Two-step relaxation mode analysis with multiple evolution times applied to all-atom molecular dynamics protein simulation

N. Karasawa,* A. Mitsutake,† and H. Takano‡
Department of Physics, Faculty of Science and Technology, Keio University, Yokohama, Kanagawa 223-8522, Japan
(Received 22 May 2017; revised manuscript received 19 August 2017; published 15 December 2017)

Proteins implement their functionalities when folded into specific three-dimensional structures, and their functions are related to the protein structures and dynamics. Previously, we applied a relaxation mode analysis (RMA) method to protein systems; this method approximately estimates the slow relaxation modes and times via simulation and enables investigation of the dynamic properties underlying the protein structural fluctuations. Recently, two-step RMA with multiple evolution times has been proposed and applied to a slightly complex homopolymer system, i.e., a single [n]polycatenane. This method can be applied to more complex heteropolymer systems, i.e., protein systems, to estimate the relaxation modes and times more accurately. In two-step RMA, we first perform RMA and obtain rough estimates of the relaxation modes and times. Then, we apply RMA with multiple evolution times to a small number of the slowest relaxation modes obtained in the previous calculation. Herein, we apply this method to the results of principal component analysis (PCA). First, PCA is applied to a 2-μs molecular dynamics simulation of hen egg-white lysozyme in aqueous solution. Then, the two-step RMA method with multiple evolution times is applied to the obtained principal components. The slow relaxation modes and corresponding relaxation times for the principal components are much improved by the second RMA.

DOI: 10.1103/PhysRevE.96.062408

I. INTRODUCTION

Biopolymers have flexible structures and various functions, which are derived from both the biopolymer structures and the structural fluctuation dynamics. Therefore, knowledge of the dynamic properties of the structural fluctuations of a biopolymer is important for understanding the interrelation between its movement and functions. Thus, many methods to analyze the dynamics and kinetics of protein simulations have been developed [1–31]. In particular, the Markov state model has been presented (see Refs. [1–7,9,15,18,19,21–23,28–31] and references cited therein) and applied to many protein systems.

Further, relaxation mode analysis (RMA) has been developed to aid investigation of the dynamic properties of spin systems [32] and homopolymer systems [33,34]. In addition, this technique has been applied to various polymer systems [35–37] in order to investigate their slow relaxation dynamics [38,39]. Recently, RMA has also been applied to biomolecular systems [13,16,24].

The relaxation modes $\{X_p\}$ satisfy

$$\langle X_p(t)X_q(0)\rangle = \delta_{p,q}e^{-\lambda_p t}. \quad (1)$$

Here, $\langle A(t)B(0)\rangle$ denotes the equilibrium correlation of A at time t and B at time 0:

$$\langle A(t)B(0)\rangle = \sum_{Q,Q'} A(Q)T_t(Q|Q')B(Q')P_{\rm eq}(Q'), \quad (2)$$

where $T_t(Q|Q')$ is the conditional probability that the system is in state Q at time t, given that it is in state Q' at time $t = 0$. Further, $P_{\rm eq}(Q')$ denotes the probability that the system is in state Q' at equilibrium. The relaxation rate of X_p is denoted by λ_p. The relaxation modes and rates are given as left eigenfunctions and eigenvalues of the time-evolution operator of the master equation of the system, respectively [24,33,34]. RMA approximately estimates the slow relaxation modes and rates from a simulation and decomposes the structural fluctuations into the slow relaxation modes, which characterize the slow relaxation dynamics of the system.

In conventional RMA, $\{\lambda_p\}$ and $\{X_p\}$ are estimated by solving the generalized eigenvalue problem of the time correlation matrices of coordinates for two different times, $C(t_0+\tau)$ and $C(t_0)$ (see Sec. II). Recently, dynamical analysis methods for molecular simulations of biopolymer systems have been developed to investigate slow dynamics. In these techniques, which include time-structure based independent component analysis (tICA) [14,17,19], time-lagged independent component analysis (TICA) [18], and dynamic component analysis (DCA) [25,27], time correlation matrices of certain physical quantities or states are used. (Note that tICA, TICA, and DCA are mathematically equivalent [18,25], and tICA is a special case of RMA with $t_0 = 0$. See Refs. [13,14] for more details on the differences between tICA and RMA. Further, the relationships between the Markov state model, tICA, TICA, and RMA are explained in Refs. [18,19,24]. From the perspective that the relaxation modes and rates in RMA are given as left eigenfunctions and eigenvalues of the time-evolution operator of the master equation of the system, respectively, RMA is related to the Markov state model. The extension of a regular Markov state model by introducing t_0, which is referred to as Markov state relaxation mode analysis, is explained in Ref. [24].) In tICA, TICA, and DCA, the time correlation functions $C(\tau)$ and $C(0)$ are used.

In practice, the relaxation times obtained from the dynamical analysis methods explained in the previous paragraph depend on the selection method for the physical quantities (or states) for the time correlation functions and the values of the τ parameter [2,18,24]. When physical quantities (or states) with slow behaviors are used to construct the time correlation

*nkarasawa@rk.phys.keio.ac.jp
†ayori@rk.phys.keio.ac.jp
‡takano@rk.phys.keio.ac.jp

functions, the slow relaxation behavior of the simulation can be investigated. The difference between RMA and the other dynamical analysis methods is introduction of the evolution time t_0; however, introduction of this parameter increases the difficulty in solving the generalized eigenvalue problem. Despite this, RMA has an advantage in that the obtained slow relaxation modes and times are improved by choosing the appropriate τ and t_0 (two adjusting parameters).

