Biophysical Reviews (2018) 10:375–389
https://doi.org/10.1007/s12551-018-0406-7

REVIEW

Relaxation mode analysis for molecular dynamics simulations of proteins

Ayori Mitsutake[1] · **Hiroshi Takano**[1]

Received: 24 December 2017 / Accepted: 6 February 2018 / Published online: 15 March 2018

Abstract
Molecular dynamics simulation is a powerful method for investigating the structural stability, dynamics, and function of biopolymers at the atomic level. In recent years, it has become possible to perform simulations on time scales of the order of milliseconds using special hardware. However, it is necessary to derive the important factors contributing to structural change or function from the complicated movements of biopolymers obtained from long simulations. Although some analysis methods for protein systems have been developed using increasing simulation times, many of these methods are static in nature (i.e., no information on time). In recent years, dynamic analysis methods have been developed, such as the Markov state model and relaxation mode analysis (RMA), which was introduced based on spin and homopolymer systems. The RMA method approximately extracts slow relaxation modes and rates from trajectories and decomposes the structural fluctuations into slow relaxation modes, which characterize the slow relaxation dynamics of the system. Recently, this method has been applied to biomolecular systems. In this article, we review RMA and its improved versions for protein systems.

Keywords Protein · Simulation · Analysis · Dynamics

Introduction

Molecular dynamics simulation is widely used for protein research. In general, the focus of this research is to extract information on the physical properties of individual proteins. The results from such simulations are then often compared with experimental results. Since these experiments are generally conducted in solvents, it is necessary to simulate protein and water molecular systems, which are complicated systems. These simulations are conducted for a variety of purposes such as to analyze the stability and dynamics of the structures around crystal structures and to determine folding from an extended structure into a native structure. There are three difficulties in current approaches for protein simulations (Freddolino et al. 2010). The first is the potential function of the protein systems. In recent years, it has become possible to evaluate the molecular force field by improving the sampling, and accuracy has consequently improved. The second problem is related to the sampling. With respect to the folding mechanism, simulation at the millisecond scale is necessary. Recently, it has become possible to perform simulations at the millisecond scale by using special hardware such as Anton (Lindorff-Larsen et al. 2011, 2012; Dror et al. 2012, Lane et al. 2013), but sampling problems still exist for complex systems such as ligand-binding systems and other even more complex systems. The third issue is related to the analysis methods. It is important to extract the characteristic degrees of freedom (order parameters) from the complex protein movements obtained from simulations, which are good indicators for analyzing trajectories.

In normal mode analysis, the normal mode near the minimum point of the potential energy of the protein molecule is obtained (Go et al. 1983; Brooks and Karplus 1983; Levitt et al. 1985). Langevin mode analysis investigates modes around the native structure, including the water effect (Lamm and Szabo 1986; Kottalam and Case 1990; Kitao et al. 1991; Hayward et al. 1993). An elastic network model and Gaussian network model approximately estimate normal modes with large amplitudes by using the harmonic potential of coarse-grained models (Tirion

This article is part of a Special Issue on 'Biomolecules to Bio-nanomachines - Fumio Arisaka 70th Birthday' edited by Damien Hall, Junichi Takagi and Haruki Nakamura.

✉ Ayori Mitsutake
ayori@mail.rk.phys.keio.ac.jp

[1] Department of Physics, Faculty of Science and Technology, Keio University, Tokyo, Japan

1996; Baher et al. 1997; Tama and Sanejouand 2001; Cui and Bahar 2005; Miyashita and Tama 2008). This method extracts collective modes with large amplitudes in the case of huge protein systems such as viruses, because huge proteins have rigid-like motions (Tama and Brooks III 2002).

Principal component analysis (PCA), also called quasi-harmonic analysis or the essential dynamics method (Levy et al. 1984; Ichiye and Karplus 1991; Abagyan and Argos 1992; Garcia 1992; Hayward et al. 1993; Amadei et al. 1993; Kitao and Go 1999), is one of the most popular methods adopted for analyzing the structural fluctuations around the average structure. The modes with large structure fluctuations are extracted and are regarded as cooperative movement, and the relation of these fluctuations with function has been widely investigated. The obtained modes are also used as the axis of the free-energy surface. Moreover, various other analysis methods have been proposed, such as full correlation analysis (Lange and Grubmüller 2007), subspace joint approximate diagonalization of eigenmatrices (Sakuraba et al. 2010), and wavelet analysis (Kamada et al. 2011), among others (Moritsugu et al. 2015; Matsunaga et al. 2015).

In recent years, it has become possible to perform an extensively long simulation; thus, development of dynamic analysis methods to identify the local minimum-energy states and analyze the transitions between them is required. Accordingly, many methods to analyze the dynamics and kinetics of protein simulations have been developed (Zuckerman 2010; Komatsuzaki et al. 2011; Bowman et al. 2014). In particular, the Markov state model has been presented and applied to many protein systems (Schütte et al. 1999; Swope et al. 2004; Singhal et al. 2004; Chodera et al. 2006, 2007; Chodera and Noé 2014; Noé et al. 2007; Noé and Fischer 2008; Noé and Clementi 2017 Buchete and Hummer 2008; Prinz et al. 2011; Pérez-Hernández et al. 2013; Schwantes and Pande 2013; Schwantes et al. 2014; Bowman et al. 2014; Wu et al. 2017). The Markov state model can analyze transitions between local minimum-energy states, which are identified from clustering analysis methods. This is a powerful method for analyzing dynamics in the context of both long and short simulations of proteins.

Relaxation mode analysis (RMA) was developed to investigate the "dynamic" properties of spin systems (Takano and Miyashita 1995) and homopolymer systems for Monte Carlo (Koseki et al. 1997) and molecular dynamics (Hirao et al. 1997) analyses, and has been applied to various polymer systems (Hagita and Takano 2002; Saka and Takano 2008; Iwaoka et al. 2015; Natori and Takano 2017) to investigate their slow relaxation dynamics (de Gennes 1984; Doi and Edwards 1986). Recently, RMA has also been applied to biomolecular systems (Mitsutake et al. 2011; Mitsutake et al. 2005; Mitsutake and Takano 2015; Nagai et al. 2009, 2013). RMA approximately estimates slow relaxation modes and rates from trajectories obtained from simulations.

The relaxation modes $\{X_p\}$ satisfy

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

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

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

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, $P_{\mathrm{eq}}(Q')$ 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 time is given by $1/\lambda_p$. Note that 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, from the viewpoint of the statistical mechanics (Hirao et al. 1997; Koseki et al. 1997; Mitsutake and Takano 2015) (see the "Relaxation modes $\{X_p\}$ and rates λ_p" section). The point of RMA is that we consider the variational problem, which is equivalent to the eigenvalue problem of the time evolution operator, and choose an appropriate trial function to estimate the slow relaxation modes and rates in the system (see the "RMA" section). From these processes, we obtain the generalized eigenvalue problem of the time correlation matrices for two different times. From the eigenvectors and eigenvalues, we approximately estimate slow relaxation modes and rates.

