

# A computational method to simulate global conformational changes of proteins induced by cosolvent

Shoichi Tanimoto<sup>1</sup>  | Koichi Tamura<sup>2</sup>  | Shigehiko Hayashi<sup>2</sup>  |  
Norio Yoshida<sup>1</sup>  | Haruyuki Nakano<sup>1</sup> 

<sup>1</sup>Department of Chemistry, Graduate School of Science, Kyushu University, Fukuoka, Japan

<sup>2</sup>Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto, Japan

## Correspondence

Norio Yoshida, Department of Chemistry, Graduate School of Science, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan.  
Email: noriwo@chem.kyushu-univ.jp

## Present address

Shoichi Tanimoto, Institute for Molecular Science, Okazaki, Aichi, Japan

Koichi Tamura, Computational Biophysics Research Team, RIKEN Center for Computational Science, Kobe, Hyogo, Japan

## Funding information

Japan Society for the Promotion of Science, Grant/Award Numbers: 18K05036, 19H02677, 19J11525; Toyota Physical and Chemical Research Institute

## Abstract

A computational method to investigate the global conformational change of a protein is proposed by combining the linear response path following (LRPF) method and three-dimensional reference interaction site model (3D-RISM) theory, which is referred to as the LRPF/3D-RISM method. The proposed method makes it possible to efficiently simulate protein conformational changes caused by either solutions of varying concentrations or the presence of cosolvent species by taking advantage of the LRPF and 3D-RISM. The proposed method is applied to the urea-induced denaturation of ubiquitin. The LRPF/3D-RISM trajectories successfully simulate the early stage of the denaturation process within the simulation time of 300 ns, whereas no significant structural change is observed even in the 1  $\mu$ s standard MD simulation. The obtained LRPF/3D-RISM trajectories reproduce the mechanism of the urea denaturation of ubiquitin reported in previous studies, and demonstrate the high efficiency of the method.

## KEYWORDS

3D-RISM, denaturation, linear response theory, LRPF, MD simulation

## 1 | INTRODUCTION

Many biological processes proceed in a solution environment, and solvent effects play an essential role in such processes. Variations in the solution environment frequently result in dramatic changes in structure and function of proteins. For example, when denaturing agents are added to a native protein solution, the protein is unfolded, and when the concentration of the denaturants is decreased, the protein folds again to recover the original three-dimensional structure.<sup>1</sup> As another example, the voltage-gated Na<sup>+</sup> channel,<sup>2</sup> which is one of the membrane proteins, opens the gate with depolarization of the membrane potential and allows Na<sup>+</sup> to flow into the cell, and closes the gate when the intracellular Na<sup>+</sup> reaches a certain concentration. The global structural changes occur in these processes on a much slower time scale (milliseconds to seconds) than the typical time scale of molecular dynamics (MD) simulation (nanoseconds to microseconds).<sup>3</sup>

Therefore, to trace such processes, efficient sampling methods are desired.

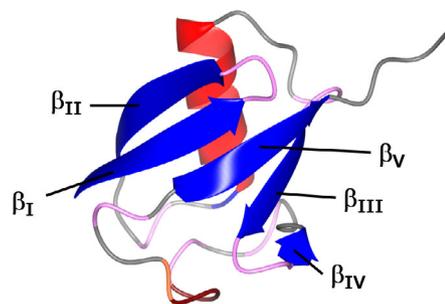
To date, many biased sampling methods have been developed to accelerate the structural sampling of processes involving global structural changes of proteins.<sup>4–10</sup> Ikeguchi et al. proposed a novel method to compute such processes based on the linear response theory (LRT).<sup>11</sup> In their method, the structural change of a protein due to the ligand binding is described as a response by atomic fluctuations in the ligand unbinding state and ligand–protein interactions. Following them, several computational methods have been proposed based on the LRT.<sup>12–18</sup>

The linear response path following (LRPF) method is one of the biased sampling methods based on the LRT, developed by Tamura and Hayashi,<sup>19</sup> which is applicable to prediction of global conformational changes without the knowledge of the target structure. In this method, the LRT is applied iteratively to move the adjacent

conformational substates of proteins, and then the pathway of global conformational changes can be found. The biasing forces are sequentially updated at each LRT step to take non-linearity of the structural change into account explicitly. A feature of this method is that the conformational changes related to the functions of proteins can be naturally considered on the basis of the LRT. The authors applied the LRPF method to structural changes of yeast calmodulin N-terminal domain upon calcium binding<sup>19</sup> and ADP/ATP membrane transporter upon the ADP binding,<sup>20</sup> and succeeded in simulating global conformational changes to structures in different functional states. Especially, in the latter case, the LRPF simulation<sup>20</sup> was used to successfully predict the structure in the alternating access state at the atomic level, which was experimentally unknown at that point and was later determined by X-ray crystallography.<sup>21</sup>

Although the LRPF method has been successfully applied to the global structural changes of proteins induced by ligand binding at a specific position, it would be difficult to apply it to processes for which the binding site is unknown. For such cases, the position or distribution of the ligand needs to be determined to evaluate the biasing forces. For this purpose, the three-dimensional reference interaction site model (3D-RISM) theory, which is a statistical mechanics theory of molecular liquids,<sup>22–24</sup> is capable of providing a spatial distribution of solvent, ions and small-ligands around the solute molecules. The distribution corresponds to the complete ensemble average over the configuration space of solvent species including ions in the thermodynamic limit under certain approximations.<sup>24,25</sup> This means that the spatial distribution computed by the 3D-RISM theory is not dependent on the initial configuration of the solvent species and is not limited by the size of the system and the number of samplings. Taking advantage of this feature, the 3D-RISM theory has succeeded in predicting the distributions of water molecules confined inside the protein and ions existing near the protein surface, which are in good agreement with the experimental data,<sup>26–28</sup> without prior information. Therefore, the 3D-RISM theory is expected to be suitable for finding the ligand position or distribution to generate the biasing forces.

In this study, a computational method for simulating global conformational changes of proteins induced by solution by combining the LRPF method and 3D-RISM theory (hereafter, referred to as the LRPF/3D-RISM method) is proposed. Hirata and co-workers proposed a theory for calculating protein structural changes by combining the LRT and 3D-RISM theory.<sup>29–34</sup> They used the generalized Langevin equation<sup>35–39</sup> to describe protein dynamics. Unlike their approach, the present LRPF/3D-RISM method employs an MD simulation to obtain a variance–covariance matrix representing the protein conformational fluctuation, which allows large molecules to be handled in a practical manner. The LRPF/3D-RISM method is applied to the urea denaturation of ubiquitin as a proof-of-concept. This denaturation process has been extensively studied, and much insight has been accumulated to date because of the medium size (76 residues) and the secondary structure with both  $\alpha$ -helix and  $\beta$ -sheets (Figure 1). Therefore, this system was suitable for the verification of the proposed method. Herein, we apply the LRPF/3D-RISM method to this



**FIGURE 1** Overview of the ubiquitin structure taken from protein data bank (PDB) entry id: 1ubq. An  $\alpha$ -helix structure and a  $\beta$ -sheet one are represented in red and blue, respectively. The  $\beta$ -sheet strands are labeled from the N-terminus

system and demonstrate the efficiency of the approach by comparing it with the standard MD simulation method.

## 2 | THEORY

### 2.1 | LRPF method

In this subsection, we briefly review the LRPF method,<sup>19</sup> which is the basis of our new method. In the LRPF simulation, the directed perturbative forces based on the LRT proposed by Ikeguchi et al.<sup>11</sup> are iteratively applied to protein atoms to simulate the global conformational changes of a protein induced by ligand binding. According to the LRT, the response of a protein to perturbations from ligands can be described by Equation (1):

$$\Delta \mathbf{R}_\alpha \approx \beta \sum_\gamma C_{\alpha\gamma} \mathbf{f}_\gamma \quad (1)$$

where  $\Delta \mathbf{R}_\alpha$  is a displacement vector of a protein atom  $\alpha$ .  $\beta = 1/k_B T$  is the inverse temperature, where  $k_B$  and  $T$  are the Boltzmann constant and the absolute temperature, respectively.  $C_{\alpha\gamma}$  is the variance–covariance matrix element of the protein atoms in the ligand-free state.  $\mathbf{f}_\gamma$  is the external force acting on a protein atom  $\gamma$ , representing the ligand binding. In the present LRPF simulation, the external force acting on the atom  $\gamma$  is obtained from Equation (2):

$$\bar{\mathbf{f}}_\gamma = \langle \mathbf{f}_{\gamma, \text{self}} \rangle_0 + \sum_{\substack{\alpha \in S(\gamma) \\ \text{except } \gamma}} \langle \mathbf{f}_\alpha \rangle_0 \quad (2)$$

where  $\bar{\mathbf{f}}_\gamma$  is the contracted force acting on the protein atom  $\gamma$ ,  $\mathbf{f}_{\gamma, \text{self}}$  is the perturbative force directly acting on the protein atom  $\gamma$ , and the second term in the right-hand side of Equation (2) is the sum of the perturbative force components acting on all atoms other than the atom  $\gamma$  in the residue  $S(\gamma)$  containing the atom.  $\langle \rangle$  indicates the thermal average of the protein conformations and the subscript “0” means the average without the perturbation. In the LRPF simulation, the

contracted force,  $\bar{f}_\gamma$ , calculated from Equation (2) is employed as the external force,  $f_\gamma$ , in Equation (1).

