics, School of Science and Technology, Meiji University, Kawasaki-shi, Kanagawa 214-8571, Japan;  [orcid.org/0000-0003-4035-4885](http://orcid.org/0000-0003-4035-4885); Email: [yutaka.maruyama@sinicx.com](mailto:yutaka.maruyama@sinicx.com)

### Authors

Ryo Igarashi – OMRON SINIC X Corporation, Tokyo 113-0033, Japan

Yoshitaka Ushiku – OMRON SINIC X Corporation, Tokyo 113-0033, Japan

Ayori Mitsutake – Department of Physics, School of Science and Technology, Meiji University, Kawasaki-shi, Kanagawa 214-8571, Japan;  [orcid.org/0000-0002-4194-7255](http://orcid.org/0000-0002-4194-7255)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jcim.2c01444>

### Author Contributions

Y.M. and A.M. performed mRMSD and 3D-RISM computations. Y.M. designed this study. All analyses were performed under the discussion of all authors. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

### Funding

This work was supported by JST-Mirai Program, Grant Number JPMJMI19G1 and by JSPS KAKENHI Grant Number JP20H03230.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Numerical calculations were conducted in part using Cygnus at the Center for Computational Sciences, University of Tsukuba. Molecular graphics were designed using UCSF Chimera, which was developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco.<sup>103</sup>

## REFERENCES

- (1) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.
- (2) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.
- (3) Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K. Accelerating molecular modeling applications with graphics processors. *J. Comput. Chem.* **2007**, *28*, 2618–2640.
- (4) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singhary, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130.
- (5) Eastman, P.; Pande, V. In *GPU Computing Gems Jade ed.*; Mei, W., Hwu, W., Eds.; Applications of GPU Computing Series; Morgan Kaufmann: Boston, 2012; pp 399–407.
- (6) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- (7) Kutzner, C.; Páll, S.; Fechner, M.; Esztermann, A.; de Groot, B. L.; Grubmüller, H. Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. *J. Comput. Chem.* **2015**, *36*, 1990–2008.
- (8) Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **2009**, *5*, 1632–1639.
- (9) Narumi, T.; Ohno, Y.; Okimoto, N.; T, K.; Suenaga, A.; Futatsugi, N.; Yanai, R.; Himeno, R.; Fujikawa, S.; Ikei, M.; Taiji, M. A 55 TFLOPS simulation of amyloid-forming peptides from yeast prion Sup35 with the special purpose computer system MDGRAPE-3. *Proceedings of the SC06 (High Performance Computing, Networking, Storage and Analysis), Tampa, FL, USA 2006*, 49.
- (10) Ohmura, I.; Morimoto, G.; Ohno, Y.; Hasegawa, A.; Taiji, M. MDGRAPE-4: a special-purpose computer system for molecular dynamics simulations. *Philos. Trans. R. Soc. A* **2014**, *372*, 20130387.
- (11) Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Ierardi, D. J.; Kolossaváry, I.; Klepeis, J. L.; Layman, T.; McLeavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C. Anton, a Special-purpose Machine for Molecular Dynamics Simulation. *Commun. ACM* **2008**, *51*, 91–97.
- (12) Shaw, D. E.; Grossman, J. P.; Bank, J. A.; Batson, B.; Butts, J. A.; Chao, J. C.; Deneroff, M. M.; Dror, R. O.; Even, A.; Fenton, C. H.; Forte, A.; Gagliardo, J.; Gill, G.; Greskamp, B.; Ho, C. R.; Ierardi, D. J.; Iserovich, L.; Kuskin, J. S.; Larson, R. H.; Layman, T.; Lee, L.-S.; Lerer, A. K.; Li, C.; Killebrew, D.; Mackenzie, K. M.; Mok, S. Y.-H.; Moraes, M. A.; Mueller, R.; Nociolo, L. J.; Petricolas, J. L.; Quan, T.; Ramot, D.; Salmon, J. K.; Scarpa, D. P.; Ben Schafer, U.; Siddique, N.; Snyder, C. W.; Spengler, J.; Tang, P. T. P.; Theobald, M.; Toma, H.; Towles, B.; Vitale, B.; Wang, S. C.; Young, C. Anton 2: Raising the Bar for Performance and Programmability in a Special-purpose Molecular Dynamics Supercomputer. *Proc. Int. Conf. for High Performance Computing, Networking, Storage and Analysis*. **2014**, 41–53.
- (13) Shaw, D. E.; Adams, P. J.; Azaria, A.; Bank, J. A.; Batson, B.; Bell, A.; Bergdorf, M.; Bhatt, J.; Butts, J. A.; Correia, T.; Dirks, R. M.; Dror, R. O.; Eastwood, M. P.; Edwards, B.; Even, A.; Feldmann, P.; Fenn, M.; Fenton, C. H.; Forte, A.; Gagliardo, J.; Gill, G.; Gorlatova, M.; Greskamp, B.; Grossman, J.; Gullingsrud, J.; Harper, A.; Hasenplaugh, W.; Heily, M.; Heshmat, B. C.; Hunt, J.; Ierardi, D. J.; Iserovich, L.; Jackson, B. L.; Johnson, N. P.; Kirk, M. M.; Klepeis, J. L.; Kuskin, J. S.; Mackenzie, K. M.; Mader, R. J.; McGowen, R.; McLaughlin, A.; Moraes, M. A.; Nasr, M. H.; Nociolo, L. J.; O'Donnell, L.; Parker, A.; Petricolas, J. L.; Pocina, G.; Predescu, C.; Quan, T.; Salmon, J. K.; Schwink, C.; Shim, K. S.; Siddique, N.; Spengler, J.; Szalay, T.; Tabladillo, R.; Tartler, R.; Taube, A. G.; Theobald, M.; Towles, B.; Vick, W.; Wang, S. C.; Wazlowski, M.; Weingarten, M. J.; Williams, J. M.; Yuh, K. A. Anton 3: Twenty Microseconds of Molecular Dynamics Simulation before Lunch.

- Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. New York, NY, USA 2021, 1–11.
- (14) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334*, 517–520.
- (15) Mitsutake, A.; Iijima, H.; Takano, H. Relaxation mode analysis of a peptide system: Comparison with principal component analysis. *J. Chem. Phys.* **2011**, *135*, 164102.
- (16) Nagai, T.; Mitsutake, A.; Takano, H. Principal component relaxation mode analysis of an all-atom molecular dynamics simulation of human lysozyme. *J. Phys. Soc. Jpn.* **2013**, *82*, 023803.
- (17) Mitsutake, A.; Takano, H. Relaxation mode analysis and Markov state relaxation mode analysis for chignolin in aqueous solution near a transition temperature. *J. Chem. Phys.* **2015**, *143*, 124111.
- (18) Karasawa, N.; Mitsutake, A.; Takano, H. Two-step relaxation mode analysis with multiple evolution times applied to all-atom molecular dynamics protein simulation. *Phys. Rev. E* **2017**, *96*, 062408.
- (19) Karasawa, N.; Mitsutake, A.; Takano, H. Improved Relaxation Mode Analysis of a Hen Egg-White Lysozyme Protein. *Biophys. J.* **2017**, *112*, 353a.
- (20) Mitsutake, A.; Takano, H. Relaxation mode analysis for molecular dynamics simulations of proteins. *Biophys. Rev.* **2018**, *10*, 375.
- (21) Karasawa, N.; Mitsutake, A.; Takano, H. Identification of slow relaxation modes in a protein trimer via positive definite relaxation mode analysis. *J. Chem. Phys.* **2019**, *150*, 084113.
- (22) Mitsutake, A.; Takano, H. Folding pathways of NuG2—a designed mutant of protein G—using relaxation mode analysis. *J. Chem. Phys.* **2019**, *151*, 044117.
- (23) Maruyama, Y.; Takano, H.; Mitsutake, A. Analysis of molecular dynamics simulations of 10-residue peptide, chignolin, using statistical mechanics: Relaxation mode analysis and three-dimensional reference interaction site model theory. *Biophys. Physicobiol.* **2019**, *16*, 407–429.
- (24) Miyakawa, T.; Sugimori, K.; Kawaguchi, K.; Takasu, M.; Nagao, H.; Morikawa, R. Relationship between Dynamics of Structures and Dynamics of Hydrogen Bonds in Hras-GTP/GDP Complex. *Proceedings of the 2020 10th International Conference on Bioscience, Biochemistry and Bioinformatics*. New York, NY, USA, **2020**, 1–7.
- (25) Mori, T.; Saito, S. Dynamic heterogeneity in the folding/unfolding transitions of Fip35. *J. Chem. Phys.* **2015**, *142*, 135101.
- (26) Mori, T.; Saito, S. Molecular Mechanism Behind the Fast Folding/Unfolding Transitions of Villin Headpiece Subdomain: Hierarchy and Heterogeneity. *J. Phys. Chem. B* **2016**, *120*, 11683–11691.
- (27) Naritomi, Y.; Fuchigami, S. Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: The case of domain motions. *J. Chem. Phys.* **2011**, *134*, 065101.
- (28) Naritomi, Y.; Fuchigami, S. Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis. *J. Chem. Phys.* **2013**, *139*, 215102.
- (29) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **2013**, *139*, 015102.
- (30) Schwantes, C. R.; Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* **2013**, *9*, 2000–2009.
- (31) Schwantes, C. R.; Pande, V. S. Modeling Molecular Kinetics with tICA and the Kernel Trick. *J. Chem. Theory Comput.* **2015**, *11*, 600–608.
- (32) Takano, H.; Miyashita, S. Relaxation modes in random spin systems. *J. Phys. Soc. Jpn.* **1995**, *64*, 3688–3698.
- (33) Koseki, S.; Hirao, H.; Takano, H. Monte Carlo Study of Relaxation Modes of a Single Polymer Chain. *J. Phys. Soc. Jpn.* **1997**, *66*, 1631–1637.
- (34) Hirao, H.; Koseki, S.; Takano, H. Molecular dynamics study of relaxation modes of a single polymer chain. *J. Phys. Soc. Jpn.* **1997**, *66*, 3399–3405.
- (35) Kufareva, I.; Abagyan, R. *Homology modeling*; Springer, 2011; pp 231–257.