Conventional RMA approximately estimates slow relaxation modes by solving the generalized eigenvalue problem of the time correlation matrices of coordinates for two different times, $C(\tau + t_0)$ and $C(t_0)$, which are calculated from the trajectory. Recently, dynamical analysis methods for molecular simulations of biopolymer systems have been developed to investigate slow dynamics. In these techniques such as time structure-based independent component analysis (tICA) (Naritomi and Fuchigami 2011, 2013), time-lagged independent component analysis (TICA) (Pérez-Hernández et al. 2013; Schwantes and Pande 2013), and dynamic component analysis (DCA) (Mori et al. 2015, 2016), time correlation matrices of certain physical quantities or states are used. (Note that tICA is a special case of RMA with $t_0 = 0$. See Mitsutake et al. (2011) and Naritomi and Fuchigami (2011) for more details on the differences between tICA and RMA.) In tICA, TICA, and DCA, the time correlation functions $C(\tau)$ and $C(0)$ are used, whereas $C(\tau + t_0)$ and $C(t_0)$ are used in RMA. 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. From this point of view, RMA is

related to Markov state models. (The relationship among the Markov state model, tICA, and TICA is explained in Pérez-Hernández et al. (2013), Schwantes and Pande (2013), and Mitsutake and Takano (2015).) The combination method of tICA and a Markov state model was also proposed (Pérez-Hernández et al. 2013; Schwantes and Pande 2013). A Markov state model was constructed from clustering in the subspace determined by tICA.

In this review, we first provide a definition of relaxation modes and rates from the viewpoint of the statistical mechanics in the "Relaxation modes $\{X_p\}$ and rates λ_p" section. The "RMA" section explains the original RMA (RMA with a single evolution time) and the process of RMA using coordinates for the trial function in detail. The "Improvement of RMA" section explains the improved versions of RMA, including RMA with multiple evolution times, principal component RMA (PCRMA), two-step RMA, and Markov-state RMA (MSRMA). Finally, in the "Application of RMA to a system with large conformational changes" section, we present results from studies in which RMA was applied to a system with large conformational changes. The "Conclusions" section provides conclusions and perspectives on the state of the field.

Relaxation modes $\{X_p\}$ and rates λ_p

In this section, we provide the definition of relaxation modes and rates from the viewpoint of the statistical mechanics (Risken 1989; Zwanzig 2001). The relaxation modes $\{X_p\}$ satisfy Eq. 1. 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. We first explain the relation in three types of simulations satisfying the detailed balance condition.

In a Monte Carlo simulation satisfying the detailed balance condition, the time evolution of the probability $P(Q;t)$ that the biomolecule is in a state $Q = \left(\boldsymbol{r}_1^{\mathrm{T}}, \boldsymbol{r}_2^{\mathrm{T}}, \cdots, \boldsymbol{r}_N^{\mathrm{T}}\right)^{\mathrm{T}}$ at time t is described by a master equation:

$$\frac{\partial}{\partial t}P(Q;t) = -\sum_{Q'}\Gamma(Q|Q')P(Q';t). \tag{3}$$

Here, $\Gamma(Q|Q')$ denotes the (Q, Q')-component of the time evolution matrix Γ, and $\sum_{Q'}$ denotes the summation over all possible states. $\Gamma(Q|Q')$ is also chosen so that the detailed balance for the equilibrium distribution function $P_{\mathrm{eq}}(Q)$ is satisfied:

$$\Gamma(Q|Q')P_{\mathrm{eq}}(Q') = \Gamma(Q'|Q)P_{\mathrm{eq}}(Q). \tag{4}$$

In the Brownian dynamics simulation, the time evolution of coordinates $\boldsymbol{r}_i$, $(i = 1, \cdots, N)$ is given by the Langevin equation for a biomolecule with N atoms:

$$\frac{d\boldsymbol{r}_i}{dt} = -\frac{1}{\zeta}\left[-\frac{\partial}{\partial \boldsymbol{r}_i}U(\{\boldsymbol{r}_j\}) + \boldsymbol{w}_i\right]. \tag{5}$$

Here, $\boldsymbol{r}_i(t)$ denotes the position of the ith atom at time t, and ζ is the friction constant. The interaction between atoms is described by the potential $U(\{\boldsymbol{r}_i\}) = U(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$. The random force $\boldsymbol{w}_i(t)$ acting on the ith atom is a Gaussian white stochastic process and satisfies

$$\langle w_{i,\alpha}(t)w_{j,\beta}(t)\rangle = 2\zeta k_B T\delta_{\alpha,\beta}\delta_{i,j}\delta(t-t'), \tag{6}$$

where $w_{i,\alpha}$, k_B, and T denote the α-component of $\boldsymbol{w}_i$ (α=x, y, or z), the Boltzmann constant, and the temperature of the system, respectively. The Smoluchowski equation equivalent to Eq. 5 can be written as

$$\begin{aligned}\tfrac{\partial}{\partial t}P(Q,t) &= -\Gamma(Q)P(Q,t)\\ &= \sum_{i=1}\frac{\partial}{\partial \boldsymbol{r}_i}\cdot\frac{1}{\zeta}\left\{k_B T\frac{\partial}{\partial \boldsymbol{r}_i} + \frac{\partial U}{\partial \boldsymbol{r}_i}\right\}P.\end{aligned} \tag{7}$$

Here, $Q = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N\}$ denotes a point in the phase space of the system, and $P(Q,t)dQ$ denotes the probability that the system is found at time t in an infinitesimal volume dQ at point Q in the phase space. The time evolution operator Γ satisfies the detailed balance condition (Risken 1989):

$$P_{\mathrm{eq}}(Q')\Gamma(Q)\delta(Q-Q') = P_{\mathrm{eq}}(Q)\Gamma^{\dagger}(Q)\delta(Q-Q'), \tag{8}$$

where $P_{\mathrm{eq}}(Q) \propto \exp\left[-\dfrac{U(\{\boldsymbol{r}_j\})}{k_B T}\right]$. Here, $\Gamma(Q)\delta(Q-Q')$ and the adjoint operator $\Gamma^{\dagger}(Q)\delta(Q-Q')$ act only on Q in $\delta(Q-Q')$. In the matrix representation, so that $\Gamma(Q)\delta(Q-Q') = \Gamma(Q|Q')$ and $\Gamma^{\dagger}(Q)\delta(Q-Q') = \Gamma(Q'|Q)$, the detailed balance condition is the same as that in Eq. 4.

In a molecular dynamics simulation with the Langevin thermostat, the time evolution of coordinates $\boldsymbol{r}_i$, $(i = 1, \cdots, N)$ is given by the Langevin equation for a biomolecule with N atoms:

$$m_i\frac{d\boldsymbol{v}_i}{dt} = -\zeta\boldsymbol{v}_i - \frac{\partial}{\partial \boldsymbol{r}_i}U(\{\boldsymbol{r}_j\}) + \boldsymbol{w}_i, \tag{9}$$

with

$$\frac{d\boldsymbol{r}_i}{dt} = \boldsymbol{v}_i. \tag{10}$$

Here, $\boldsymbol{r}_i(t)$ and $\boldsymbol{v}_i(t)$ denote the position and the velocity of the ith atom at time t, respectively. The mass of the ith atom is denoted by m_i and ζ is the friction constant.