RMA requires relatively high statistical precision of the time correlation matrices because the generalized eigenvalue problem is treated; thus, it is difficult for RMA to handle a large number of degrees of freedom directly. We must reduce the number of degrees of freedom automatically. Therefore, to reduce the degrees of freedom, we previously proposed principal component RMA (PCRMA) using two evolution times [16]. In the proposed technique, principal component analysis (PCA) is first implemented. Then, RMA is applied to a small number of principal components with large fluctuations. (Note that PCRMA using all principal components is mathematically equivalent to RMA.) This method can systematically reduce the degrees of freedom. Recently, we also developed a different RMA, which is referred to as "two-step RMA with multiple evolution times" [40]. We have applied two-step RMA using multiple evolution times to a single [n]polycatenane, which is a homopolymer system consisting of n ring polymers topologically interlocked with neighboring rings, in order to investigate its dynamics. This technique uses a similar approach to PCRMA.

In this paper, we have applied this method to a protein, i.e., a complex heteropolymer system. The effectiveness of the method was first demonstrated for a single homopolymer system, as mentioned above. However, its effectiveness for a protein system is not trivial at all because the structural fluctuations of proteins are considerably more complex than those of single homopolymers. A process chart of the method is shown in Fig. 1(a). In this method, RMA with a single evolution time using small t_0 and τ is first implemented and the relaxation modes and times are roughly estimated. Then, we again apply a second RMA to a small number of the obtained slowest relaxation modes. This approach not only reduces the number of degrees of freedom automatically, but also uses the physical quantities that exhibit slow relaxations to construct time correlation functions in the second RMA. For the second RMA, we also use the recently presented RMA with multiple evolution times [40] because the relaxation modes obtained from the first RMA each have different relaxation times. The evolution times for the second RMA can be estimated from the relaxation times obtained from the first RMA. Hence, we can automatically determine the evolution times. Using this method, the estimation accuracy of the relaxation modes and times can be improved.

In this paper, we briefly describe the method for the case of a molecular dynamics (MD) simulation satisfying the detailed balance condition and considering position coordinates only (for details, see Ref. [24]). In order to compare the results of the present method with those yielded by the PCRMA [16], we apply this two-step RMA method to PCA results. First, PCA of a 2-μs MD simulation of hen egg-white lysozyme in aqueous solution is conducted. Then, the two-step RMA method with multiple evolution times is applied to the obtained principal components. [This means that, in Fig. 1(a), $\boldsymbol{R}$ is replaced with principal components, i.e., the $\boldsymbol{\Phi}_c$ of Eq. (24) in this work.] The slow relaxation modes and corresponding relaxation times for the principal components are much improved following use of the second RMA.

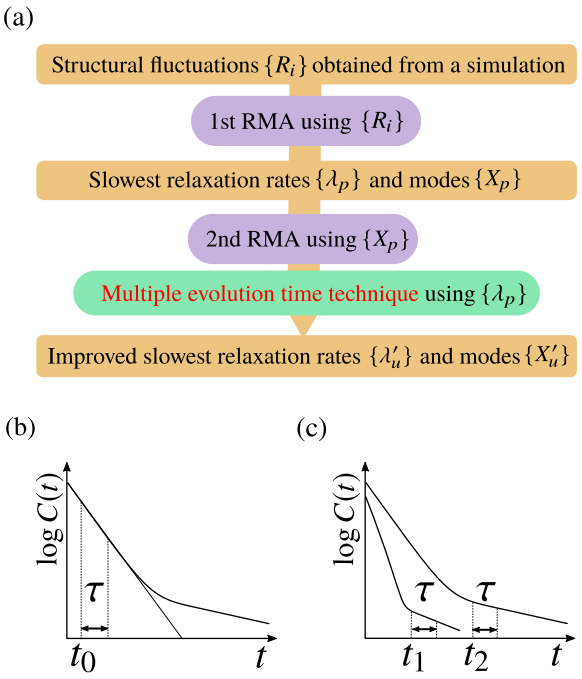

FIG. 1. (a) Process chart of two-step RMA with multiple evolution times. (b), (c) Schematics showing t_0, τ, and t_i selection using semilog plot of time-displaced autocorrelation function $C(t)$ versus t.

II. METHODS

A. Conventional RMA

We consider a biopolymer composed of N atoms. Further, we assume that $\boldsymbol{R}$ is a $3N$-dimensional column vector that consists of a set of atomic coordinates relative to their average coordinates

$$\boldsymbol{R}^{\mathrm{T}} = \left(\boldsymbol{r}_1'^{\mathrm{T}}, \boldsymbol{r}_2'^{\mathrm{T}}, \ldots, \boldsymbol{r}_N'^{\mathrm{T}}\right) = (x_1', y_1', z_1', \ldots, x_N', y_N', z_N'), \quad (3)$$

with

$$\boldsymbol{r}_i' = \boldsymbol{r}_i - \langle \boldsymbol{r}_i \rangle, \quad (4)$$

where $\boldsymbol{r}_i$ is the coordinate of the ith atom of the biopolymer in the center-of-mass coordinate system and $\langle \boldsymbol{r}_i \rangle$ is its average coordinate. In conventional RMA, X_p is approximated by a trial function, which is constructed as a linear combination of relevant physical quantities that are time evolved for $t_0/2$:

$$X_p(Q) = \sum_{i=1}^{3N} f_{p,i} R_i(t_0/2; Q) \quad (5)$$

with

$$R_i(t; Q) = \sum_{Q'} R_i(Q') T_t(Q'|Q). \quad (6)$$

Here, $R_i(Q)$ is the ith component of $\boldsymbol{R}$ in state Q. The quantity $R_i(t; Q)$ is the expectation value of R_i after a period t, starting

from state Q. The evolution time $t_0/2$ is introduced to reduce the relative weight of the faster modes contained in R_i, and to estimate slow relaxation times with greater precision.