The biasing force directly acting on the protein atom  $\alpha$  introduced in the LRPF simulation is given by Equation (3):

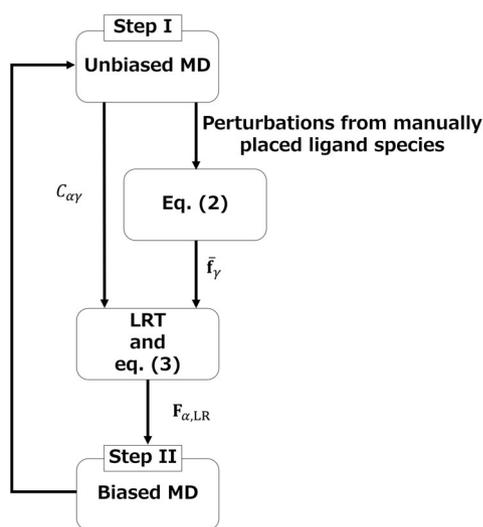
$$\mathbf{F}_{\alpha,LR} = \eta \Delta \mathbf{R}_\alpha \quad (3)$$

where  $\eta$  is a scaling factor that controls the strength of the biasing force. The work applied by the biasing forces,  $W$ , is defined by Equation (4):

$$W = d \sum_{\alpha} |\mathbf{F}_{\alpha,LR}| \quad (4)$$

where  $d$  represents the root-mean-squared deviation (RMSD) of the backbone chain atoms undergoing forced displacement in the biased MD simulation, which is set to 3 Å in this study in accordance with the original article.<sup>19</sup> During the biased MD simulation, the value of  $\eta$  is determined such that  $W$  takes a specific value. The other details are presented in the original article.<sup>19</sup>

The LRPF simulation consists of the following series of iterative MD simulation cycles (see Scheme 1). The first step is an unbiased MD simulation for obtaining the biasing forces,  $\mathbf{F}_{\alpha,LR}$ , in Equation (3). In this step, the variance-covariance matrix element of protein atoms in the ligand-free state,  $C_{\alpha\gamma}$ , and the external forces,  $\bar{f}_\alpha$ , in Equation (2) are evaluated with the trajectory obtained by the equilibrium MD simulation. In the second step, a biased MD simulation is performed using the biasing forces obtained in the first step. After that, an unbiased MD simulation of the first step of the next cycle is performed to relax the system and obtain the biasing forces used in the next cycle. By repeating the above series of cycles and successively updating the biasing forces applied to the system, the LRPF



**SCHEME 1** Simulation procedure of the original LRPF method to compute the global conformational changes of a protein due to the ligand binding

simulation makes it possible to simulate nonlinear global conformational changes of proteins without detailed information on the target structure.<sup>19</sup>

## 2.2 | 3D-RISM theory

We briefly explain the 3D-RISM equations as well as related equations required in this study; details are presented in the various literature.<sup>22–24,40–44</sup>

In this theory, one can obtain the spatial distribution function of solvent site  $i$ ,  $g_i(\mathbf{r})$ , around a solute molecule by solving the 3D-RISM equation under the given solute structure. In the present study, the protein is regarded as “solute” whereas water, ions, and denaturant molecules are “solvent.” The 3D-RISM equation is given by Equation (5)<sup>22,23</sup>:

$$h_i(\mathbf{r}) = \sum_j c_j(\mathbf{r}) * \{ \bar{\omega}_{ji}(|\mathbf{r}|) + \rho_j \bar{h}_{ji}(|\mathbf{r}|) \} \quad (5)$$

where  $h$  is the solute–solvent total correlation function, which is related to the spatial distribution function by  $g(\mathbf{r}) = h(\mathbf{r}) + 1$ ,  $c$  is the solute–solvent direct correlation function,  $\bar{\omega}$  is the intramolecular correlation function of the solvent species,  $\rho$  is the number density of solvent species, and  $\bar{h}$  is the total correlation function of solvent species obtained by solving the one-dimensional RISM (1D-RISM) equation for the solvent system in advance. The Roman subscripts,  $i, j, \dots$ , refer to the interaction sites of solvent molecules, and the asterisk denotes the convolution integral. One can solve the 3D-RISM equation numerically coupled with the Kovalenko–Hirata (KH) closure (Equation [6])<sup>24,41,44</sup>:

$$g_i(\mathbf{r}) = \begin{cases} \exp\{d_i(\mathbf{r})\}, & \text{if } d_i(\mathbf{r}) \leq 0 \\ 1 + d_i(\mathbf{r}), & \text{if } d_i(\mathbf{r}) > 0 \end{cases} \quad (6)$$

$$d_i(\mathbf{r}) = -\beta u_i(\mathbf{r}) + h_i(\mathbf{r}) - c_i(\mathbf{r})$$

where  $u$  is the solute–solvent pair interaction potential function given by Equation (7):

$$u_i(\mathbf{r}) = \sum_{\alpha} \left[ 4\epsilon_{\alpha i} \left\{ \left( \frac{\sigma_{\alpha i}}{|\mathbf{r} - \mathbf{R}_{\alpha}|} \right)^{12} - \left( \frac{\sigma_{\alpha i}}{|\mathbf{r} - \mathbf{R}_{\alpha}|} \right)^6 \right\} + \frac{q_{\alpha} q_i}{|\mathbf{r} - \mathbf{R}_{\alpha}|} \right] \quad (7)$$

where  $\mathbf{R}_{\alpha}$  is the position of the solute atom  $\alpha$ , and  $\epsilon$ ,  $\sigma$ , and  $q$  are the Lennard–Jones (LJ) parameters and partial charges, respectively, with the usual meaning.

In the LRPF/3D-RISM scheme proposed in the subsequent subsection, the mean force on the solute site  $\alpha$  induced by the solvent species  $l$ ,  $\mathbf{F}_{\alpha,\text{sol},l}$ , is determined based on the distribution of the solvent species around the protein obtained by the 3D-RISM theory. Here,  $l$  includes the solvent species assumed to be a ligand that induces conformational changes of the protein. The force is calculated by Equation (8)<sup>45,46</sup>:

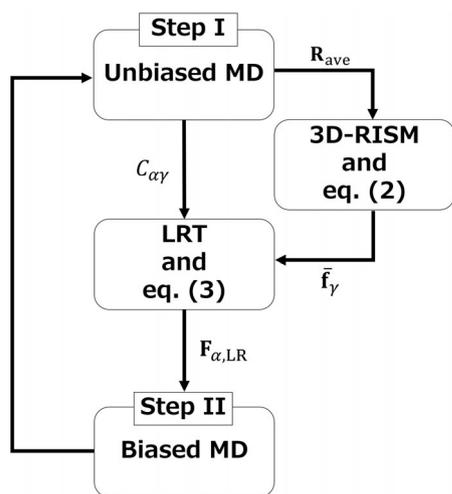
$$\mathbf{F}_{\alpha,\text{solv}} = -\frac{\partial \Delta\mu_l}{\partial \mathbf{R}_\alpha} = -\sum_{i \in l} \rho_i \int d\mathbf{r} \frac{\partial u_i(\mathbf{r})}{\partial \mathbf{R}_\alpha} g_i(\mathbf{r}) \quad (8)$$

where  $\Delta\mu_l$  is the solvation free energy component of the solvent species  $l$ . Since urea is regarded as a ligand in the present study, the summation in the right-hand side of Equation (8) runs over the interaction sites of urea.

### 2.3 | LRPF/3D-RISM method

In the original LRPF simulation for the ion-binding protein,<sup>19</sup> the external forces exerted by the ion binding are evaluated with bound ions that were intuitively placed based on experimental evidence without thorough statistical sampling of the ion positions. Hence, in principle, the arbitrariness of the ligand binding positions remains. Furthermore, in the case of conformational changes caused by varying the solution environment, it is not certain to which part of the protein the external forces triggering global conformational changes act. In the method proposed here, the LRPF/3D-RISM method, the mean forces induced by the solvent species,  $\mathbf{F}_{\alpha,\text{solv}}$ , evaluated by the 3D-RISM theory, are employed for the external forces representing the ligand binding or change of solution components, which makes it applicable to the cases described above. The 3D-RISM theory provides a distribution probability of a ligand having a conspicuous peak at a location corresponding to the binding site detected in the experiment.<sup>27</sup> Therefore, the forces induced by the solvent species acting on a protein can be determined a priori without experimental information regarding the binding site of the ligand or the sites that make a major contribution to conformational changes.