The Kramers equation, equivalent to Eqs. 9 and 10, can be written as

$$\begin{aligned}&\tfrac{\partial}{\partial t}P(Q,t) = -\Gamma(Q)P(Q,t)\\ &= \sum_{i=1}^{N}\left\{\frac{\partial}{\partial \boldsymbol{r}_i}\cdot\boldsymbol{v}_i - \frac{1}{m_i}\frac{\partial}{\partial \boldsymbol{v}_i}\cdot\frac{\partial U}{\partial \boldsymbol{r}_i} - \frac{\zeta}{m_i}\frac{\partial}{\partial \boldsymbol{v}_i}\cdot\left(\boldsymbol{v}_i + \frac{k_B T}{m_i}\frac{\partial}{\partial \boldsymbol{v}_i}\right)\right\}P.\end{aligned} \tag{11}$$

Here, $Q = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\}$ denotes a point in the phase space of the system. The time evolution operator Γ satisfies the detailed balance condition:

$$P_{\rm eq}(Q')\Gamma(Q)\delta(Q - Q') = P_{\rm eq}(\epsilon Q)\Gamma^{\dagger}(\epsilon Q)\delta(\epsilon Q - \epsilon Q'), \tag{12}$$

where $P_{\rm eq}(Q) \propto \exp\left(-\frac{1}{k_B T}\left[\frac{1}{2}\sum_i m_i \boldsymbol{v}_i^2 + U(\{\boldsymbol{r}_j\})\right]\right)$ and $P_{\rm eq}(Q) = P_{\rm eq}(\epsilon Q)$. Here, ϵQ denotes the time-reversed state of the state Q, namely, $\epsilon Q = \{\epsilon_1 \boldsymbol{r}_1, \ldots, \epsilon_N \boldsymbol{r}_N, .\epsilon_{N+1}\boldsymbol{v}_1, \ldots, \epsilon_{2N}\boldsymbol{v}_N\}$ with

$$\epsilon_i = \begin{cases} 1 & \text{for } i = 1, \ldots, N, \\ -1 & \text{for } i = N+1, \ldots, 2N. \end{cases} \tag{13}$$

In the matrix representation, the detailed balance condition is written as follows:

$$\Gamma(Q|Q')P_{\rm eq}(Q') = \Gamma(\epsilon Q'|\epsilon Q)P_{\rm eq}(\epsilon Q). \tag{14}$$

The time evolution equation of $P(Q;t)$ of Eqs. 7 and 11 corresponds to Eq. 3 in the matrix representation. In Monte Carlo and Brownian dynamics, because only coordinates are the degrees of freedom in the system, $\epsilon Q = Q$, the detailed balance condition in all three cases is given by Eq. 14.

We now consider the eigenvalue problem of the time evolution operator $\Gamma(Q|Q')$ of the master equation:

$$\sum_Q \phi_n(Q)\Gamma(Q|Q') = \lambda_n \phi_n(Q'). \tag{15}$$

$$\sum_{Q'} \Gamma(Q|Q')\psi_n(Q') = \lambda_n \psi_n(Q). \tag{16}$$

Here, $\phi_n(Q)$ and $\psi_n(Q)$ are the left and right eigenfunctions of the time evolution operator Γ with eigenvalue λ_n, respectively. When we define a quantity $\hat{\phi}_n(Q)$ through

$$\psi_n(Q) = \hat{\phi}_n(Q)P_{\rm eq}(Q), \tag{17}$$

then $\hat{\phi}_n(Q) = \phi_n(\epsilon Q)$. The eigenfunctions are chosen to satisfy the orthonormal condition:

$$\sum_Q \phi_m(Q)\psi_n(Q) = \sum \phi_m(Q)\hat{\phi}_n P_{\rm eq}(Q) = \langle \phi_m \hat{\phi}_n \rangle = \delta_{m,n}. \tag{18}$$

The equilibrium time-displaced correlation function of $\phi_n(Q)$ and $\hat{\phi}_m(Q)$ is given by the following:

$$\begin{aligned} \langle \phi_m(t)\hat{\phi}_n(0)\rangle &= \sum_Q \sum_{Q'} \phi_m(Q)T_t(Q|Q')\hat{\phi}_n(Q')P_{\rm eq}(Q') \\ &= \sum_Q \sum_{Q'} \phi_m(Q)e^{-\Gamma t}(Q|Q')\hat{\phi}_n(Q')P_{\rm eq}(Q') \\ &= \sum_Q \sum_{Q'} \phi_m(Q)e^{-\Gamma t}(Q|Q')\psi_n(Q') \\ &= \sum_Q \phi_m(Q)e^{-\lambda_n t}\psi_n(Q) \\ &= \delta_{m,n}e^{-\lambda_n t}, \end{aligned} \tag{19}$$

where $T_t(Q|Q') = e^{-\Gamma\tau}(Q|Q')$ is the conditional probability that the system is found at time t at Q given that the system is at Q' at time 0.

If two quantities $A(Q)$ and $B(Q)$ are expanded as

$$A(Q) = \sum_n a_n \phi_n(Q) \text{ and } B(Q) = \sum_n \hat{b}_n \hat{\phi}_n(Q), \tag{20}$$

then the time correlation function of A and B in the equilibrium state is given by

$$\langle A(t)B(0)\rangle = \sum_n a_n \hat{b}_n \exp(-\lambda_n t). \tag{21}$$

Thus, in terms of $\phi_n(Q)$ and $\hat{\phi}_n(Q)$, the correlation function $\langle A(t)B(0)\rangle$ is decomposed into a sum of exponentially relaxing contributions. Therefore, we use two sets of functions, $\{\phi_n(Q)\}$ and $\{\hat{\phi}_n(Q)\}$, as relaxation modes, and refer to $\{\lambda_n\}$ as their relaxation rates. 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.

RMA

RMA with a single evolution time, t_0

RMA approximately estimates slow relaxation modes and rates from trajectories obtained from simulations. Herein, we explain how to obtain the slow relaxation modes and rates. The point of this method is that we consider the variational problem, which is equivalent to the eigenvalue problem of the time evolution operator, and choose an appropriate trial function in order to estimate the slow relaxation modes and rates in the system.

We consider the equations for the conditional probability:

$$\sum_{Q} \phi_n(Q) T_\tau(Q|Q') = e^{-\lambda_n \tau} \phi_n(Q'), \tag{22}$$

$$\sum_{Q'} T_\tau(Q|Q') \psi_n(Q') = e^{-\lambda_n \tau} \psi_n(Q). \tag{23}$$

The eigenvalue problem in Eqs. 22 and 23 is equivalent to the variational problem

$$\delta \mathcal{R} = 0 \tag{24}$$

with

$$\mathcal{R}[\phi_n] = \frac{\langle \phi_n(\tau) \hat{\phi}_n(0) \rangle}{\langle \phi_n(0) \hat{\phi}_n(0) \rangle}, \tag{25}$$

and the stationary value of $\mathcal{R}$ gives the eigenvalue $\exp(-\lambda_n \tau)$. RMA treats the variational problem of Eqs. 24 and 25 using trial functions instead of the eigenvalue problem of Eqs. 22 and 23. To choose the trial function given by a linear combination of important relevant quantities, we can evaluate the relaxation modes and rates from simulation data.