For this trial function, we consider the variational problem:

$$\delta\mathcal{R} = 0, \tag{7}$$

with

$$\mathcal{R}[X_p] = \frac{\langle X_p(\tau)X_p(0)\rangle}{\langle X_p(0)X_p(0)\rangle}, \tag{8}$$

where the stationary value of $\mathcal{R}$ gives the value $\exp(-\lambda_p\tau)$. The variational problem becomes a generalized eigenvalue problem of the time correlation matrices of the physical quantities for two different times, t_0 and $t_0+\tau$. In practice, the time correlation matrices for the two different times are calculated through simulations. Then, by solving the generalized eigenvalue problem, the $\{\lambda_p\}$ and $\{X_p\}$ are obtained from the eigenvalues and eigenvectors, respectively. In order to examine the validity of the present analysis, the autocorrelation functions $C_{i,i}(t)$ are reproduced from the estimated eigenvalues and eigenvectors and compared with those directly calculated via simulation. We refer to this method as the "RMA method with a single evolution time," which is $t_0/2$.

The relaxation times $\{1/\lambda_p\}$ and the $\{X_p\}$ obtained via RMA depend on the manner in which t_0 and τ are selected in practice. For simple understanding, we consider the case of one physical quantity R. From the variational problem of Eqs. (7) and (8), the relaxation time $1/\lambda$ is obtained from the gradient of the straight line connecting two points at $t = t_0$ and $t_0+\tau$ in the semilog plot of the correlation function $C(t) = \langle R(t)R(0)\rangle - \langle R\rangle^2$ versus t, as shown in Fig. 1(b). If the time correlation function of the physical quantity contains several $\{1/\lambda_p\}$, and if we choose $t_0 = 0$ (tICA case) or small t_0 and small τ as shown in Fig. 1(b), the obtained $1/\lambda$ does not correspond to the slow relaxation behavior of $\log C(t)$ at long times. To investigate the slow relaxation, we wish to choose values of t_0 and τ that are as large as possible. On the other hand, the choice of longer t_0 and τ is also limited because of the decreasing accuracy of the time correlation function at long times. Therefore, we must choose the appropriate t_0 and τ.

When the relevant physical quantities $\{R_i\}$ in the trial function exhibit different relaxations, it is preferable to use different evolution times for the different physical quantities, as shown in Fig. 1(c). That is, if we know the characteristic time scales of the relevant physical quantities, we can choose a specific evolution time t_i for each relevant physical quantity R_i based on its characteristic time scale. This RMA method is referred to as the "RMA with multiple evolution times," and we describe its formulation in the next section.

B. RMA with multiple evolution times

In this section, the RMA with multiple evolution times $\{t_i/2\}$ is explained. We use the following function as an approximate relaxation mode:

$$X_p(Q) = \sum_{i=1}^{3N} f_{p,i}\, R_i(t_i/2; Q). \tag{9}$$

The parameter t_i is introduced in order to reduce the relative weight of the faster modes contained in R_i. Further, it is expected that Eq. (9) yields a superior approximation for larger t_i.

We now consider the variational problem of Eqs. (7) and (8). Then, the variational problem becomes a generalized eigenvalue problem:

$$\sum_{j=1}^{3N} C_{i,j}\left(\frac{t_i+t_j}{2}+\tau\right) f_{p,j} = \exp(-\lambda_p\tau)\sum_{j=1}^{3N} C_{i,j}\left(\frac{t_i+t_j}{2}\right) f_{p,j}. \tag{10}$$

Here, $C_{i,j}(t) = \langle R_i(t)R_j(0)\rangle$ and the orthonormal condition (8) for X_p is expressed as

$$\sum_{i=1}^{3N}\sum_{j=1}^{3N} f_{p,i}C_{i,j}\left(\frac{t_i+t_j}{2}\right) f_{q,j} = \delta_{p,q}. \tag{11}$$

Equations (9), (10), and (11) determine the relaxation rates λ_p and the corresponding relaxation modes. We choose the indices of λ_p such that $0 < \lambda_1 \leqslant \lambda_2 \leqslant \cdots$ holds. The inverse transformation of Eq. (9) is given by

$$R_i(t_i/2; Q) = \sum_{p=1}^{3N-6} g_{i,p}X_p(Q) \tag{12}$$

with

$$g_{i,p} = \sum_{j=1}^{3N} C_{i,j}\left(\frac{t_i+t_j}{2}\right) f_{p,j}. \tag{13}$$

Note that the number of meaningful relaxation modes is $3N-6$ because we remove the translational and rotational degrees of freedom in the calculation of $\{\langle \boldsymbol{r}_i\rangle\}$ [13]. The time correlation functions of R_i are given by

$$\begin{aligned}\langle R_i(t)R_j(0)\rangle &= \sum_p\sum_q g_{i,p}g_{j,q}\left\langle X_p\left(t-\frac{t_i+t_j}{2}\right)X_q(0)\right\rangle \\ &\simeq \sum_p g_{i,p}g_{j,p}\exp\left[-\lambda_p\left(t-\frac{t_i+t_j}{2}\right)\right] \\ &= \sum_p \tilde{g}_{i,p}\tilde{g}_{j,p}\exp(-\lambda_p t),\end{aligned} \tag{14}$$

for $t \geqslant (t_i+t_j)/2$. Here,

$$\tilde{g}_{i,p} = g_{i,p}\exp(\lambda_p t_i/2). \tag{15}$$

Because we are considering position coordinates only, the detailed balance condition [41] yields the following consequences: $C(t)$ is a symmetric matrix, $C_{i,j}(t) = C_{j,i}(t)$; the $\{\lambda_p\}$ are real and positive, which corresponds to pure relaxation.