- (36) Carugo, O. Statistical validation of the root-mean-square-distance, a measure of protein structural proximity. *Protein Eng. Des. Sel.* **2007**, *20*, 33–37.
- (37) Harano, Y.; Roth, R.; Sugita, Y.; Ikeguchi, M.; Kinoshita, M. Physical basis for characterizing native structures of proteins. *Chem. Phys. Lett.* **2007**, *437*, 112–116.
- (38) Kinoshita, M. Importance of translational entropy of water in biological self-assembly processes like protein folding. *Int. J. Mol. Sci.* **2009**, *10*, 1064–1080.
- (39) Oshima, H.; Kinoshita, M. A highly efficient hybrid method for calculating the hydration free energy of a protein. *J. Comput. Chem.* **2016**, *37*, 712–723.
- (40) Kajiwara, Y.; Yasuda, S.; Takamuku, Y.; Murata, T.; Kinoshita, M. Identification of thermostabilizing mutations for a membrane protein whose three-dimensional structure is unknown. *J. Comput. Chem.* **2017**, *38*, 211–223.
- (41) Maruyama, Y.; Mitsutake, A. Stability of unfolded and folded protein structures using a 3D-RISM with the RMDFT. *J. Phys. Chem. B* **2017**, *121*, 9881–9885.
- (42) Maruyama, Y.; Mitsutake, A. Analysis of Structural Stability of Chignolin. *J. Phys. Chem. B* **2018**, *122*, 3801–3814.
- (43) Maruyama, Y.; Koroku, S.; Imai, M.; Takeuchi, K.; Mitsutake, A. Mutation-induced change in chignolin stability from  $\pi$ -turn to  $\alpha$ -turn. *RSC Adv.* **2020**, *10*, 22797–22808.
- (44) Maruyama, Y.; Mitsutake, A. Structural Stability Analysis of Proteins Using End-to-End Distance: A 3D-RISM Approach. *J.* **2022**, *S*, 114–125.
- (45) Harada, R.; Sladek, V.; Shigeta, Y. Nontargeted Parallel Cascade Selection Molecular Dynamics Based on a Nonredundant Selection Rule for Initial Structures Enhances Conformational Sampling of Proteins. *J. Chem. Inf. Model.* **2019**, *59*, 5198–5206.
- (46) Harada, R.; Sladek, V.; Shigeta, Y. Nontargeted Parallel Cascade Selection Molecular Dynamics Using Time-Localized Prediction of Conformational Transitions in Protein Dynamics. *J. Chem. Theory Comput.* **2019**, *15*, 5144–5153.
- (47) Yasuda, T.; Morita, R.; Shigeta, Y.; Harada, R. Independent Nontargeted Parallel Cascade Selection Molecular Dynamics (Ino-PaCS-MD) to Enhance the Conformational Sampling of Proteins. *J. Chem. Theory Comput.* **2021**, *17*, 5933–5943.
- (48) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-residue protein. *Nat. Struct. Biol.* **2002**, *9*, 425–430.
- (49) Barua, B.; Lin, J. C.; Williams, V. D.; Kummler, P.; Neidigh, J. W.; Andersen, N. H. The Trp-cage: optimizing the stability of a globular miniprotein. *Protein Eng. Des. Sel.* **2008**, *21*, 171–185.
- (50) Nauli, S.; Kuhlman, B.; Baker, D. Computer-based redesign of a protein folding pathway. *Nat. Struct. Biol.* **2001**, *8*, 602.
- (51) Beglov, D.; Roux, B. An integral equation to describe the solvation of polar molecules in liquid water. *J. Phys. Chem. B* **1997**, *101*, 7821–7826.
- (52) Kovalenko, A.; Hirata, F. Three-dimensional density profiles of water in contact with a solute of arbitrary shape: A RISM approach. *Chem. Phys. Lett.* **1998**, *290*, 237–244.
- (53) Matubayasi, N.; Nakahara, M. Theory of solutions in the energetic representation. I. Formulation. *J. Chem. Phys.* **2000**, *113*, 6070–6081.
- (54) Matubayasi, N.; Nakahara, M. Theory of solutions in the energy representation. II. Functional for the chemical potential. *J. Chem. Phys.* **2002**, *117*, 3605–3616.
- (55) Matubayasi, N.; Nakahara, M. Theory of solutions in the energy representation. III. Treatment of the molecular flexibility. *J. Chem. Phys.* **2003**, *119*, 9686–9702.
- (56) Nguyen, C. N.; Kurtzman Young, T.; Gilson, M. K. Grid inhomogeneous solvation theory: Hydration structure and thermodynamic properties of proteins. *J. Chem. Inf. Model.* **2023**, *63*, 1529–1541.