Herein, we consider a biopolymer composed of N atoms and only treat the coordinates, because the velocities have faster relaxations ($\sim$ picosecond order) than coordinates in protein systems. We assume that $\boldsymbol{R}$ is a $3N$-dimensional column vector that consists of a set of atomic coordinates relative to their average coordinates

$$\boldsymbol{R}^{\mathrm{T}} = (\boldsymbol{r}_1'^{\mathrm{T}}, \boldsymbol{r}_2'^{\mathrm{T}}, \ldots, \boldsymbol{r}_N'^{\mathrm{T}}) = (x_1', y_1', z_1', \ldots, x_N', y_N', z_N'), \tag{26}$$

with

$$\boldsymbol{r}_i' = \boldsymbol{r}_i - \langle \boldsymbol{r}_i \rangle, \tag{27}$$

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. Note that because we consider the coordinates only, $\hat{\phi}_n(Q) = \phi_n(\epsilon Q) = \phi_n(Q)$ holds.

In RMA, we use the following function as an approximate relaxation mode:

$$X_p(Q) = \sum_{i=1}^{3N} f_{p,i} R_i(t_0/2; Q), \tag{28}$$

with

$$R_i(t; Q) = \sum_{Q'} R_i(Q') T_t(Q'|Q). \tag{29}$$

Here, $R_i(Q)$ is the ith component of $\boldsymbol{R}$. The quantity $R_i(t; Q)$ is the expectation value of R_i after a period t starting from a state Q and satisfies $R_i(t; Q)|_{t=0} = R_i(Q)$. The parameter t_0 is introduced in order to reduce the relative weight of the faster modes contained in $\boldsymbol{R}$, and it is expected that Eq. 28 becomes a better approximation as t_0 becomes larger.

For the trial function (28), $\mathcal{R}$ defined by Eq. 25 is given by

$$\mathcal{R}[X_p] = \frac{\sum_{i=1}^{3N} \sum_{j=1}^{3N} f_{p,i} C_{i,j}(t_0 + \tau) f_{p,j}}{\sum_{i=1}^{3N} \sum_{j=1}^{3N} f_{p,i} C_{i,j}(t_0) f_{p,j}}, \tag{30}$$

where $C_{i,j}(t)$ is a component of a $3N \times 3N$ symmetric matrix $C(t)$ defined by

$$C_{i,j}(t) = \langle R_i(t) R_j(0) \rangle. \tag{31}$$

Then, the variational problem of Eq. 25 becomes a generalized eigenvalue problem

$$\sum_{j=1}^{3N} C_{i,j}(t_0 + \tau) f_{p,j} = \exp(-\lambda_p \tau) \sum_{j=1}^{3N} C_{i,j}(t_0) f_{p,j}. \tag{32}$$

The orthonormal condition of Eq. 18 for X_p is written as

$$\sum_{i=1}^{3N} \sum_{j=1}^{3N} f_{p,i} C_{i,j}(t_0) f_{p,j} = \delta_{p,q}. \tag{33}$$

Equations 32 and 33 determine the relaxation rates λ_p and the corresponding relaxation modes $f_{p,i}$. We chose the indices of λ_p so that $0 < \lambda_1 \leq \lambda_2 \leq \cdots$ holds. Here, the relation

$$T_t(Q|Q') P_{\mathrm{eq}}(Q') = T_t(Q'|Q) P_{\mathrm{eq}}(Q), \tag{34}$$

which is equivalent to the detailed balance condition of Eq. 14 with $\epsilon Q = Q$, and the Markovian property

$$\sum_{Q'} T_{t_1}(Q|Q') T_{t_2}(Q'|Q'') = T_{t_1+t_2}(Q|Q'') \tag{35}$$

are used.

The inverse transformation of Eq. 28 is given by

$$R_i(t_0/2; Q) = \sum_{p=1}^{3N} g_{i,p} X_p(Q) \tag{36}$$

with

$$g_{i,p} = \sum_{j=1}^{3N} C_{i,j}(t_0) f_{p,j}. \tag{37}$$

The time correlation functions of R_i are reproduced by

$$\begin{aligned} \langle R_i(t) R_j(0) \rangle &= \sum_p \sum_q g_{i,p} g_{j,q} \left\langle X_p(t - t_0) X_q(0) \right\rangle, \\ &\simeq \sum_p g_{i,p} g_{j,p} \exp\left[-\lambda_p (t - t_0)\right], \\ &= \sum_p \tilde{g}_{i,p} \tilde{g}_{j,p} \exp(-\lambda_p t), \end{aligned} \tag{38}$$

for $t \geq t_0$. Here,

$$\tilde{g}_{i,p} = g_{i,p} \exp(\lambda_p t_0/2). \tag{39}$$

Because we are considering position coordinates only, the detailed balance condition yields the following consequences: $C(t)$ is a symmetric matrix, $C_{i,j}(t) = C_{j,i}(t)$; $\{\lambda_p\}$ are real and positive, which corresponds to pure relaxation. We refer to this method as the "RMA method with a single evolution time," which is $t_0/2$.

In practice, the time correlation matrices for the two different times are calculated through simulations. Then, by solving the generalized eigenvalue problem, $\{\lambda_p\}$ and $\{X_p\}$ are obtained from the eigenvalues and eigenvectors, respectively. To examine the validity of the present analysis, the autocorrelation functions $C_{i,i}(t)$ are reconstructed from the estimated eigenvalues and eigenvectors and are compared with those directly calculated via simulation.

Herein, we comment on the trial function. When RMA was first introduced to a spin system, states of spins on a lattice were used as the trial function (Takano and Miyashita (1995)). When RMA was first introduced to polymer systems, the coordinates of polymers were used as the trial function (Koseki et al. (1997); Hirao et al. (1997)). In polymer systems, the Rouse modes, which were derived from the theory of polymer physics (Doi and Edwards 1986), correspond to the relaxation modes. Rouse modes are given as linear combinations of coordinates. Thus, when RMA was applied to polymer systems, the modes obtained by RMA were compared with the Rouse modes. In protein systems, PCA using coordinates has been widely used. In PCA, the eigenvalue problem of the covariance matrix of coordinates is solved. Therefore, when we first applied RMA to a hetero polymer system (protein system), it seemed to be better to use coordinates as trial functions. The results of RMA and PCA were directly compared with each other. Recently, we have proposed to use physical quantities with slow motions as the trial functions and PCRMA and two-step RMA have been introduced (see the "Improvement of RMA" section). However, RMA using coordinates as the trial functions has an advantage that we can easily convert the information on the slow relaxation modes to the information in coordinate space.