C. Two-step RMA with multiple evolution times

In this technique, RMA with a single evolution time using small t_0 and τ is first implemented and the $\{X_p\}$ and $\{\lambda_p\}$ are roughly estimated. Note that all the $\{t_i\}$ in the previous subsection are set to $t_0/2$. Then, we apply the second RMA

to a small number of the obtained slowest $\{X_p\}$. We denote the number of $\{X_p\}$ used in the second RMA as N_m. In the second RMA, we use the previously presented RMA with multiple evolution times technique because the characteristic time scales of the $\{X_p\}$ obtained from the first RMA are known to correspond to their $\{1/\lambda_p\}$. We use the following trial function:

$$X'_u(Q) = \sum_{p=1}^{N_\mathrm{m}} f'_{u,p} X_p(t'_p/2; Q). \tag{16}$$

Here, $X_p(Q)$ is the relaxation mode obtained from the first RMA. Then, the generalized eigenvalue problem (10) becomes

$$\sum_{q=1}^{N_\mathrm{m}} C'_{p,q}\left(\frac{t'_p + t'_q}{2} + \tau'\right) f'_{u,q} = \exp(-\lambda'_u \tau') \sum_{q=1}^{N_\mathrm{m}} C'_{p,q}\left(\frac{t'_p + t'_q}{2}\right) f'_{u,q}, \tag{17}$$

where

$$C'_{p,q}(t) = \langle X_p(t) X_q(0)\rangle = \sum_{i=1}^{3N}\sum_{i=j}^{3N} f_{p,i} C_{i,j}(t_0 + t) f_{q,j} \tag{18}$$

is the time correlation matrix of the relaxation modes obtained from the first RMA. The orthonormal condition (11) becomes

$$\sum_{p=1}^{N_\mathrm{m}}\sum_{q=1}^{N_\mathrm{m}} f'_{u,p} C'_{p,q}\left(\frac{t'_p + t'_q}{2}\right) f'_{v,q} = \delta_{u,v}. \tag{19}$$

The original time correlation matrix $C_{i,j}(t)$ is reconstructed from the results of the second RMA as

$$C_{i,j}(t) \simeq \sum_{u=1}^{N_\mathrm{m}} \tilde{\gamma}_{i,u}\tilde{\gamma}_{j,u} \exp(-\lambda'_u t), \tag{20}$$

with

$$\tilde{\gamma}_{i,u} = \sum_{p=1}^{N_\mathrm{m}} \exp[\lambda'_u (t_0 + t'_p)/2] g_{i,p} g'_{p,u}, \tag{21}$$

where

$$g_{i,p} = \sum_{j=1}^{3N} C_{i,j}(t_0) f_{p,j} \tag{22}$$

and

$$g'_{p,u} = \sum_{q=1}^{N_\mathrm{m}} C'_{p,q}\left(\frac{t'_p + t'_q}{2}\right) f'_{u,q}. \tag{23}$$

The evolution times t'_p for the second RMA are estimated from the $\{1/\lambda_p\}$ obtained from the first RMA. The relaxation modes $\{X'_u\}$ and times $\{1/\lambda'_u\}$ are improved in the second RMA.

III. COMPUTATIONAL DETAILS

In this study, an MD simulation is performed using the AMBER package (AMBER 14.0) with GPU, along with the ff14SB force field and TIP3P model [42]. A hen egg-white lysozyme, which consists of 129 amino acid residues, is considered. The lysozyme [Protein Data Bank (PDB) identification (ID): 6LYZ [43]] is solvated with a 10-Å buffer of TIP3P water around the protein in each direction. The numbers of atoms in the lysozyme and water molecules are 1960 and 28 923 (9641 water molecules), respectively. Eight chloride ions (Cl^-) are also included in the system, yielding a net-neutral system. Thus, the total number of atoms in the system is 30 891. After energy minimization and heating, with equilibration at a constant pressure (1 atm) and at 298.15 K, a 2-μs MD simulation is performed at 298.15 K. A 2-fs time step is employed. Further, for the production run, a Berendsen thermostat is used to generate a constant-pressure, constant-temperature (NPT) ensemble. The cutoff is 10 Å, which is used to limit the direct space sum for the Particle Mesh Ewald (PME) method of AMBER. For the equilibration and production run, the pmemd code with GPU for MD simulations is used [44–46]. For analysis, the coordinates are saved in 1-ps intervals. The number of samples is 2×10^6. We use the coordinates of the heavy atoms, and the number of the degrees of freedom is 3003 (=1001 × 3). While a few months are required to implement several microsecond MD simulations in aqueous solution on GPU, the simulation time is shorter here because the large conformational changes of the protein systems occur on the millisecond time scale. However, many MD simulations of protein systems are conducted on the microsecond time scale and the rare events occurring during the limited simulations are interesting. In the authors' experience, even when RMA is applied to a limited-time simulation, the rare events occurring during the simulation are extracted. In the simulation, a few conformational changes are observed. Thus, we can show the effectiveness of the improved analysis method in the simulation system.