The computational procedure of the LRPF/3D-RISM method is schematically depicted in Scheme 2. Similar to the original LRPF method, this method consists of iterative cycles that include



**SCHEME 2** Simulation procedure of the LRPF/3D-RISM method to compute the global conformational changes of a protein induced by interaction with solvent components

two MD simulation steps. The difference is that, after an unbiased MD simulation to compute the variance-covariance matrix elements of protein atoms in the ligand-free state,  $C_{\alpha\gamma}$ , of step I, the 3D-RISM calculation is performed using the obtained averaged structure,  $\mathbf{R}_{\text{ave}}$ , to evaluate  $\mathbf{F}_{\alpha,\text{solv}}$ . Then, the external forces,  $\bar{\mathbf{f}}_\gamma$ , are evaluated from the calculated  $\mathbf{F}_{\alpha,\text{solv}}$  and Equation (2). Thereafter, the biasing forces,  $\mathbf{F}_{\alpha,\text{LR}}$ , are determined from the LRT (Equation (1)) and Equation (3), and the biased MD simulation is performed as step II. After that, an unbiased MD simulation of step I of the next cycle to relax the system and obtain the biasing forces used in the next cycle is performed. Steps I and II are repeated until the target structure is obtained.

## 3 | COMPUTATIONAL DETAILS

*Preparatory equilibration of ubiquitin.* We employed the crystal structure of ubiquitin (PDB entry id: 1ubq) as initial structure.<sup>47</sup> All MD simulations were performed using the program NAMD<sup>48</sup> with the CHARMM36m force field for the protein.<sup>49</sup> TIP3P parameters<sup>50,51</sup> were employed for water molecules. Hydrogen atoms were added using the psfgen plugin in NAMD. The entire protein was immersed in a rectangular box of size ca.  $55^3 \text{ \AA}^3$ . The atomic coordinates of the simulation box filled with water molecules were set up with the solvate plugin in VMD.<sup>52</sup> A nonbonded cutoff of  $12 \text{ \AA}$  was used, with the switching function beginning at  $10 \text{ \AA}$ . Long-range electrostatic interactions were calculated with the particle mesh Ewald method<sup>53</sup> with tolerance of  $1 \times 10^{-6}$  and the grid size was set to  $64^3$ . Periodic boundary conditions were used. The internal degrees of freedom of water molecules were constrained with the SETTLE algorithm.<sup>54</sup> Bonds including a hydrogen atom were constrained by the RATTLE algorithm.<sup>55</sup> Hereafter, the above calculation conditions were used unless otherwise noted.

The system was first minimized with 10,000 steps of the conjugate gradient method. All MD simulations below were performed under the constant NPT condition with  $P = 1 \text{ bar}$  and  $T = 298 \text{ K}$  unless otherwise noted. The Langevin heat bath with a damping coefficient of  $2 \text{ ps}^{-1}$  was used to keep the temperature constant. The pressure was maintained with the Berendsen's barostat<sup>56</sup> using a relaxation time of 100 fs. The integration time step was set to 2 fs for all MD simulations except for the heating phase, for which a time step of 1 fs was used. The energetically minimized system corresponding to 0 K was gradually heated by reassigning velocities at every 1 ps. The temperature of the system was linearly elevated to 298 K in 298 ps. After the heating phase, further equilibration at 298 K lasted for 102 ps. To further equilibrate solvents around the protein, protein atoms were restrained at their initial positions by harmonic potentials, and a restrained simulation was carried out for 2 ns. The force constants of restraining potentials were gradually decreased from 10 to  $0.00625 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  in the course of the simulation. After the restrained simulation, an equilibrium simulation without restraint was conducted for 120 ns. Trajectories were recorded every 100 fs for analysis.

**LRPF/3D-RISM simulations.** Five snapshots were taken from the 120 ns equilibrium trajectory from which LRPF/3D-RISM simulations were started. The first cycles of the LRPF1, LRPF2, LRPF3, LRPF4, and LRPF5 were started from the snapshots at 120, 108, 96, 84, and 72 ns, respectively. The 108–120, 96–108, 84–96, 72–84, and 60–72 ns portions of the 120 ns trajectory were used to calculate the initial variance–covariance matrices in the LRPF1, LRPF2, LRPF3, LRPF4, and LRPF5, respectively. In this study, only  $C_{\alpha}$  atoms were used to calculate the matrices. From the second cycle onward, the last snapshot of the unbiased MD simulation in each cycle was used for the initial structure of the next biased MD simulation, and the last 12 ns trajectory of the unbiased MD simulation was used for the calculation of the variance–covariance matrix. The adjustable parameter  $W$  in Equation (4), which governs the efficiency of the LRPF-part of the simulation, was determined on the basis of an observation obtained from the previous study<sup>57</sup>: the maximum RMSD of ubiquitin denaturation induced by guanidinium chloride is about 12 Å. It is thus reasonable to set the maximum value to the upper-limit of RMSD realized by a single LRPF cycle and adjust the value of  $W$  according to the results of trial simulations. For example, we started with  $W = 90$  for the 1st cycle of the LRPF1 trajectory, but it underwent large conformational change and the RMSD exceeded the upper-limit of RMSD. So, we reset the parameter to  $W = 60$ . In the next cycle, starting from the values adopted in the previous cycle, the  $W$  was adjusted:  $W$  was reduced if the upper limit of the RMSD was exceeded and increased if no significant structural change was observed. Although the way to control the value of  $W$  may seem ad hoc, it is difficult to determine the appropriate value of  $W$  a priori before the calculation because of complex nature of the coupling between the linear response force and the nonlinear unfolding dynamics. Because it is not our purpose to predict the very precise conformational change pathway of the unfolding event and the purpose of the present work is to obtain a rough picture of how mechanical unfolding proceeds in a limited simulation time, the protocol is acceptable. In the biased MD simulation, the magnitude of the biasing force was gradually raised by  $W = 1 \text{ kcal mol}^{-1}$  every 200 ps. When  $W$  reached a predetermined value, the biasing force was kept constant. Then, the force was gradually decreased by  $W = 1 \text{ kcal mol}^{-1}$  every 200 ps until the biasing force completely disappeared.

Hereafter, an unbiased MD simulation of the first step of the next cycle was carried out. All the trajectories were recorded every 100 fs for analysis. The simulation times for the unbiased MD simulation following the biased MD simulation were chosen randomly in the range of 20–45 ns with reference to the previous study.<sup>19</sup> When the preset simulation time has elapsed, the RMSD was checked to see if it has plateaued for at least 12 ns as the variance–covariance matrix needs to be calculated for an equilibrium trajectory, and if not, the simulation time was extended. Details of each of the LRPF/3D-RISM simulations are summarized in Table 1.

The biasing forces were obtained as follows. In this study, we applied the forces only to  $76 - 2 = 74 C_{\alpha}$  atoms (excluding the first and the last residues). All 3D-RISM calculations for obtaining the mean forces induced by urea were conducted with an in-house RISM/3D-RISM code.<sup>58</sup> We employed the KH closure for solving the 1D- and 3D-RISM equations.<sup>24,41,44</sup> All the explicit water molecules were omitted from the snapshots. The same potential parameters as in the equilibrium MD simulation were employed for the protein and water molecules with modified hydrogen parameters ( $\sigma_{\text{H}} = 0.4 \text{ \AA}$  and  $\varepsilon_{\text{H}} = 0.046 \text{ kcal mol}^{-1}$ ) for the TIP3P water model. Urea parameters were taken from references.<sup>59,60</sup> The Lorentz–Berthelot combination rule was applied in the calculations of the LJ parameters between different sites. Water/urea mixed solution (2 M) in the ambient condition was assumed at 298 K. The 1D-RISM calculation for evaluating the correlation functions of solvent species was performed with 8192 grid points with a spacing of 0.05 Å. The number of grid points in the 3D-RISM calculations was  $256^3$  with a spacing of 0.5 Å. The modified direct inversion in the iterative subspace (MDIIS)<sup>40</sup> was used to converge the 1D- and 3D-RISM equations, where the convergence threshold was  $10^{-8}$  with respect to the RMSD of the correlation functions.