RMA for protein systems

In homopolymer systems, relaxation of the positions of a polymer relative to the center of the mass is investigated. This means that the translational degrees of freedom are removed from the coordinates of the polymer. Because the rotational degrees of freedom remain, the rotational relaxation of the polymer is observed as slow relaxations. In protein systems, it is of interest to evaluate fluctuations of the conformations of a biomolecule around its average conformation. Thus, the translational and rotational degrees of freedom are removed from the sampled conformations of a biomolecule. In practice, treatment of the generalized-eigenvalue problem for removing the translational degrees of freedom in the homopolymer system was given by Koseki et al. (1997). Herein, we explain how to treat the generalized eigenvalue problem for removing the translational and rotational degrees of freedom when using the coordinates for the trial function (Mitsutake et al. 2011). The point of this process is that the generalized eigenvalue problem for real symmetric matrices can be easily solved numerically if the matrices are positive definite. Therefore, we shift the zero eigenvalues to finite positive values without changing the other eigenvalues and the corresponding eigenvectors.

A schematic illustration of the process for RMA using coordinates for the trial function is shown in Fig. 1. First, we remove the translational and rotational degrees of freedom as well as conduct PCA (Eckart 1935; McLachlan 1979). After the average structure converges, the origin of the coordinate system is chosen to be the center of the mass of the average positions, $\langle \boldsymbol{r}_i \rangle$ with $i = 1, \ldots, N$, and the axes of the coordinate system are chosen to be the principal axes of the moment of the inertia tensor of the average positions. We calculate $C_{i,j}(t) = \dfrac{C_{i,j}(t) + C_{j,i}(t)}{2}$ and $C'(t)$:

$$C'(t) = C(t) + \sum_{\alpha=x,y,z} \exp(-\lambda_\alpha^{\mathrm{tr}}(t - t_0)) \boldsymbol{d}_\alpha^{\mathrm{tr}} \boldsymbol{d}_\alpha^{\mathrm{tr}\,\mathrm{T}} + \sum_{\alpha=x,y,z} \exp(-\lambda_\alpha^{\mathrm{rot}}(t - t_0)) \boldsymbol{d}_\alpha^{\mathrm{rot}} \boldsymbol{d}_\alpha^{\mathrm{rot}\,\mathrm{T}}, \tag{40}$$

where $\boldsymbol{d}_x^{\mathrm{tr}}$, $\boldsymbol{d}_y^{\mathrm{tr}}$, and $\boldsymbol{d}_z^{\mathrm{tr}}$ are unit vectors given by

$$\begin{aligned} \boldsymbol{d}_x^{\mathrm{tr}} &= \frac{1}{\sqrt{N}}(1, 0, 0, 1, 0, 0, \cdots, 1, 0, 0)^{\mathrm{T}}, \\ \boldsymbol{d}_y^{\mathrm{tr}} &= \frac{1}{\sqrt{N}}(0, 1, 0, 0, 1, 0, \cdots, 0, 1, 0)^{\mathrm{T}}, \\ \boldsymbol{d}_z^{\mathrm{tr}} &= \frac{1}{\sqrt{N}}(0, 0, 1, 0, 0, 1, \cdots, 0, 0, 1)^{\mathrm{T}}, \end{aligned} \tag{41}$$

and $\boldsymbol{d}_x^{\mathrm{rot}}$, $\boldsymbol{d}_y^{\mathrm{rot}}$, and $\boldsymbol{d}_z^{\mathrm{rot}}$ are unit vectors given by

$$\begin{aligned} \boldsymbol{d}_x^{\mathrm{rot}} &= \frac{1}{\sqrt{\sum_{i=1}^N (\langle z_i \rangle^2 + \langle y_i \rangle^2)}} \\ &\quad \times (0, -\langle z_1 \rangle, \langle y_1 \rangle, 0, -\langle z_2 \rangle, \langle y_2 \rangle, \cdots, 0, -\langle z_N \rangle, \langle y_N \rangle)^{\mathrm{T}}, \\ \boldsymbol{d}_y^{\mathrm{rot}} &= \frac{1}{\sqrt{\sum_{i=1}^N (\langle z_i \rangle^2 + \langle x_i \rangle^2)}} \\ &\quad \times (\langle z_1 \rangle, 0, -\langle x_1 \rangle, \langle z_2 \rangle, 0, -\langle x_2 \rangle, \cdots, \langle z_N \rangle, 0, -\langle x_N \rangle)^{\mathrm{T}}, \text{ and} \\ \boldsymbol{d}_z^{\mathrm{rot}} &= \frac{1}{\sqrt{\sum_{i=1}^N (\langle y_i \rangle^2 + \langle x_i \rangle^2)}} \\ &\quad \times (-\langle y_1 \rangle, \langle x_1 \rangle, 0, -\langle y_2 \rangle, \langle x_2 \rangle, 0, \cdots, -\langle y_N \rangle, \langle x_N \rangle, 0)^{\mathrm{T}}. \end{aligned} \tag{42}$$

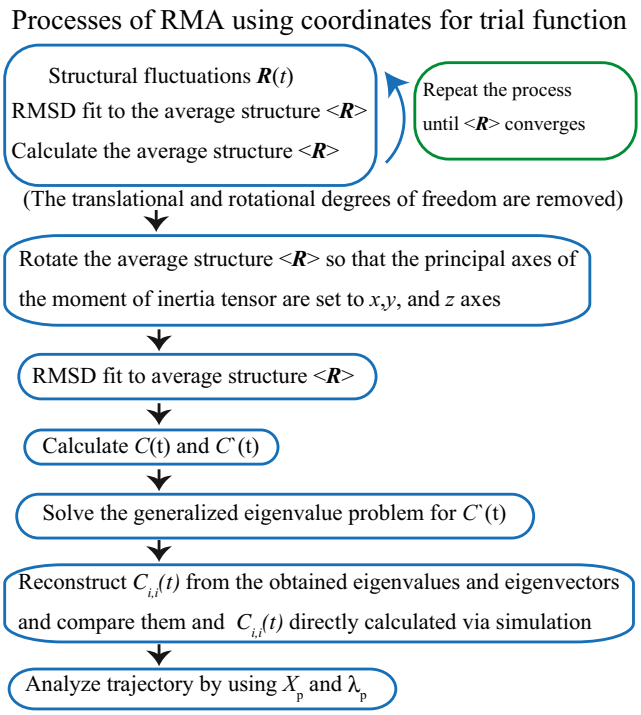

Fig. 1 Schematic illustration of the RMA process using the coordinate $\boldsymbol{R}$ for the trial function