The first RMA is applied to the principal components with large fluctuations $\boldsymbol{\Phi}_\mathrm{c}$ reported in Ref. [13], instead of $\boldsymbol{R}$. Here, $\boldsymbol{\Phi}_\mathrm{c} = (\Phi_1, \Phi_2, \ldots, \Phi_{N_\mathrm{c}})^\mathrm{T}$ and

$$\Phi_n = \boldsymbol{F}_n^\mathrm{T}\boldsymbol{R} = \sum_{i=1}^{3N} F_{n,i} R_i, \tag{24}$$

where $\boldsymbol{F}_n$ is the orthonormal eigenvector of the correlation matrix ($\langle R_i R_j\rangle$), with the eigenvalue Λ_n obtained via PCA. We set the eigenvalue indices such that the relation $\Lambda_1 \geqslant \Lambda_2 \geqslant \cdots$ holds. Note that $C(t)$ in the previous section now represents the $N_\mathrm{c} \times N_\mathrm{c}$ time correlation matrix $\langle \boldsymbol{\Phi}_\mathrm{c}(t)\boldsymbol{\Phi}_\mathrm{c}^\mathrm{T}(0)\rangle$.

IV. RESULTS AND DISCUSSION

After removing the translational and rotational motions from the coordinates [47,48], PCA is first implemented for the heavy atoms. The number of total modes yielded by the PCA is 2997 (=3003 − 6). Further, the relative contributions of the variances of the first 10, 30, and 100 principal components to the total variance are approximately 30%, 60%, and 80%, respectively. The normalized time-displaced autocorrelation functions of the principal components with the ninth-largest fluctuations are shown in Fig. 2. Note that each of the principal components includes different relaxation processes. In particular, for a short time region, we observe fast relaxation processes, and the principal component with the largest fluctuation does not correspond to the mode with the

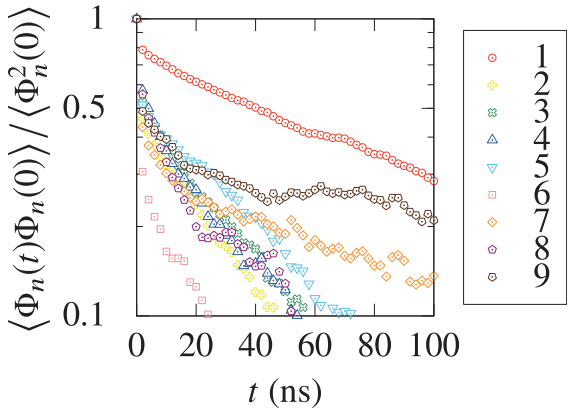

FIG. 2. Normalized time-displaced autocorrelation functions of $\Phi_n(t)$, $\langle\Phi_n(t)\Phi_n(0)\rangle/\langle\Phi_n^2(0)\rangle$ $(n = 1, \ldots, 9)$, obtained via PCA.

slowest relaxation. From Fig. 2, the seventh and ninth principal components seem to have slow relaxations at long times.

The first RMA is applied to N_c principal components with larger fluctuations, with $\mathbf{\Phi}_c$ in place of $\boldsymbol{R}$ in Eq. (9), where $N_c = 100$. The number of degrees of freedom is then reduced from $3N - 6$ to N_c. Here, we set $t_0 = 0$ ps and $\tau = 700$ ps for the first RMA (this corresponds to tICA because $t_0 = 0$). The $\{X_p\}$ and $\{1/\lambda_p\}$ are obtained from the generalized eigenvalue problem of principal components $\mathbf{\Phi}_c$ (instead of $\boldsymbol{R}$) of Eqs. (10) and (11), where all evolution times $\{t_i/2\}$ are set to $t_0/2$. The 10 slowest $\{1/\lambda_p\}$ are listed in the second column of Table I.

In the authors' experience, the slow $\{X_p\}$ obtained from the conventional RMA with small t_0 and τ contain the true slow $\{X_p\}$ [13], although the $\{1/\lambda_p\}$ are underestimated. This can be seen by calculating the time-displaced autocorrelation function $C'_{p,p}(t)$ of the pth slow relaxation mode X_p directly from simulations. The autocorrelation $C'_{p,p}(t)$ exhibits slow relaxation at long times. The relaxation time estimated from the long-time behavior of $C'_{p,p}(t)$ is usually significantly longer than $1/\lambda_p$, which is estimated from the short-time behavior of $C'_{p,p}(t)$, as $C'_{p,p}(\tau) = \exp(-\lambda_p\tau)$ with small t_0 and τ. The time-displaced $C'_{p,p}(t)$ of the nine slowest X_p obtained from the first RMA are directly calculated from the simulation and shown in Fig. 3(a). Note that $C'_{p,p}(t)$ is calculated from

TABLE I. 10 slowest relaxation times obtained from first and second RMAs.

Mode number p or u	Relaxation time $1/\lambda_p$ (1st RMA) (ns)	Relaxation time $1/\lambda'_u$(2nd RMA) (ns)
1	2.8×10	4.1×10^2
2	2.5×10	1.5×10^2
3	2.1×10	8.3×10
4	1.5×10	6.9×10
5	1.4×10	5.0×10
6	1.0×10	3.6×10
7	8.1	3.1×10
8	7.5	2.5×10
9	6.5	2.4×10
10	4.7	1.8×10

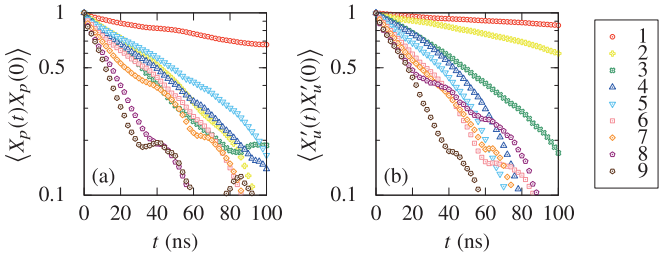

FIG. 3. Time-displaced autocorrelation functions (a) $\langle X_p(t)X_p(0)\rangle$ $(p = 1, \ldots, 9)$ obtained via first RMA and (b) $\langle X'_u(t)X'_u(0)\rangle$ $(u = 1, \ldots, 9)$ obtained via second RMA.