**Standard MD simulation for the water/urea mixed solution system.** We also performed standard MD simulations in water/urea mixed solution to compare the results. The same potential parameters as in the equilibrium MD simulation and 3D-RISM calculations were employed for all the species. Unless otherwise noted, the same simulation settings as the equilibrium MD simulation and the same calculation conditions as the above 3D-RISM calculations were used.

| Cycle no. | LRPF1                    | LRPF2                    | LRPF3                    | LRPF4                    | LRPF5                    |
|-----------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 1         | $W = 60$<br>66 (30) ns   | $W = 50$<br>69 (40) ns   | $W = 45$<br>42 (20) ns   | $W = 40$<br>44 (20) ns   | $W = 55$<br>53 (20) ns   |
| 2         | $W = 55$<br>72 (45) ns   | $W = 70$<br>62 (20) ns   | $W = 60$<br>72 (45) ns   | $W = 45$<br>72 (45) ns   | $W = 65$<br>69 (30) ns   |
| 3         | $W = 45$<br>127 (100) ns | $W = 50$<br>130 (100) ns | $W = 90$<br>154 (100) ns | $W = 50$<br>130 (100) ns | $W = 40$<br>47 (30) ns   |
| 4         |                          |                          |                          |                          | $W = 45$<br>127 (100) ns |
| Total     | 265 ns                   | 261 ns                   | 268 ns                   | 246 ns                   | 296 ns                   |

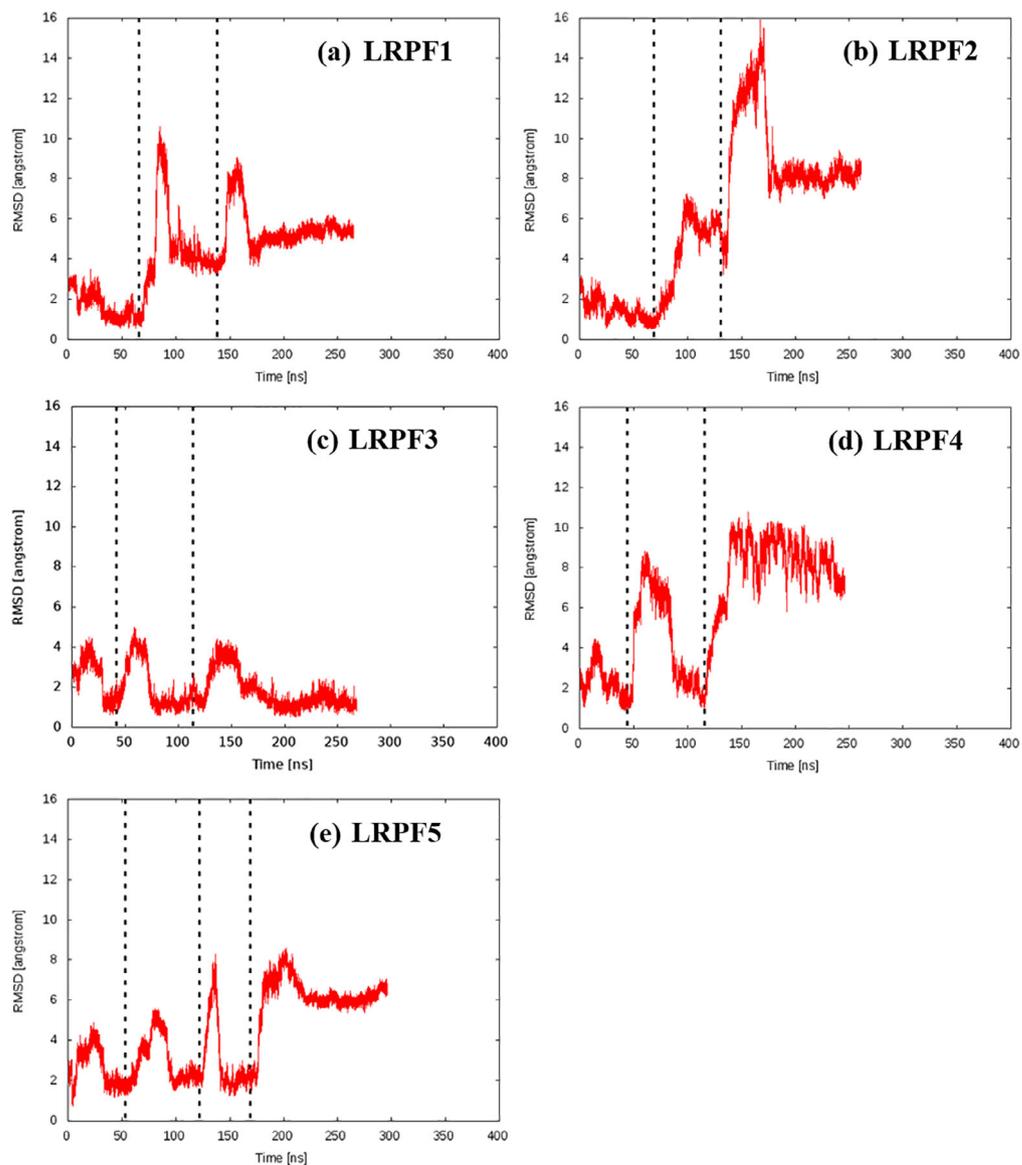
**TABLE 1** Summary of the LRPF/3D-RISM simulations

Note: The simulation time for the unbiased MD simulation step is given in parentheses. The unit of  $W$  is given in  $\text{kcal mol}^{-1}$ . The magnitude of the biasing forces is defined in Equation (4).

The five standard MD simulations were started with the same structure as the initial structure of the first cycle of the five LRPF/3D-RISM simulations. To determine the initial placement of urea molecules, the 3D-RISM calculations of the protein in the water/urea mixed solution (1 M) in the ambient condition was performed, and the Placevent algorithm<sup>61</sup> written in Python 2.7 utilizing the SciPy package<sup>62</sup> was used. The program was set to stop placing atoms when the highest remaining population grid was less than 1.5 times that of bulk. Then, the protein and placed urea molecules were immersed in a rectangular box with a size of approximately  $65^3 \text{ \AA}^3$ . Long-range electrostatic interactions were calculated with the particle mesh Ewald method<sup>53</sup> with tolerance of  $1 \times 10^{-6}$  and the grid size of  $81^3$ . To relax the solvents (urea and water molecules) around the protein, the system was first gradually heated by reassigning velocities at every 1 ps with the protein fixed at its position and the system volume kept constant. The temperature of the system was linearly elevated to 298 K in 298 ps. After the heating phase, further equilibration at 298 K lasted for 10.702 ns in the NVT ensemble with the protein fixed at its

position. To further equilibrate solvents, protein atoms were restrained at their initial positions by harmonic potentials, and a restrained simulation was carried out for 10 ns in the NPT ensemble. The force constants of restraining potentials were linearly decreased from 10 to 0 kcal mol<sup>-1</sup> Å<sup>-2</sup> in the course of the simulation. After the restrained simulation, an equilibrium simulation without restraint was conducted for 10 ns in the NPT ensemble. The final concentration of each system was in the range of 7.4 to 8.0 M. Then, a production run simulation was performed in the NPT ensemble for 1 μs.

**Trajectory analysis.** The RMSDs for 74 C<sub>α</sub> atoms were calculated with reference to the crystal structure using the cpptraj program.<sup>63</sup> In addition, the root-mean-squared fluctuations (RMSFs) for all C<sub>α</sub> atoms in the LRPF/3D-RISM simulations were also calculated using the cpptraj program. The secondary structure of the protein was analyzed using the DSSP program<sup>64</sup> implemented in the cpptraj program. The visual analysis of the trajectories and conformations was done using VMD,<sup>52</sup> the protein conformations in Figures 1, 5, and S5–S7 were depicted using CCP4mg,<sup>65</sup> and the protein conformations and



**FIGURE 2** Time evolution of the RMSDs in the LRPF/3D-RISM simulation. Vertical dashed lines represent cycles of the LRPF/3D-RISM simulations

distribution functions of atoms of the urea molecule in Figures 6–8 and S8 were depicted using the UCSF Chimera package.<sup>66</sup> All plots were depicted using gnuplot version 4.6.