The values of $\lambda_\alpha^{\mathrm{tr}}$ and $\lambda_\alpha^{\mathrm{rot}}$ are usually set to zero. These unit vectors satisfy the following relations:

$$\boldsymbol{d}_\alpha^a \cdot \boldsymbol{d}_\beta^b = \boldsymbol{d}_\alpha^{a\mathrm{T}} \boldsymbol{d}_\beta^b = \delta_{\alpha,\beta}\delta_{a,b} \tag{43}$$

and

$$C(t)\boldsymbol{d}_\alpha^a = 0, \tag{44}$$

where $\alpha, \beta = x, y, z$ and $a, b = \mathrm{tr}, \mathrm{rot}$. Then, we solve the generalized eigenvalue problem for $C'(t_0+\tau)$ and $C'(t_0)$, $C'(t_0+\tau)\boldsymbol{v}'_p = \exp(-\lambda'_p\tau)C'(t_0)\boldsymbol{v}'_p$, with the orthonormal condition $\boldsymbol{v}'^{\mathrm{T}}_p C'(t_0)\boldsymbol{v}'_q = \delta_{p,q}$. The unit vectors $\boldsymbol{d}_\alpha^a$ are eigenvectors of this generalized eigenvalue problem with eigenvalues $\exp(-\lambda_\alpha^a\tau)$. We denote $\boldsymbol{f}'_p$ as the eigenvectors other than $\boldsymbol{d}_\alpha^a$. Because $\boldsymbol{d}_\alpha^{a\mathrm{T}} C'(t)\boldsymbol{f}'_p = \exp(-\lambda_\alpha^a(t-t_0))\boldsymbol{d}_\alpha^{a\mathrm{T}}\boldsymbol{f}'_p = 0$, $C'(t)\boldsymbol{f}'_p = C(t)\boldsymbol{f}'_p$ holds. Therefore, $\boldsymbol{f}'_p$ are identical with the eigenvectors $\boldsymbol{f}_p = (f_{p,1}, f_{p,2}, \ldots, f_{p,3N})^{\mathrm{T}}$ of the generalized-eigenvalue problem for $C(t_0+\tau)$ and $C(t_0)$ with the same eigenvalues $\exp(-\lambda_p\tau)$. Thus, $\boldsymbol{f}_p$ and $\exp(-\lambda_p\tau)$ can be obtained by solving the generalized eigenvalue problem for $C'(t_0+\tau)$ and $C'(t_0)$, which are real symmetric positive definite matrices.

After obtaining relaxation modes and rates, we confirm whether or not the slow relaxation modes and rates obtained using τ and t_0 are appropriate. For this purpose, the convergences of slow relaxation times as a function of τ are examined. The autocorrelation functions $C_{i,i}(t)$ are reconstructed from the estimated eigenvalues and eigenvectors and are compared with those directly calculated via simulation (especially the slow relaxation behavior). After examining the validity, we use the obtained relaxation modes and rates for analysis.

Improvement of RMA

Selection of τ and t_0 and relevant quantities for the trial function

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 simplification, we here consider the case of one physical quantity, R. From the variational problem of Eqs. 24 and 25, the relaxation time $1/\lambda$ is obtained from the gradient of the straight line connecting two points at $t = t_0$ and $t = t_0+\tau$ in the semi-log plot of the correlation function $C(t) = \langle R(t)R(0)\rangle - \langle R\rangle^2$ versus t, as shown in Fig. 2a. If the time correlation function of the physical quantity contains several $\{1/\lambda_p\}$, and if we choose $t_0 = 0$ (tICA case) or a small t_0 and small τ, as shown in Fig. 2a (green line), 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, as shown in Fig. 2a (blue line). However, the choice of a longer t_0 and τ is also limited, because of the decreasing accuracy of the time correlation function over long time periods. Therefore, we must choose the appropriate t_0 and τ.

We can improve the RMA explained above by using two different approaches: introduction of multiple evolution times and using the different relevant physical quantities obtained from coordinates (and velocities) for the trial function. For the first improvement, we describe two types of methods with multiple evolution times, as shown in Fig. 2b, c. (The detailed descriptions are given by Nagai et al. (2013), Natori and Takano (2017), and Karasawa et al. (2017).) For the second improvement, we describe the

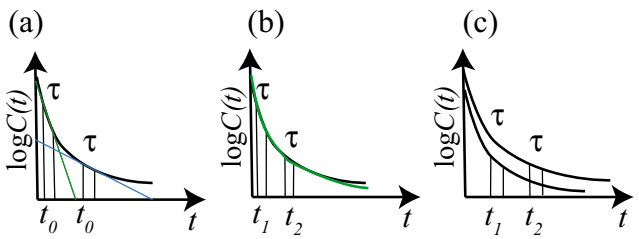

Fig. 2 Schematic illustration of RMA with a single evolution time t_0 (**a**), and multiple evolution times (1) using t_1 and t_2 (**b**) and (2) using t_i (**c**)

PCRMA (Nagai et al. 2013), in which the relevant physical quantities for the trial function are given by the PC modes with large structural fluctuations and the two-step RMA (Natori and Takano 2017; Karasawa et al. 2017), which are in turn given by the slowest relaxation modes roughly obtained by RMA. Moreover, the MSRMA (Mitsutake and Takano 2015) is also proposed. We will describe these two improved RMAs in detail below.

RMA with multiple evolution times

RMA with multiple evolution times t_1 and t_2(1)

In this method, the following trial functions are used as approximate relaxation modes:

$$X_p(Q) = \sum_{i=1}^{3N} f_{p,i}^1 R_i(t_1/2;\, Q) + \sum_{i=1}^{3N} f_{p,i}^2 R_i(t_2/2;\, Q). \quad (45)$$

Note that two evolution times, $t_1/2$ and $t_2/2$, are used instead of a single evolution time, $t_0/2$. Because the contributions of faster modes in $\boldsymbol{R}$ time-evolved for $t_1/2$ and those for $t_2/2$ are different, the approximate relaxation modes can extract the faster modes, which cannot be extracted by the approximate relaxation modes using a single evolution time (see Fig. 2b). Using Eq. 45 as a trial function for the variational problem, the following generalized eigenvalue problem is obtained:

$$\sum_{j=1}^{6N} C_{i,j}(t_0+\tau) f_{p,j} = \exp(-\lambda_p \tau) \sum_{j=1}^{6N} C_{i,j}(t_0) f_{p,j}, \quad (46)$$

with $\boldsymbol{f}_p = (\boldsymbol{f}_p^{1\,\mathrm{T}}, \boldsymbol{f}_p^{2\,\mathrm{T}})^{\mathrm{T}}$. Here, $C(t)$ is a $6N \times 6N$ matrix defined by

$$C(t) = \begin{pmatrix} C^{1,1}(t) & C^{1,2}(t) \\ C^{2,1}(t) & C^{2,2}(t) \end{pmatrix}, \quad (47)$$

and $C^{\mu_1,\mu_2}(t)$ is an $3N \times 3N$ matrix defined by

$$C_{i,j}^{\mu_1,\mu_2}(t) = \left\langle R_i\left(\frac{t_{\mu_1}}{2} + \frac{t_{\mu_2}}{2} + t\right) R_j(0) \right\rangle, \quad (48)$$

where $\mu_1, \mu_2 = 1$ or 2. The orthonormal condition is written as

$$\sum_{i=1}^{6N}\sum_{j=1}^{6N} f_{p,i} C_{i,j}(0) f_{p,j} = \delta_{p,q}. \quad (49)$$

The inverse transformation of Eq. 45 is given by

$$R_i(t_1/2;\, Q) = \sum_{p=1}^{6N} g_{i,p}^1 X_p(Q)$$
$$R_i(t_2/2;\, Q) = \sum_{p=1}^{6N} g_{i,p}^2 X_p(Q) \quad (50)$$

with

$$g_{i,p} = \sum_{j=1}^{6N} C_{i,j}(0) f_{p,j}, \quad (51)$$

where $\boldsymbol{g}_p = \left(\boldsymbol{g}_p^{1\,\mathrm{T}}, \boldsymbol{g}_p^{2\,\mathrm{T}}\right)^{\mathrm{T}}$. The time correlation functions of R_i are reproduced by

$$\langle R_i(t) R_j(0)\rangle \simeq \sum_{p=1}^{6N} \tilde{g}_{i,p}^{av} \tilde{g}_{j,p}^{av} \exp\left(-\lambda_p t\right), \quad (52)$$

where

$$\tilde{g}_{i,p}^{av} = (\exp(\lambda_p t_1/2) g_{i,p}^1 + \exp(\lambda_p t_2/2) g_{i,p}^2)/2. \quad (53)$$