$C_{i,j}(t_0 + t)$ using Eq. (18) and that $C'_{p,p}(0) = 1$ is obtained from the orthonormal condition (11) with $t_i = t_j = t_0$. These $C'_{p,p}(t)$ exhibit slow relaxation processes at long times. The relaxation times estimated from their long-time behavior, which, for example, can be the time at which $C'_{p,p}(t)$ becomes 0.5, are significantly longer than the $1/\lambda_p$ values shown in Table I. For the first mode, the time at which $C'_{1,1}(t)$ becomes 0.5 is approximately 200 ns (data not shown), while $1/\lambda_1$ is 28 ns, as shown in Table I. The slow $\{X_p\}$ yielded by the first RMA indeed contain the true slow $\{X_p\}$, even if we choose small t_0 and τ.

To improve the relaxation mode accuracy and to obtain more accurate relaxation times, the second RMA is applied to N_m of the slowest $\{X_p\}$ obtained from the first RMA. Here, we set $N_m = 20$. The number of degrees of freedom is reduced to N_m. The evolution times $\{t'_p\}$ are estimated from the results of the first RMA and chosen as $t'_p \simeq r_t/\lambda_p$ with $r_t = 0.03$. Note that we can also estimate $\{t'_p\}$ from the long-time behavior of the autocorrelation functions of $\{X_p\}$, $C'_{p,p}(t)$, directly calculated from the simulation. The time interval τ' is chosen to be $\tau' = 10\,000$ ps. Note that we can select a τ' that is longer than τ because the number of degrees of freedom is reduced and the physical quantities $\{X_p\}$ exhibit slow relaxations. The relaxation times $1/\lambda'_u$ obtained from Eq. (17) are listed in the third column of Table I. By comparing the second and third columns of Table I, we find that the relaxation times are improved significantly. The time-displaced autocorrelation functions of the relaxation modes yielded by the second RMA $\{X'_u\}$ are shown in Fig. 3(b). The relaxation of $\langle X'_u(t)X'_u(0)\rangle$ gradually accelerates as u becomes large. The second RMA also improves the relaxation mode accuracy.

In order to examine the validity of the RMA techniques, the time-displaced autocorrelation functions of principal components reproduced from the first and second RMAs are shown in Fig. 4. We also show results yielded by the PCRMA, which was previously introduced in Ref. [16]. Here, the PCRMA parameters are chosen to be $N_c = 20$, $t_1 = 40$ ps, $t_2 = 1000$ ps, and $\tau = 350$ ps. Note that the time-displaced autocorrelation functions reproduced from the first RMA are in disagreement with those obtained from the simulation directly because the $\{1/\lambda_p\}$ are underestimated, as seen previously. However, the time-displaced autocorrelation functions reproduced from the second RMA are in good agreement with those obtained from the simulation directly. In particular, the slow relaxation processes of the seventh and ninth principal components are also well reproduced by the second RMA. The $\{X'_u\}$ and $\{1/\lambda'_u\}$ obtained from the second RMA are more accurate

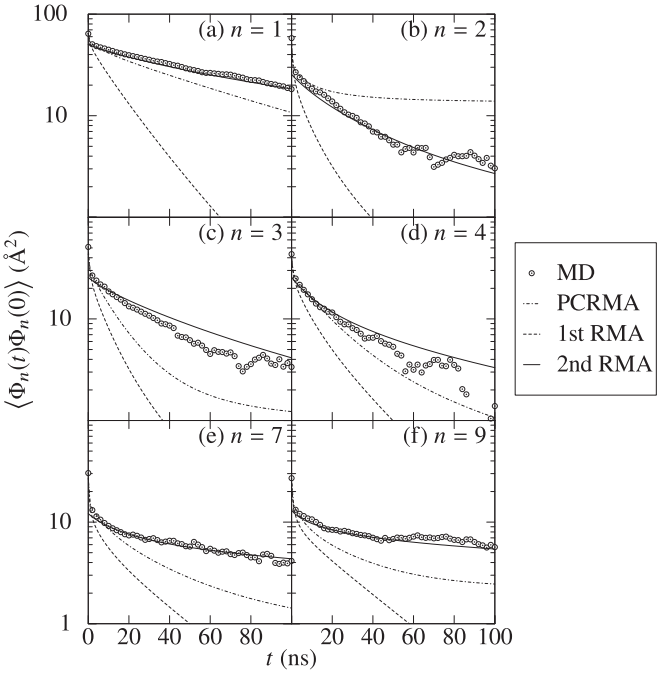

FIG. 4. Time-displaced autocorrelation functions of $\langle \Phi_n(t)\Phi_n(0)\rangle$ for n values of (a)–(f) 1, 2, 3, 4, 7, and 9, respectively, calculated directly via simulation (circles) and reproduced by first RMA (dashed lines), second RMA (solid lines), and PCRMA (dashed-dotted lines). The values for the seventh and ninth principal components are shown because they exhibit slower relaxation processes. For the PCRMA introduced in Ref. [16], we set N_c = 20, t_1 = 40 ps, t_2 = 1000 ps, and τ = 350 ps.