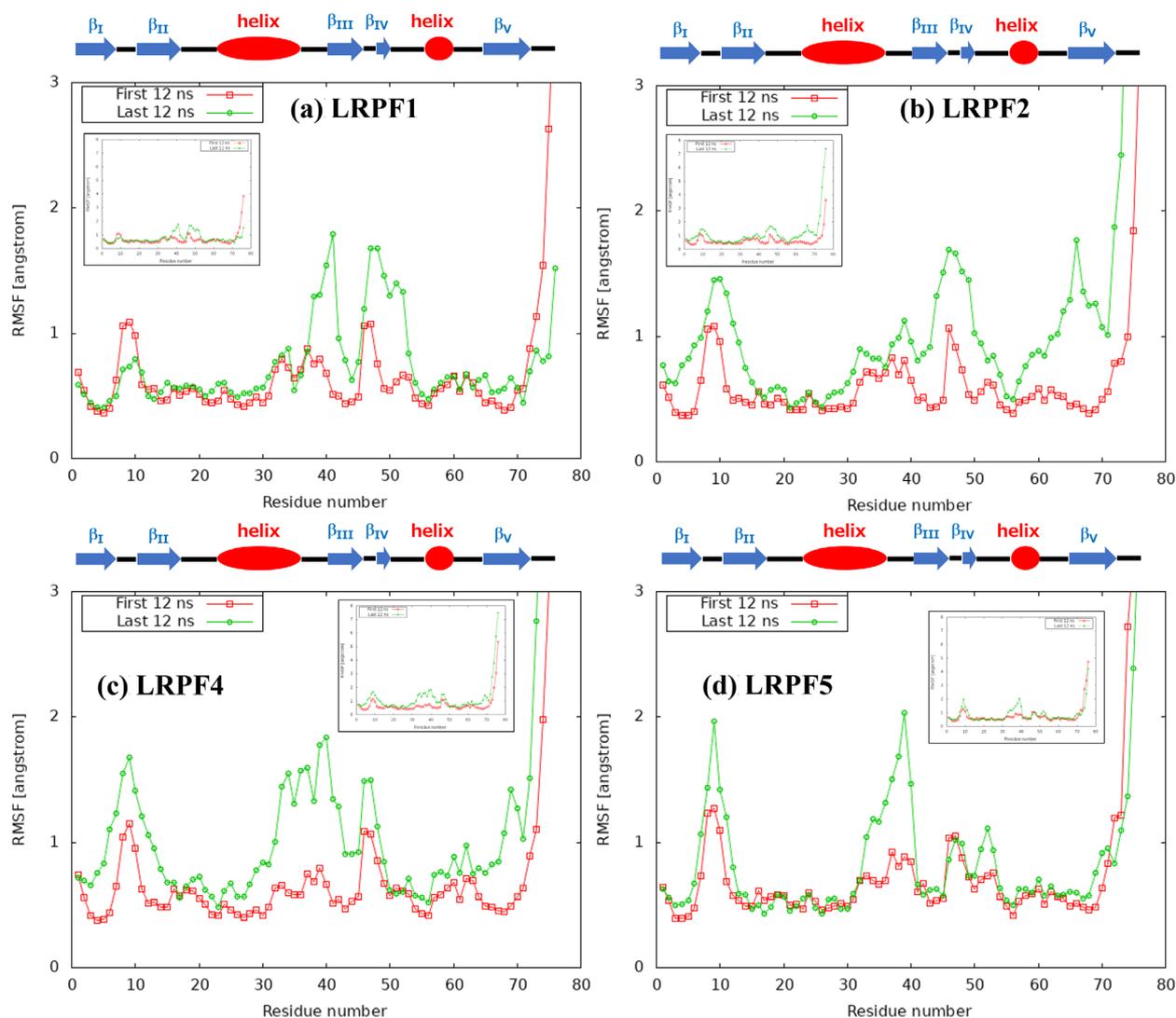
## 4 | RESULTS AND DISCUSSION

### 4.1 | Structure change

The time evolution of the RMSDs of 74  $C_{\alpha}$  atoms from the crystal structure in five LRPF/3D-RISM trajectories is depicted in Figure 2. For comparison, Figure S1 shows the corresponding plots in five standard MD trajectories of the water/urea mixed solution system. In the

LRPF/3D-RISM trajectories, the protein structures show remarkable changes (RMSD  $>5$  Å) within 300 ns except for the LRPF3. On the other hand, the standard MD trajectories do not show any significant conformational changes (RMSD  $<4$  Å) within the simulation time of 1  $\mu$ s. These results suggest that the LRPF/3D-RISM method can describe conformational changes of proteins induced by the surrounding solution environment more efficiently than the standard MD simulations.

To examine the detail of the conformational changes for the four trajectories (LRPF1, LRPF2, LRPF4, and LRPF5), Figure 3 shows the RMSF values of all the  $C_{\alpha}$  atoms from the averaged structure in the 12 ns trajectory used for the calculation of the first cycle variance-covariance matrix and averaged structure in the last 12 ns trajectory



**FIGURE 3** RMSFs of all the  $C_{\alpha}$  atoms in four of five LRPF/3D-RISM trajectories. Native secondary structures of ubiquitin are schematically depicted at the top of each panel where the red circles, blue arrows, and black lines represent helical structures,  $\beta$ -sheet strands, and random-coil structures, respectively. Red lines with open squares: RMSFs from the averaged structure in the 12 ns trajectory used for the calculation of the first cycle variance-covariance matrix (the 108–120 ns portion of the 120 ns equilibrium simulation for the LRPF1, 96–108 ns portion of the 120 ns equilibrium simulation for the LRPF2, 72–84 ns portion of the 120 ns equilibrium simulation for the LRPF4, and 60–72 ns portion of the 120 ns equilibrium simulation for the LRPF5); green lines with open circles: RMSFs from the averaged structure in the last 12 ns of each of the four LRPF/3D-RISM trajectories. The inset in each panel shows the wide  $y$ -range of the same plot

of each of four LRPF/3D-RISM trajectories. For the first 12 ns, the loop structures between  $\beta_I$  and  $\beta_{II}$ ,  $\beta_{III}$ , and  $\beta_{IV}$ , and the C-terminal part are more flexible (RMSF  $>1$  Å) than the other parts. On the other hand, the fluctuation of all the secondary structures is small, indicating that these are relatively rigid in aqueous environment. After the LRPF/3D-RISM simulation, some of the helices and  $\beta$ -sheets have larger fluctuations than those before the LRPF/3D-RISM simulation when looking at the four trajectories as a whole. In particular, the second helix,  $\beta_{III}$ , and  $\beta_{IV}$  are more flexible than the other parts folded in secondary structural conformations. Furthermore, in the LRPF2 and LRPF4, the entire structure is extensively fluctuated, and the RMSF values of two N-terminal  $\beta$ -sheets ( $\beta_I$  and  $\beta_{II}$ ) and the C-terminal  $\beta$ -sheet ( $\beta_V$ ) also become larger than those before the LRPF/3D-RISM simulation (see Figure 3 (b) and (c)). These features are in good agreement with those of the early stage of the denaturation reported in previous studies.<sup>67,68</sup>

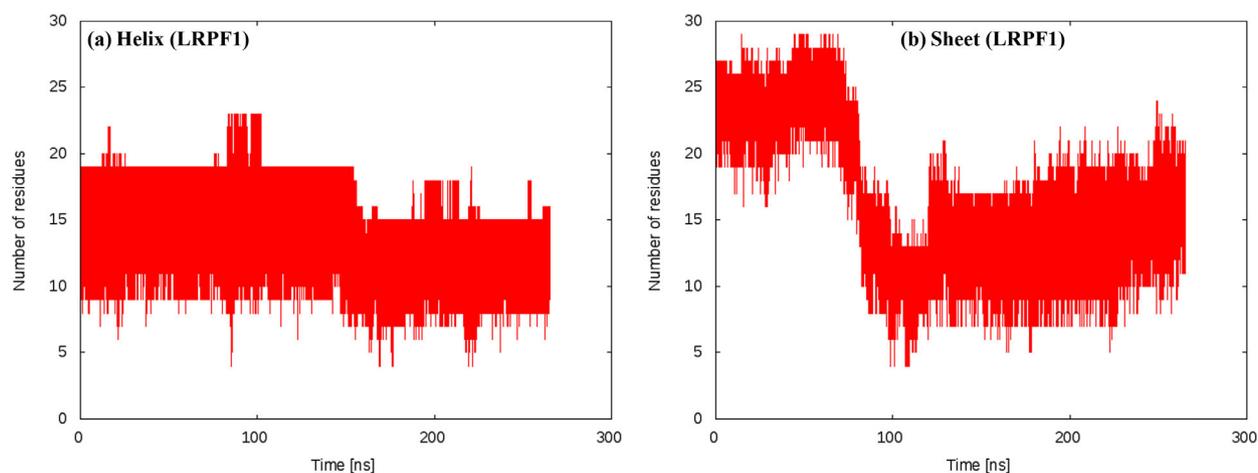
We next analyzed the secondary structure using the DSSP program<sup>64</sup> implemented in the cpptraj program<sup>63</sup> to characterize the process of the secondary structural changes in the LRPF/3D-RISM trajectories. The results of the structure analysis on the LRPF1 trajectory are plotted in Figure 4 as a time evolution diagram of the number of residues constituting the helical and sheet structures. The corresponding plots for the LRPF2, LRPF4, and LRPF5 are shown in Figures S2–S4, respectively. The number of residues constituting the  $\beta$ -sheets rapidly decreases at an early time ( $<100$  ns). On the other hand, although the helix conformation is slightly reduced around 150 ns, the helical structure is almost completely maintained over the entire simulation time. A similar tendency is observed in the other trajectories (see Figures S2–S4). Furthermore, unlike the helical structure, the number of residues constituting the  $\beta$ -sheet structure varies considerably over the entire simulation time. To see which part was disrupted, the averaged structure of the last 12 ns in each cycle of the LRPF1 trajectory is shown in Figure 5 along with time evolution of the RMSDs. The corresponding figures for the LRPF2, LRPF4, and LRPF5 are shown in Figures S5–S7. It is clear from Figures 5 and S5–S7 that the structure of the  $\beta$ -sheets, especially the  $\beta$ -sheets on

the C-terminal side ( $\beta_{III}$ – $\beta_V$ ) is almost completely disrupted. The structure of the N-terminal  $\beta$ -sheet ( $\beta_{II}$ ) is also somewhat disrupted. In contrast, the structure of the  $\alpha$ -helix is almost completely intact. These consistent results also indicate the high reproducibility of the LRPF/3D-RISM method for studying the mechanism of the denaturation by urea.