- namics of the miniature receptor cucurbit[7]uril. *J. Chem. Phys.* **2012**, *137*, 044101.
- (57) Nguyen, C. N.; Young, T. K.; Gilson, M. K. Erratum: "Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril" [J. Chem. Phys. 137, 044101 (2012)]. *J. Chem. Phys.* **2012**, *137*, 149901.
- (58) Maruyama, Y.; Harano, Y. Does water drive protein folding? *Chem. Phys. Lett.* **2013**, *581*, 85–90.
- (59) Pronk, S.; Päll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.
- (60) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prothom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (61) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
- (62) 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.
- (63) Sumi, T.; Mitsutake, A.; Maruyama, Y. A solvation-free-energy functional: A reference-modified density functional formulation. *J. Comput. Chem.* **2015**, *36*, 1359–1369.
- (64) Sumi, T.; Mitsutake, A.; Maruyama, Y. Erratum: "A solvation-free-energy functional: A reference-modified density functional formulation" [J. Comput. Chem. 2015, 36, 1359–1369]. *J. Comput. Chem.* **2015**, *2009*.
- (65) Maruyama, Y. Correction terms for the solvation free energy functional of three-dimensional reference interaction site model based on the reference-modified density functional theory. *J. Mol. Liq.* **2019**, *291*, 111160.
- (66) Maruyama, Y.; Hirata, F. Modified Anderson method for accelerating 3D-RISM calculations using graphics processing unit. *J. Chem. Theory Comput.* **2012**, *8*, 3015–3021.
- (67) Nukada, A.; Maruyama, Y.; Matsuoka, S. High performance 3-D FFT using multiple CUDA GPUs. *Proceedings of the 5th Annual Workshop on General Purpose Processing with Graphics Processing Units - GPGPU-5 2012*, *57*–63.
- (68) Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B* **2007**, *111*, 4553–4559.
- (69) Doshi, U.; Hamelberg, D. Achieving Rigorous Accelerated Conformational Sampling in Explicit Solvent. *J. Phys. Chem. Lett.* **2014**, *5*, 1217–1224.
- (70) Zaborowski, B.; Jagiela, D.; Czaplewski, C.; Halabis, A.; Lewandowska, A.; Źmudzińska, W.; Oldziej, S.; Karczyńska, A.; Omieczynski, C.; Wirecki, T.; Liwo, A. A Maximum-Likelihood Approach to Force-Field Calibration. *J. Chem. Inf. Model.* **2015**, *55*, 2050–2070.
- (71) Miao, Y.; Feixas, F.; Eun, C.; McCammon, J. A. Accelerated molecular dynamics simulations of protein folding. *J. Comput. Chem.* **2015**, *36*, 1536–1549.
- (72) Kamiya, M.; Sugita, Y. Flexible selection of the solute region in replica exchange with solute tempering: Application to protein-folding simulations. *J. Chem. Phys.* **2018**, *149*, 072304.
- (73) Harada, R.; Shigeta, Y. Temperature-Shuffled Structural Dissimilarity Sampling Based on a Root-Mean-Square Deviation. *J. Chem. Inf. Model.* **2018**, *58*, 1397–1405.
- (74) Harada, R.; Shigeta, Y. Selection Rules for Outliers in Outlier Flooding Method Regulate Its Conformational Sampling Efficiency. *J. Chem. Inf. Model.* **2019**, *59*, 3919–3926.
- (75) Omelyan, I.; Kovalenko, A. Enhanced solvation force extrapolation for speeding up molecular dynamics simulations of complex biochemical liquids. *J. Chem. Phys.* **2019**, *151*, 214102.
- (76) 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.
- (77) Chowdhury, S.; Lee, M. C.; Xiong, G.; Duan, Y. Ab initio Folding Simulation of the Trp-cage Mini-protein Approaches NMR Resolution. *J. Mol. Biol.* **2003**, *327*, 711–717.