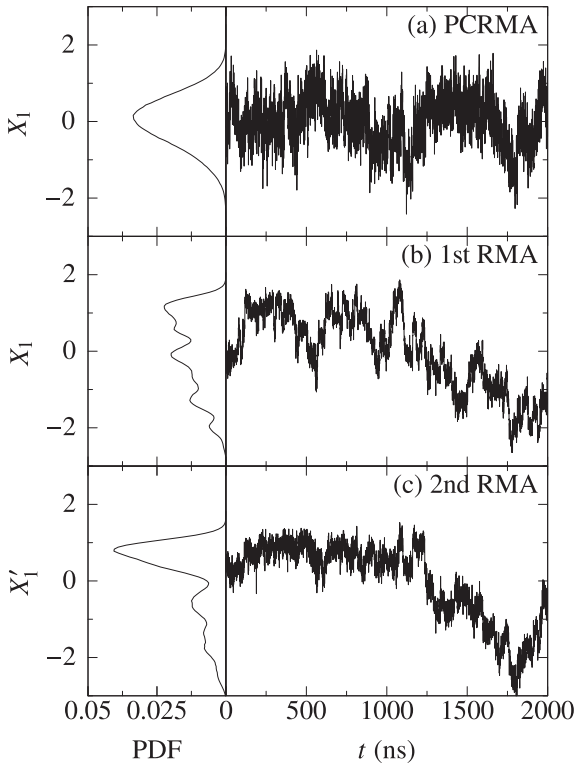

FIG. 5. Time series and probability density function (PDF) of first relaxation mode for (a) PCRMA, (b) first RMA, and (c) second RMA.

than those yielded by the first RMA. Comparison of the results of the PCRMA presented in Ref. [16] and those of the present method (two-step RMA) indicate that the present method can better reproduce the long-time behavior of the time-displaced autocorrelation functions. These results demonstrate that we obtain relatively accurate relaxation modes and times from the simulation using the two-step RMA.

In Fig. 5, the time series and probability density function (PDF) of the first slowest relaxation mode for the PCRMA, the first RMA, and the second RMA are shown. The fluctuations for the second RMA are reduced to a greater extent than those for the PCRMA and first RMA. The second RMA clearly extracts a rare event during the simulation. The large change near 1200 ns relates to local structural changes near Asn46, Arg73, and Asn103. The dynamics of proteins on the microsecond time scale usually corresponds to local conformational changes of the residues [49]. The second RMA extracts the rare conformational changes of these residues during the simulation.

V. CONCLUSIONS

In this paper, we have demonstrated the effectiveness of two-step RMA with multiple evolution times for a protein system. In this method, RMA with a single evolution time using small t_0 and τ is first implemented, yielding rough estimates of the relaxation modes and rates. Then, the second RMA is applied to a small number of the slowest relaxation modes obtained in the previous stage. In the second RMA, we use RMA with multiple evolution times because the relaxation modes obtained from the first RMA each have different relaxation times. Herein, we applied this method to the results of a PCA. That is, PCA was first applied to a 2-μs MD simulation of hen egg-white lysozyme in aqueous solution. Then, the two-step RMA method with multiple evolution times was applied to the principal components obtained via that method. The time-displaced autocorrelation functions of the principal components reproduced from the second RMA were in good agreement with those obtained from the simulation directly. Further, the slow relaxation modes and corresponding relaxation times for the principal components were much improved by the second RMA. Note that, if the slow modes and times are not sufficient to represent the dynamics of the system, we can also repeat the RMA, i.e., implementing three-step RMA, and so on. We can calculate the motions of proteins along the obtained modes using the relations between the modes $\boldsymbol{R}$, $\boldsymbol{\Phi}_c$, $\{X_p\}$, and $\{X'_u\}$. The free energy surfaces obtained using the relaxation modes can also be calculated, in the same manner as in Ref. [13]. The directions of the slower relaxation modes will reflect the transitions between the free energy minimum states. A combined method featuring the Markov state model and tICA or TICA was proposed in Refs. [18,19]. In that approach, a Markov state model is constructed from clustering in the subspace determined by the tICA or TICA. For a simulation of a 10-residue small protein, chignolin, near its transition temperature, Markov state relaxation mode analysis, which is the extension of a regular Markov state model obtained by introducing t_0, was

also applied to several characteristic states: native, misfolded, intermediate, and unfolded states. These states were classified by an original RMA [24]. The refined slow relaxation modes obtained via the present method can be also used to cluster states in order to construct several kinetic models, such as Markov state models and milestone methods [50]. This method is a powerful technique for investigating the dynamics and kinetics of a system through long simulations.

ACKNOWLEDGMENTS

The authors would like to thank S. Natori for fruitful discussion. This work was supported by JST PRESTO (JPMJPR13LB). This work was also partially supported by JSPS KAKENHI Grant No. JP24540441. N.K. acknowledges a Research Grant from the Keio Leading-edge Laboratory of Science and Technology.

[1] C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, J. Comput. Phys. **151**, 146 (1999).

[2] W. C. Swope, J. W. Pitera, and F. Suits, J. Phys. Chem. B **108**, 6571 (2004).

[3] N. Singhal, C. D. Snow, and V. S. Pande, J. Chem. Phys. **121**, 415 (2004).

[4] J. D. Chodera, W. C. Swope, J. W. Pitera, and K. A. Dill, Multiscale Modell. Simul. **5**, 1214 (2006).

[5] J. D. Chodera, N. Singhal, V. S. Vande, K. A. Dill, and W. C. Swope, J. Chem. Phys. **126**, 155101 (2007).

[6] F. Noé, I. Horenko, C. Schütte, and J. C. Smith, J. Chem. Phys. **126**, 155102 (2007).