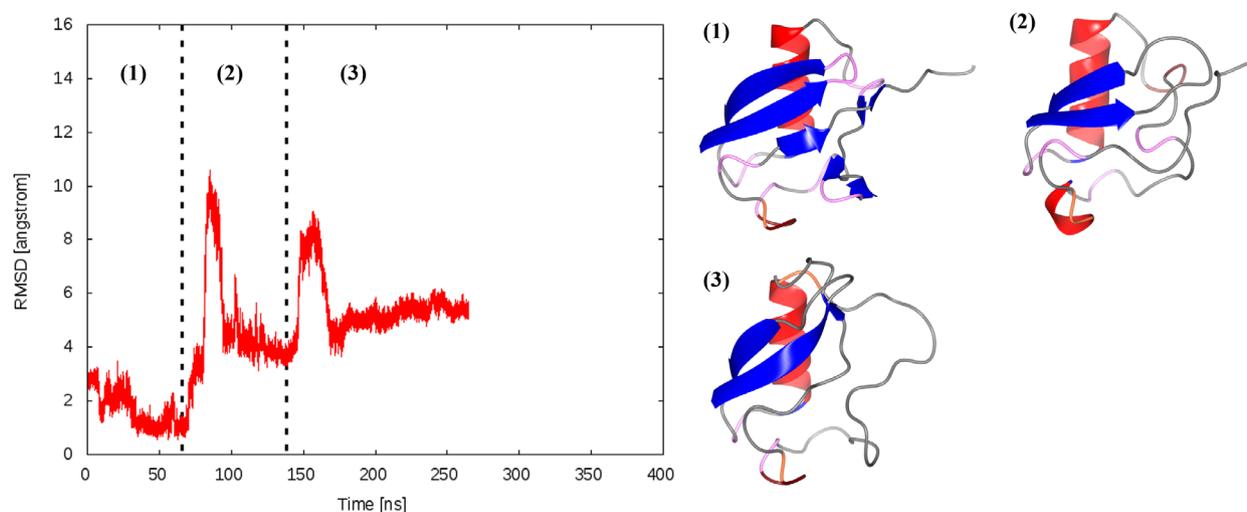
## 4.2 | Urea distribution

To examine the contribution of urea to conformational changes in the LRPF/3D-RISM simulations, the distribution functions of urea around the C-terminal  $\beta$ -sheet ( $\beta_V$ ; residue number: 64–72) of the averaged structure of the last 12 ns in the first and second cycles of the LRPF2 trajectory are depicted in Figures 6 and 7, respectively, because significant conformational changes of ubiquitin were observed in the second and third cycles of the LRPF2 trajectory (see Figure 2). The prominent peak of the urea oxygen is observed close to the amide hydrogen of Leu71 backbone in Figure 6 (see also Figure S8(a)). A slight distribution of urea hydrogen is also found around the carbonyl oxygen of Leu71. These distributions correspond to the hydrogen bonds between Leu71 backbone and urea, which are probably responsible for the increase of the RMSF value in the second cycle of the LRPF2 trajectory (the difference in the RMSF value between the first and second cycles ( $\Delta$ RMSF[1  $\rightarrow$  2]) of Leu71 is 0.6906 Å). A distribution of the urea hydrogen is also observed around the carbonyl oxygen of Arg72 backbone. In addition, the conspicuous peak of the urea oxygen is observed around the side chains of Arg42 and Arg72. These distributions may contribute to the increases of the RMSF values of these residues in the second cycle of the LRPF2 trajectory ( $\Delta$ RMSF[1  $\rightarrow$  2] of Arg42 is 0.5006 Å, and that of Arg72 is 0.5371 Å).

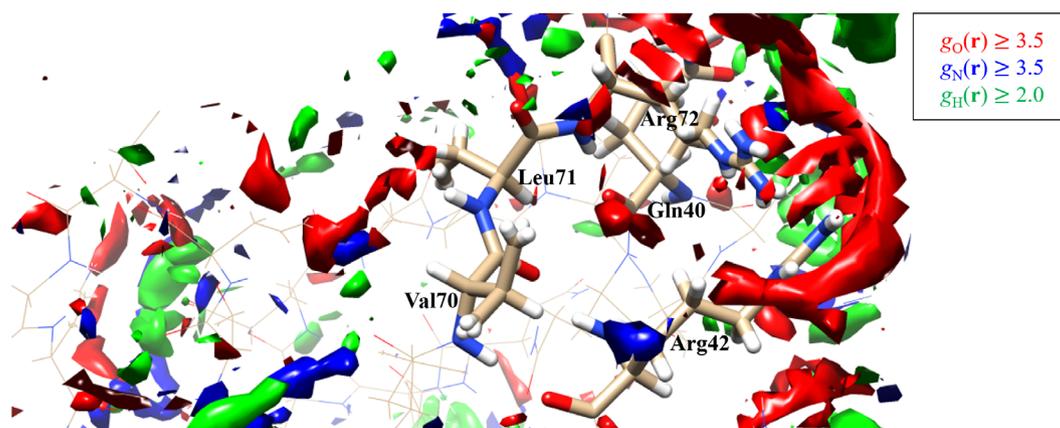
In Figure 7, the conspicuous peak of the urea oxygen is observed close to the amide hydrogens of Ser65 and Thr66 backbones (see also Figure S8(b)). The distribution of urea nitrogen is also observed around the amide hydrogen of Ser65. These hydrogen bonds between the backbone and urea may be responsible for the increases of the



**FIGURE 4** Time evolutions of the numbers of residues constituting the (a) helical and (b) sheet structures in the LRPF1 trajectory



**FIGURE 5** Time evolution of the RMSDs and the averaged structures of the last 12 ns in cycles of the LRPF1 trajectory. Vertical dashed lines represent cycles of the LRPF/3D-RISM simulation. The averaged structure corresponding to each cycle is shown on the right. Helical structures and  $\beta$ -sheet ones are represented in red and blue, respectively

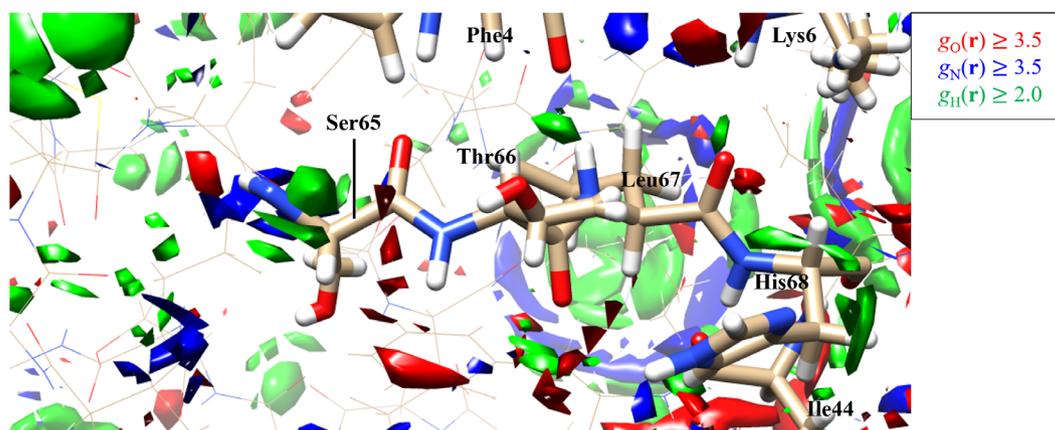


**FIGURE 6** Distribution functions of oxygen, nitrogen, and hydrogen atoms of urea around the C-terminal  $\beta$ -sheet ( $\beta_V$ ) of the averaged structure of the last 12 ns in the first cycle of the LRPF2 trajectory. Isosurfaces of the distribution functions  $g(r) \geq 3.5$  for the oxygen and nitrogen atoms, and  $g(r) \geq 2.0$  for the hydrogen atom are plotted in red, blue, and green, respectively. Residues constituting  $\beta_V$  (Val70, Leu71, and Arg72) and  $\beta_{III}$  (Gln40 and Arg42) are represented as stick models, whereas the other residues are represented as wire models

RMSF values in the third cycle of the LRPF2 trajectory (the differences in the RMSF values between the second and third cycles ( $\Delta$ RMSFs[2  $\rightarrow$  3]) of Ser65 and Thr66 are 0.5901 Å and 1.1017 Å, respectively). A distribution of urea nitrogen and oxygen is observed around the hydroxy groups of the side chains of Ser65 and Thr66, respectively. In addition, the distribution of urea hydrogen is also found close to the oxygen atom of the side chain of Thr66. These interactions between these residues and urea may also contribute to the increases of the RMSFs of Ser65 and Thr66. Peaks of the urea oxygen and hydrogen are observed around the side chains of His68, and these interactions may be responsible for the increase of the RMSF value of the residue in the third cycle of the LRPF2 trajectory ( $\Delta$ RMSF[2  $\rightarrow$  3] of His68 is 0.5867 Å).