- (78) Paschek, D.; Hempel, S.; Garcia, A. E. Computing the stability diagram of the Trp-cage miniprotein. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17754–17759.
- (79) Andryushchenko, V. A.; Chekmarev, S. F. A hydrodynamic view of the first-passage folding of Trp-cage miniprotein. *Eur. Biophys. J.* **2016**, *45*, 229–243.
- (80) Yasuda, T.; Shigeta, Y.; Harada, R. The Folding of Trp-cage is Regulated by Stochastic Flip of the Side Chain of Tryptophan. *Chem. Lett.* **2021**, *S0*, 162–165.
- (81) Zhou, R. Trp-cage: Folding free energy landscape in explicit water. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280–13285.
- (82) Ding, F.; Buldyrev, S. V.; Dokholyan, N. V. Folding Trp-Cage to NMR Resolution Native Structure Using a Coarse-Grained Protein Model. *Biophys. J.* **2005**, *88*, 147–155.
- (83) Iavarone, A. T.; Patriksson, A.; van der Spoel, D.; Parks, J. H. Fluorescence Probe of Trp-Cage Protein Conformation in Solution and in Gas Phase. *J. Am. Chem. Soc.* **2007**, *129*, 6726–6735.
- (84) Patriksson, A.; Adams, C. M.; Kjeldsen, F.; Zubarev, R. A.; van der Spoel, D. A Direct Comparison of Protein Structure in the Gas and Solution Phase: The Trp-cage. *J. Phys. Chem. B* **2007**, *111*, 13147–13150.
- (85) Scian, M.; Lin, J. C.; Le Trong, I.; Makhatadze, G. I.; Stenkamp, R. E.; Andersen, N. H. Crystal and NMR structures of a Trp-cage mini-protein benchmark for computational fold prediction. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 12521–12525.
- (86) Kannan, S.; Zacharias, M. Role of Tryptophan Side Chain Dynamics on the Trp-Cage Mini-Protein Folding Studied by Molecular Dynamics Simulations. *PLoS One* **2014**, *9*, e88383.
- (87) Byrne, A.; Williams, D. V.; Barua, B.; Hagen, S. J.; Kier, B. L.; Andersen, N. H. Folding Dynamics and Pathways of the Trp-Cage Miniproteins. *Biochemistry* **2014**, *53*, 6011–6021.
- (88) Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations. *PLoS Comput. PLoS Comput. Biol.* **2009**, *5*, e1000452.
- (89) Marino, K. A.; Bolhuis, P. G. Confinement-Induced States in the Folding Landscape of the Trp-cage Miniprotein. *J. Phys. Chem. B* **2012**, *116*, 11872–11880.
- (90) Han, W.; Schulten, K. Characterization of Folding Mechanisms of Trp-Cage and WW-Domain by Network Analysis of Simulations with a Hybrid-Resolution Model. *J. Phys. Chem. B* **2013**, *117*, 13367–13377.
- (91) Beauchamp, K. A.; McGibbon, R.; Lin, Y.-S.; Pande, V. S. Simple few-state models reveal hidden complexity in protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17807–17813.
- (92) Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 17874–17879.
- (93) Zheng, W.; Best, R. B. Reduction of All-Atom Protein Folding Dynamics to One-Dimensional Diffusion. *J. Phys. Chem. B* **2015**, *119*, 15247–15255.
- (94) Schwantes, C. R.; Shukla, D.; Pande, V. S. Markov State Models and tICA Reveal a Nonnative Folding Nucleus in Simulations of NuG2. *Biophys. J.* **2016**, *110*, 1716–1719.
- (95) Becerra, D.; Butyaev, A.; Waldspühl, J. Fast and flexible coarse-grained prediction of protein folding routes using ensemble modeling and evolutionary sequence variation. *Bioinformatics* **2019**, *36*, 1420–1428.
- (96) Shao, Q.; Zhu, W. Nonnative contact effects in protein folding. *Phys. Chem. Chem. Phys.* **2019**, *21*, 11924–11936.