[7] N. Buchete and G. Hummer, J. Phys. Chem. B **112**, 6057 (2008).

[8] O. F. Lange and H. Grubmüller, Proteins **70**, 1294 (2008).

[9] F. Noé and S. Fischer, Curr. Opin. Struct. Biol. **18**, 154 (2008).

[10] D. M. Zuckerman, *Statistical Physics of Biomolecules: An Introduction* (CRC Press, Boca Raton, FL, 2010).

[11] S. Sakuraba, Y. Joti, and A. Kitao, J. Chem. Phys. **133**, 185102 (2010).

[12] M. Kamada, M. Toda, M. Sekijima, M. Takata, and J. Joe, Chem. Phys. Lett. **502**, 241 (2011).

[13] A. Mitsutake, H. Iijima, and H. Takano, J. Chem. Phys. **135**, 164102 (2011); Seibutsu Butsuri (Biophysics), **45**, Supplement S214 (2005) (Abstracts for the 43th Annual Meeting, The Biophysical Society of Japan) (in Japanese).

[14] Y. Naritomi and S. Fuchigami, J. Chem. Phys. **134**, 065101 (2011).

[15] J. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, J. Chem. Phys. **134**, 174105 (2011).

[16] T. Nagai, A. Mitsutake, and H. Takano, J. Phys. Soc. Jpn. **82**, 023803 (2013); Seibutsu Butsuri (Biophysics), **49**, Supplement S75, (Abstracts for the 47the Annual Meeting, The Biophysical Society of Japan) (2009).

[17] Y. Naritomi and S. Fuchigami, J. Chem. Phys. **139**, 215102 (2013).

[18] G. Pérez-Hernández, F. Paul, T. G. Giorgino, D. Fabritiis, and F. Noé, J. Chem. Phys. **139**, 015102 (2013).

[19] C. R. Schwantes and V. S. Pande, J. Chem. Theory Comput. **9**, 2000 (2013).

[20] Y. Matsunaga, A. Baba, C. Li, J. E. Straub, M. Toda, T. Komatsuzaki, and R. S. Berry, J. Chem. Phys. **139**, 215101 (2013).

[21] J. D. Chodera and F. Noé, Curr. Opin. Struct. Biol. **25**, 135 (2014).

[22] C. R. Schwantes, R. T. McGibbon, and V. S. Pande, J. Chem. Phys. **141**, 090901 (2014).

[23] *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, edited by G. R. Bowman, V. S. Pande, F. Noé (Springer, Dordrecht, 2014).

[24] A. Mitsutake and H. Takano, J. Chem. Phys. **143**, 124111 (2015).

[25] T. Mori and S. Saito, J. Chem. Phys. **142**, 135101 (2015).

[26] Y. Matsunaga, A. Kidera, and Y. Sugita, J. Chem. Phys. **142**, 214115 (2015).

[27] T. Mori and S. Saito, J. Phys. Chem. B **120**, 11683 (2016).

[28] C. T. Leahy, R. D. Murphy, G. Hummer, E. Rosta, and N. Buchete, J. Phys. Chem. Lett. **7**, 2676 (2016).

[29] F. Noé and C. Clementi, Curr. Opin. Struct. Biol. **43**, 141 (2017).

[30] H. Wu, F. Nüske, F. Paul, S. Klus, P. Koltai, and F. Noé, J. Chem. Phys. **146**, 154104 (2017).

[31] L. Martini, A. Kells, R. Covino, G. Hummer, N. Buchete, and E. Rosta, Phys. Rev. X **7**, 031060 (2017).

[32] H. Takano and S. Miyashita, J. Phys. Soc. Jpn. **64**, 3688 (1995).

[33] S. Koseki, H. Hirao, and H. Takano, J. Phys. Soc. Jpn. **66**, 1631 (1997).

[34] H. Hirao, S. Koseki, and H. Takano, J. Phys. Soc. Jpn. **66**, 3399 (1997).

[35] K. Hagita and H. Takano, J. Phys. Soc. Jpn. **71**, 673 (2002).

[36] S. Saka and H. Takano, J. Phys. Soc. Jpn. **77**, 034001 (2008).

[37] N. Iwaoka, K. Hagita, and H. Takano, J. Phys. Soc. Jpn. **84**, 044801 (2015) and references therein.

[38] P. G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1984).

[39] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Oxford University Press, Oxford, 1986).

[40] S. Natori and H. Takano, J. Phys. Soc. Jpn. **86**, 043003 (2017).

[41] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, 2nd ed. (Springer, Berlin, 1989).

[42] D. A. Case, V. Babin, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Götz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossvry, A. Kovalenko, T. S. Lee, S. Le Grand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, and P. A. Kollman, AMBER 14, University of California, San Francisco, 2014.

[43] R. Diamond, J. Mol. Biol. **82**, 371 (1974).

[44] A. W. Götz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand, and R. C. Walker, J. Chem. Theory Comput. **8**, 1542 (2012).

[45] R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, and R. C. Walker, J. Chem. Theory Comput. **9**, 3878 (2013).
[46] S. Le Grand, A. W. Götz, and R. C. Walker, Comput. Phys. Commun. **184**, 374 (2013).
[47] C. Eckart, Phys. Rev. **47**, 552 (1935).
[48] A. D. McLachlan, J. Mol. Biol. **128**, 49 (1979).
[49] R. O. Dror, R. M. Dirks, J. P. Grossman, H. Xu, and D. E. Shaw, Annu. Rev. Biophys. **41**, 429 (2012).
[50] A. K. Faradjian and R. Elber, J. Chem. Phys. **120**, 10880 (2004).