RMA with multiple evolution times t_i (2)

When the relevant physical quantities $\boldsymbol{R}$ in the trial function exhibit different relaxations, it is preferable to use different evolution times for the different physical quantities, as shown in Fig. 1c. 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 "RMA with multiple evolution times $\{t_i/2\}$." In this method, we use the following trial function:

$$X_p(Q) = \sum_{i=1}^{3N} f_{p,i} R_i(t_i/2;\, Q). \quad (54)$$

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. 54 would yield a superior approximation for larger t_i values.

The variational problem becomes a generalized-eigenvalue problem:

$$\sum_{j=1}^{3N} C_{i,j}\left(\frac{t_i+t_j}{2} + \tau\right) f_{p,j} = \exp(-\lambda_p \tau) \sum_{j=1}^{3N} C_{i,j}\left(\frac{t_i+t_j}{2}\right) f_{p,j}. \quad (55)$$

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

$$\sum_{i=1}^{3N}\sum_{j=1}^{3N} f_{p,i} C_{i,j}\left(\frac{t_i+t_j}{2}\right) f_{q,j} = \delta_{p,q}. \quad (56)$$

Equations 54, 55, and 56 determine the relaxation rates λ_p and the corresponding relaxation modes. We chose the

indices of λ_p such that $0 < \lambda_1 \leq \lambda_2 \leq \cdots$ holds. The inverse transformation of Eq. 54 is given by

$$R_i(t_i/2;\, Q) = \sum_{p=1}^{3N-6} g_{i,p} X_p(Q), \tag{57}$$

with

$$g_{i,p} = \sum_{j=1}^{3N} C_{i,j} \left(\frac{t_i + t_j}{2} \right) f_{p,j}. \tag{58}$$

The time correlation functions of R_i are given by

$$\begin{aligned} \langle R_i(t) R_j(0) \rangle &= \sum_p \sum_q g_{i,p} g_{j,q} \left\langle X_p \left(t - \frac{t_i + t_j}{2} \right) X_q(0) \right\rangle, \\ &\simeq \sum_p g_{i,p} g_{j,p} \exp\left[-\lambda_p \left(t - \frac{t_i + t_j}{2} \right) \right], \\ &= \sum_p \tilde{g}_{i,p} \tilde{g}_{j,p} \exp(-\lambda_p t), \end{aligned} \tag{59}$$

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

$$\tilde{g}_{i,p} = g_{i,p} \exp(\lambda_p t_i/2). \tag{60}$$

RMAs to automatically reduce the degrees of freedom of relevant quantities for the trial function

RMA requires relatively high statistical precision of the time correlation matrices because of treatment for the generalized eigenvalue problem; thus, it is difficult for RMA to handle a large number of degrees of freedom directly. We must therefore reduce the number of degrees of freedom automatically.

In an original RMA, the coordinates (and velocity) are used for the trial function. The results may change depending on which relevant quantities are used for the trial function because their correlation functions are fitted using t_0 and τ. (For the Markov state model, the dependence of relaxation times on the selection of states is discussed in Swope et al. (2004) and Pérez-Hernández et al. (2013).) It is better to use the relevant quantities that include the slow behavior. For the second improvement, we describe the PCRMA in which the relevant quantities are given by the PC modes with large structural fluctuations, and the two-step RMA in which the quantities are given by the slowest relaxation modes roughly obtained by the first RMA. A schematic illustration of PCRMA and two-step RMA is given in Fig. 3.

PCRMA

To apply RMA to a protein system by reducing its degrees of freedom, we proposed an improved method, which is referred to as the PCRMA method (Nagai et al. 2013). In this method, PCA is carried out first, and then, RMA is applied to a small number of principal components with large fluctuations ($\boldsymbol{\Phi} = (\Phi_1, \Phi_2, \cdots, \Phi_{N_c})^T$). We use the following function as an approximate relaxation mode:

$$X_p(Q) = \sum_{i=1}^{N_c} f_{p,i} \Phi_i(t_0/2;\, Q). \tag{61}$$

Because the degrees of freedom is reduced to N_c and the relevant quantities with large variance tend to have slow relaxations, the slow relaxation times can be estimated by setting t_0 and τ as large values. Note that because the selected principal components also contain faster relaxation modes, as shown in Fig. 4, Nagai et al. (2013) also combined PCRMA with the RMA using multiple evolution times (1) explained above. Note that in PCRMA, if the N_cth or more PC modes (with relatively small fluctuations) have slow relaxation, the slow behaviors may not be extracted; thus, there is a possibility that the slow relaxations would not be estimated with small structural fluctuations.

Two-step RMA

Using a similar process to that of PCRMA, we proposed a two-step RMA method (Natori and Takano 2017; Karasawa et al. 2017). Based on our experience, the slow $\{X_p\}$ obtained from the conventional RMA with small t_0 and τ contains the true slow $\{X_p\}$ (Mitsutake et al. 2011), although the $\{1/\lambda_p\}$ values are underestimated. The slow relaxation modes obtained by the first RMA may contain the true slow relaxation modes. Thus, we use the slow relaxation modes roughly obtained from the first RMA as the relevant quantities for the trial function. In this technique, RMA with a single evolution time using small t_0 and τ is implemented first, and $\{X_p\}$ and $\{\lambda_p\}$ are roughly estimated. We then 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 also use the previously presented technique of RMA with multiple evolution times (2), because the characteristic time scales of the $\{X_p\}$ obtained from the first RMA are roughly given by the relaxation times $\{1/\lambda_p\}$. In the second RMA, we use the following trial function:

$$X'_u(Q) = \sum_{p=1}^{N_\mathrm{m}} f'_{u,p} X_p(t'_p/2;\, Q). \tag{62}$$

Here, $X_p(Q)$ is the relaxation mode obtained from the first RMA and t'_p is determined from $1/\lambda_p$. A detailed explanation is given by Natori and Takano (2017) and Karasawa et al. (2017).