- (97) Chang, L.; Perez, A.; Miranda-Quintana, R. A. Improving the analysis of biological ensembles through extended similarity measures. *Phys. Chem. Chem. Phys.* **2021**, *24*, 444–451.
- (98) Chang, L.; Perez, A. Deciphering the Folding Mechanism of Proteins G and L and Their Mutants. *J. Am. Chem. Soc.* **2022**, *144*, 14668–14677.
- (99) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
- (100) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millán, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876.
- (101) Mohan, A.; Oldfield, C. J.; Radivojac, P.; Vacic, V.; Cortese, M. S.; Dunker, A. K.; Uversky, V. N. Analysis of Molecular Recognition Features (MoRFs). *J. Mol. Biol.* **2006**, *362*, 1043–1059.
- (102) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32*, 2319–2327.
- (103) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

## □ Recommended by ACS

### Achieving Accurate Standard Protein–Protein Binding Free Energy Calculations through the Geometrical Route and Ergodic Sampling

Haohao Fu, Wensheng Cai, *et al.*

APRIL 12, 2023

JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ ▶

### Generating Protein Folding Trajectories Using Contact-Map-Driven Directed Walks

Ziad Fakhoury, Scott Habershon, *et al.*

MARCH 30, 2023

JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ ▶

### Resolution of Maximum Entropy Method-Derived Posterior Conformational Ensembles of a Flexible System Probed by FRET and Molecular Dynamics Simulations

Jonas Dittrich, Holger Gohlke, *et al.*

APRIL 06, 2023

JOURNAL OF CHEMICAL THEORY AND COMPUTATION

READ ▶

### MDexciteR: Enhanced Sampling Molecular Dynamics by Excited Normal Modes or Principal Components Obtained from Experiments

Mauricio G. S. Costa, David Perahia, *et al.*

JANUARY 09, 2023

JOURNAL OF CHEMICAL THEORY AND COMPUTATION

READ ▶

Get More Suggestions >