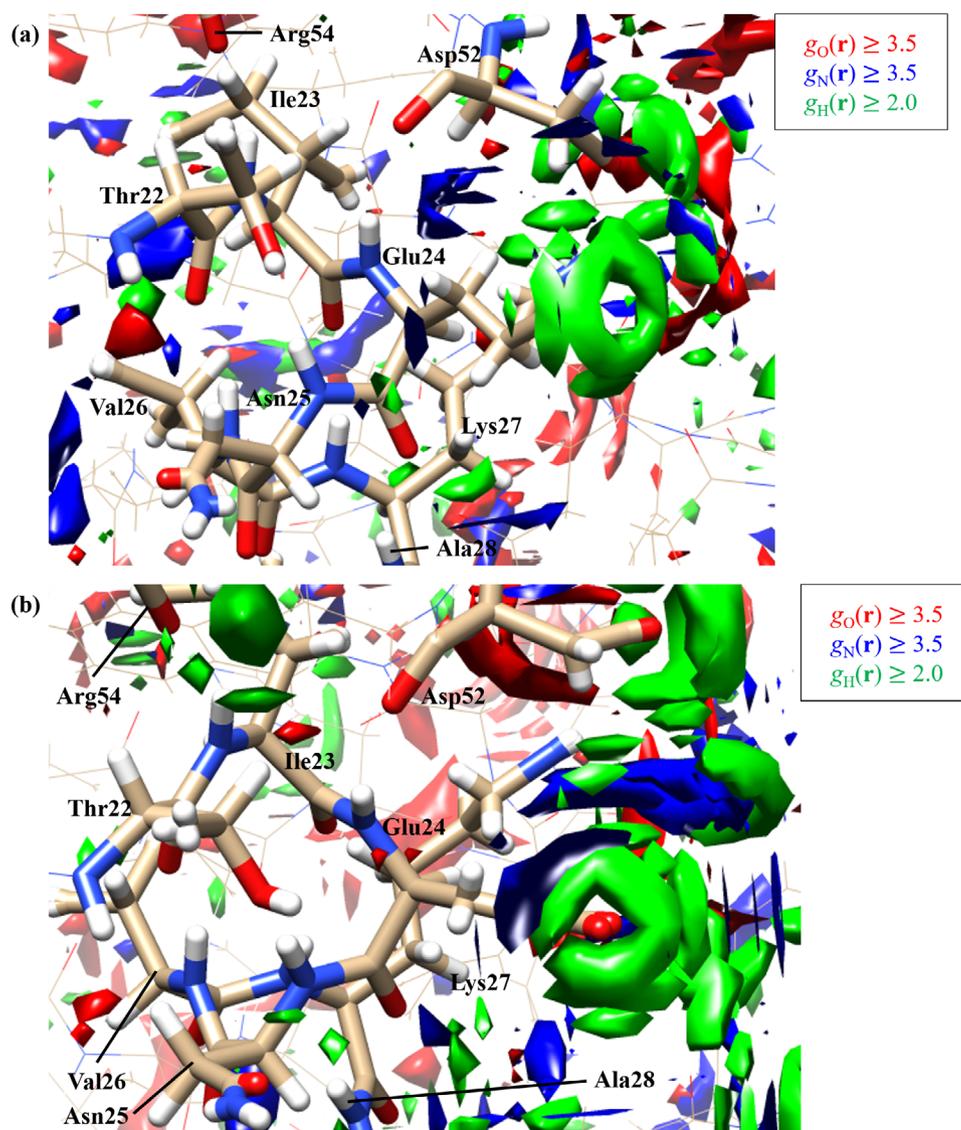
These observations that hydrogen bonds between the backbone and urea induce the denaturation correspond to the “direct mechanism” reported in previous studies.<sup>67–71</sup>

The distribution functions of urea around a part of the  $\alpha$ -helix (residue number: 23–34) in the averaged structure of the last 12 ns in the first and second cycles of the LRPF2 trajectory are also examined in Figure 8. The notable peaks of urea distributions are found around the side chains of the residues constituting the  $\alpha$ -helix, particularly that of Glu24. On the other hand, the amide atoms of the backbone form hydrogen bonds with other residues, and no prominent distribution of urea around the backbone atoms, as in the case of  $\beta$ -sheets, are observed. This result indicates that hydrogen bonds are unlikely to be formed between those atoms and urea, and may be related to the



**FIGURE 7** Distribution functions of oxygen, nitrogen, and hydrogen atoms of urea around the C-terminal  $\beta$ -sheet ( $\beta_V$ ) of the averaged structure of the last 12 ns in the second cycle of the LRPF2 trajectory. Residues constituting  $\beta_V$  (Ser65, Thr66, Leu67, and His68),  $\beta_I$  (Phe4 and Lys6), and  $\beta_{III}$  (Ile44) are represented as stick models, whereas the other residues are represented as wire models. The other representations are the same as those in Figure 6

**FIGURE 8** Distribution functions of oxygen, nitrogen, and hydrogen atoms of urea around a part of the  $\alpha$ -helix structure (residue number: 23–34) in the averaged structure of the last 12 ns in the (a) first and (b) second cycles of the LRPF2 trajectory. Isosurfaces of distribution functions  $g(r) \geq 3.5$  for the oxygen and nitrogen atoms, and  $g(r) \geq 2.0$  for the hydrogen atom are depicted in red, blue, and green, respectively. Residues constituting the  $\alpha$ -helix structure (Ile23, Glu24, Asn25, Val26, Lys27, and Ala28) and random-coil structures (Thr22, Asp52, and Arg54) are drawn in a stick representation, while the other residues are drawn in a wire representation



small fluctuations of the  $\alpha$ -helix compared with other parts (see Figure 3(b); the values of the  $\Delta\text{RMSF}(1 \rightarrow 2)$  and  $\Delta\text{RMSF}(2 \rightarrow 3)$  of the residues constituting the  $\alpha$ -helix were all less than 0.5 Å). These features are in good agreement with a previous study.<sup>67</sup>

As is clear from the above, the 3D-RISM theory provides distribution functions of solvent molecules that properly takes into account the detailed protein–solvent interactions such as hydrogen bonds, at the molecular level in a well-converged statistical ensemble of the solvent. This feature of the 3D-RISM theory enabled us to obtain an appropriate urea configuration and thus to describe the urea denaturation process in good agreement with the experimental results.

## 5 | CONCLUSIONS

In this study, we have proposed the LRP/3D-RISM method, which facilitates the calculation of conformational changes of proteins induced by varying the solution environment. The LRP method is available if only one of the two end conformations before and after the structural changes is known, and the 3D-RISM theory is easily applicable even for solutions of varying concentration or with cosolvents. Utilizing both the features, the LRP/3D-RISM method expands the scope of the application of the original LRP method and describes the dynamics of proteins in more complex solution environments. We applied the method to the urea-induced denaturation process of ubiquitin. Four out of five LRP/3D-RISM trajectories successfully simulated the early stage of the denaturation process within the simulation time of 300 ns, while standard MD simulations for the water/urea mixed solution system did not show significant conformational changes within the simulation time of 1  $\mu\text{s}$ . The obtained LRP/3D-RISM trajectories consistently reproduced the mechanism of the urea-induced denaturation of ubiquitin reported in previous studies, in which  $\beta$ -sheets preferentially unfold in an aqueous urea solution. These results indicate that the proposed method opens up the possibility for gaining detailed understanding of the mechanism of protein conformational changes induced by various solvent species.

Since a ligand should be treated as a “solvent” species in the present LRP/3D-RISM method, it is limited to small molecules like urea that can be treated as ligands. However, by combining with the solute–solute 3D-RISM or fragment-based ligand binding methods, it may be possible to apply the LRP/3D-RISM method to larger ligands.<sup>72,73</sup>

## ACKNOWLEDGMENTS

The authors are grateful for the financial support from the Japan Society for the Promotion of Science (JSPS) KAKENHI (Grant nos. 18K05036 and 19H02677). NY acknowledges the Toyota Riken Scholar from the Toyota Physical and Chemical Research Institute. ST acknowledges the financial support of JSPS Grants-in-Aid for JSPS Fellows (Grant no. 19J11525). Numerical calculations were conducted, in part, at the Research Center for Computational Science, Institute for Molecular Science, National Institutes of Natural Sciences. Molecular graphics were depicted with the UCSF Chimera,

developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco.<sup>66</sup>

## ORCID

Shoichi Tanimoto  <https://orcid.org/0000-0001-6267-085X>

Koichi Tamura  <https://orcid.org/0000-0002-9472-7555>

Shigehiko Hayashi  <https://orcid.org/0000-0003-2658-3171>

Norio Yoshida  <https://orcid.org/0000-0002-2023-7254>

Haruyuki Nakano  <https://orcid.org/0000-0002-7008-0312>

## REFERENCES

- [1] C. B. Anfinsen, *Science* **1973**, 181, 223.
- [2] J. Payandeh, T. Scheuer, N. Zheng, W. A. Catterall, *Nature* **2011**, 475, 353.
- [3] K. Henzler-Wildman, D. Kern, *Nature* **2007**, 450, 964.
- [4] J. Kästner, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, 1, 932.
- [5] J. Schlitter, M. Engels, P. Krüger, E. Jacoby, A. Wollmer, *Mol. Simul.* **1993**, 10, 291.
- [6] B. Isralewitz, M. Gao, K. Schulten, *Curr. Opin. Struct. Biol.* **2001**, 11, 224.
- [7] D. Hamelberg, J. Mongan, J. A. McCammon, *J. Chem. Phys.* **2004**, 120, 11919.
- [8] N. Nakajima, H. Nakamura, A. Kidera, *J. Phys. Chem. B* **1997**, 101, 817.
- [9] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99, 12562.
- [10] Y. Sugita, Y. Okamoto, *Chem. Phys. Lett.* **1999**, 314, 141.
- [11] M. Ikeguchi, J. Ueno, M. Sato, A. Kidera, *Phys. Rev. Lett.* **2005**, 94, 1.
- [12] C. Groß, K. Hamacher, K. Schmitz, S. Jäger, *J. Chem. Inf. Model.* **2017**, 57, 243.
- [13] C. Atilgan, A. R. Atilgan, *PLoS Comput. Biol.* **2009**, 5, e1000544.
- [14] F. Jalalpour, O. Sensoy, C. Atilgan, *J. Chem. Theory Comput.* **2020**, 16, 3825.
- [15] B. Erman, *Proteins* **2018**, 86, 1001.
- [16] M. Tiberti, A. Pandini, F. Fraternali, A. Fornili, *Bioinformatics* **2018**, 34, 207.
- [17] M. Yu, Y. Chen, Z. Le Wang, Z. Liu, *Phys. Chem. Chem. Phys.* **2019**, 21, 5200.
- [18] N. Matthews, A. Kitao, S. Laycock, S. Hayward, *J. Chem. Inf. Model.* **2019**, 59, 2900.
- [19] K. Tamura, S. Hayashi, *J. Chem. Theory Comput.* **2015**, 11, 2900.
- [20] K. Tamura, S. Hayashi, *PLoS One* **2017**, 12, e0181489.
- [21] J. J. Ruprecht, M. S. King, T. Zögg, A. A. Aleksandrova, E. Pardon, P. G. Crichton, J. Steyaert, E. R. S. Kunji, *Cell* **2019**, 176, 435.
- [22] D. Beglov, B. Roux, *J. Phys. Chem. B* **1997**, 101, 7821.
- [23] A. Kovalenko, F. Hirata, *Chem. Phys. Lett.* **1998**, 290, 237.
- [24] F. Hirata, *Molecular Theory of Solvation*, Kluwer Academic Publishers, Dordrecht **2004**.
- [25] J.-P. Hansen, I. R. McDonald, *Theory of Simple Liquids*, 2nd ed., Academic Press, Inc., San Diego **1990**.
- [26] N. Yoshida, *J. Chem. Inf. Model.* **2017**, 57, 2646.
- [27] N. Yoshida, S. Phongphanphanee, Y. Maruyama, T. Imai, F. Hirata, *J. Am. Chem. Soc.* **2006**, 128, 12042.
- [28] T. Imai, R. Hiraoka, A. Kovalenko, F. Hirata, *J. Am. Chem. Soc.* **2005**, 127, 15334.
- [29] B. Kim, F. Hirata, *J. Chem. Phys.* **2013**, 138, 054108.
- [30] F. Hirata, K. Akasaka, *J. Chem. Phys.* **2015**, 142, 044110.
- [31] F. Hirata, B. Kim, *J. Mol. Liq.* **2016**, 217, 23.
- [32] F. Hirata, *J. Chem. Phys.* **2016**, 145, 234106.
- [33] F. Hirata, *J. Mol. Liq.* **2018**, 270, 218.
- [34] F. Hirata, M. Sugita, M. Yoshida, K. Akasaka, *J. Chem. Phys.* **2018**, 148, 020901.
- [35] H. Mori, *Phys. Rev.* **1958**, 112, 1829.

- [36] H. Mori, *Prog. Theor. Phys.* **1962**, 28, 763.
- [37] H. Mori, *Prog. Theor. Phys.* **1965**, 33, 423.
- [38] H. Mori, *Prog. Theor. Phys.* **1965**, 34, 399.
- [39] U. Balucani, M. Zoppi, *Dynamics of the Liquid State*, Clarendon Press, Oxford **1994**.
- [40] A. Kovalenko, S. Ten-no, F. Hirata, *J. Comput. Chem.* **1999**, 20, 928.
- [41] A. Kovalenko, F. Hirata, *J. Chem. Phys.* **1999**, 110, 10095.
- [42] A. Kovalenko, F. Hirata, *J. Chem. Phys.* **2000**, 112, 10391.
- [43] A. Kovalenko, F. Hirata, *J. Chem. Phys.* **2000**, 112, 10403.
- [44] A. Kovalenko, F. Hirata, *J. Phys. Chem. B* **1999**, 103, 7942.
- [45] S. Gusarov, T. Ziegler, A. Kovalenko, *J. Phys. Chem. A* **2006**, 110, 6083.
- [46] N. Yoshida, F. Hirata, *J. Comput. Chem.* **2006**, 27, 453.
- [47] S. Vijay-kumar, C. E. Bugg, W. J. Cook, *J. Mol. Biol.* **1987**, 194, 531.
- [48] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, 26, 1781.
- [49] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. De Groot, H. Grubmüller, A. D. MacKerell Jr., *Nat. Methods* **2016**, 14, 71.
- [50] A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, 102, 3586.
- [51] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, 79, 926.
- [52] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, 14, 33.
- [53] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, 98, 10089.
- [54] S. Miyamoto, P. A. Kollman, *J. Comput. Chem.* **1992**, 13, 952.
- [55] H. C. Andersen, *J. Comput. Phys.* **1983**, 52, 24.
- [56] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **1984**, 81, 3684.
- [57] M. Mandal, C. Mukhopadhyay, *Phys. Chem. Chem. Phys.* **2014**, 16, 21706.
- [58] Y. Maruyama, N. Yoshida, H. Tadano, D. Takahashi, M. Sato, F. Hirata, *J. Comput. Chem.* **2014**, 35, 1347.
- [59] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. MacKerell Jr., *J. Comput. Chem.* **2010**, 31, 671.
- [60] W. Yu, X. He, K. Vanommeslaeghe, A. D. MacKerell Jr., *J. Comput. Chem.* **2012**, 33, 2451.
- [61] D. J. Sindhikara, N. Yoshida, F. Hirata, *J. Comput. Chem.* **2012**, 33, 1536.
- [62] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, *Nature Methods*, **2020**, 7, 261.
- [63] D. R. Roe, T. E. Cheatham III., *J. Chem. Theory Comput.* **2013**, 9, 3084.
- [64] W. Kabsch, C. Sander, *Biopolymers* **1983**, 22, 2577.
- [65] S. McNicholas, E. Potterton, K. S. Wilson, M. E. M. Noble, *Acta Cryst.* **2011**, D67, 386.
- [66] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, 25, 1605.
- [67] S. Parui, B. Jana, *J. Phys. Chem. B* **2019**, 123, 8889.
- [68] C. Camilloni, A. Guerini Rocco, I. Eberini, E. Gianazza, R. A. Broglia, G. Tiana, *Biophys. J.* **2008**, 94, 4654.
- [69] E. P. O'Brien, R. I. Dima, B. Brooks, D. Thirumalai, *J. Am. Chem. Soc.* **2007**, 129, 7346.
- [70] L. Hua, R. Zhou, D. Thirumalai, B. J. Berne, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105, 16928.
- [71] W. K. Lim, J. Rösgen, S. W. Englander, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, 106, 2595.
- [72] Y. Kiyota, N. Yoshida, F. Hirata, *J. Chem. Theory Comput.* **2011**, 7, 3803.
- [73] T. Imai, K. Oda, A. Kovalenko, F. Hirata, A. Kidera, *J. Am. Chem. Soc.* **2009**, 131, 12430.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Tanimoto S, Tamura K, Hayashi S, Yoshida N, Nakano H. A computational method to simulate global conformational changes of proteins induced by cosolvent. *J Comput Chem.* 2021;42:552–563. <https://doi.org/10.1002/jcc.26481>