In the second RMA, the time interval τ can be chosen to be large, because the number of degrees of freedom is reduced and the physical quantities $\{X_p\}$

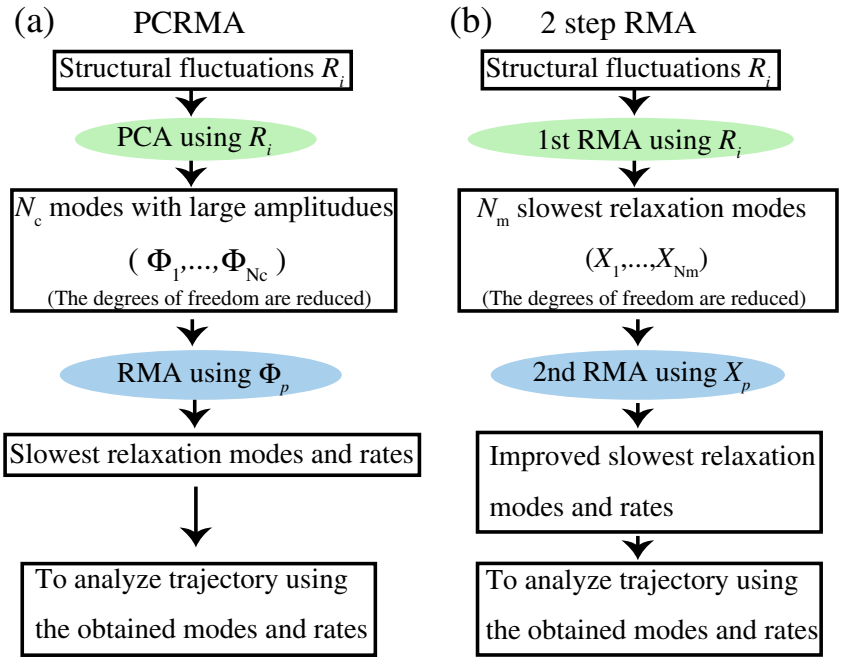

Fig. 3 Schematic illustration of PCRMA (**a**) and two-step RMA (**b**)

exhibit slow relaxations. Using the second RMA, the estimation accuracy of the relaxation modes and times can be improved.

Markov state RMA

As mentioned above, in RMA, 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. From this point of view, RMA is related to Markov state models. Herein, we consider the relation between RMA and Markov state models and propose the new method of MSRMA.

In the simplest Markov state model, the phase space of the system, where only the position coordinates are considered, is divided into clusters (subsets) S_i, $i = 1, \ldots, n$. First, the joint probability $\bar{P}_{i,j}(\tau) = P(Q \in S_i, \tau; Q \in S_j, 0)$ that the state of the system Q is in the jth cluster at time 0 and is in the ith cluster at time $\tau > 0$ is calculated in a simulation. Second, the transition probability $\bar{T}_{i,j}(\tau)$ that the state of the system is found in the ith cluster after time τ starting from a state in the jth cluster is calculated by

$$\bar{T}_{i,j}(\tau) = \bar{P}_{i,j}(\tau)/\bar{p}_j, \tag{63}$$

where $\bar{p}_j = P(Q \in S_j)$ is the probability that the state of the system is found in the jth cluster, which is estimated in the simulation. Then, by solving the eigenvalue problem

$$\bar{\boldsymbol{f}}_p{}^{\mathrm{T}}\bar{T}(\tau) = \bar{\boldsymbol{f}}_p{}^{\mathrm{T}}\bar{\Lambda}_p \tag{64}$$

for the transition matrix $\bar{T}(\tau) = (\bar{T}_{i,j}(\tau))$, the pth eigenvector $\bar{\boldsymbol{f}}_p$ and its eigenvalue $\bar{\Lambda}_p$ are obtained. The eigenvector $\bar{\boldsymbol{f}}_1 \propto (1, 1, \ldots, 1)^{\mathrm{T}}$ corresponds to the equilibrium state and its eigenvalue $\bar{\Lambda}_1 = 1$. Other eigenvectors $\bar{\boldsymbol{f}}_p$ represent structural transitions and the corresponding eigenvalues $\bar{\Lambda}_p$ give their relaxation time scales $\bar{\tau}_p$ as

$$\bar{\tau}_p = -\frac{\tau}{\ln \bar{\Lambda}_p}. \tag{65}$$

Note that in the Markov description, it is important that the states are defined in a kinetically meaningful way (Swope et al. 2004; Pérez-Hernández et al. 2013). We need to define the states that are classified by order parameters representing the dynamics and kinetics of the system. Even with a good choice of states, in order for a Markov description of the process to be accurate, the time interval τ should also be chosen carefully. In other words, for the Markov description to work, the time interval of the transition matrix τ must be chosen appropriately so that

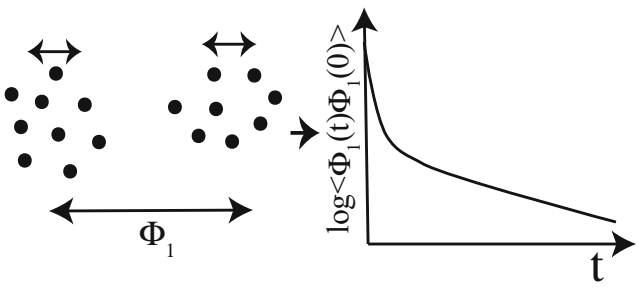

Fig. 4 Schematic illustration for PCRMA

it is as large as the slowest relaxation time of the states. When plotting $\bar{\tau}_p$ as a function of τ, $\bar{\tau}_p$ slowly converges to the appropriate time scale when τ is increased. In addition, when a much longer τ than the slowest relaxation time of the states is used, the Markov state model is not expected to be accurate. Thus, we usually set the time interval τ to the value when the variation of τ_p is sufficiently flat (Swope et al. 2004; Pérez-Hernández et al. 2013).

The abovementioned procedure of the Markov state model is related to the following procedure of RMA. We consider an approximate relaxation mode given by

$$\bar{X}_p = \sum_{i=1}^{n} f_{p,i}\delta_i(t_0/2;\, Q), \tag{66}$$

where $\delta_i(t;\, Q)$ is defined in the same way as $R_i(t;\, Q)$ in Eq. 29 from $\delta_i(Q)$ given as a function of the state Q of the system by

$$\delta_i(Q) = \begin{cases} 1 \text{ for } Q \in S_i, \\ 0 \text{ for } Q \notin S_i. \end{cases} \tag{67}$$

Then, the generalized eigenvalue problem is given by

$$\sum_j \bar{C}_{i,j}(t_0+\tau)\bar{f}_{p,j} = \mathrm{e}^{-\bar{\lambda}_p\tau}\sum_j \bar{C}_{i,j}(t_0)\bar{f}_{p,j}, \tag{68}$$

with

$$\sum_{i,j} \bar{f}_{p,i}\bar{C}_{i,j}(t_0)\bar{f}_{q,j} = \delta_{p,q}, \tag{69}$$

where

$$\bar{C}_{i,j}(t) = \langle \delta_i(t)\delta_j(0)\rangle. \tag{70}$$

According to the definition of $\delta_i(Q)$, it follows that $\bar{C}_{i,j}(t)$ is the joint probability $\bar{P}_{i,j}(t)$.

If we set $t_0 = 0$, the generalized eigenvalue problem (68) becomes the eigenvalue problem (64) with $\bar{\Lambda}_p = \mathrm{e}^{-\bar{\lambda}_p\tau}$ or $\bar{\tau}_p = 1/\bar{\lambda}_p$, because $\bar{C}(0) = \mathrm{diag}(\bar{p}_1, \ldots, \bar{p}_n)$ and $\bar{C}(\tau)\bar{C}(0)^{-1} = \bar{T}(\tau)$. Thus, the Markov state model is a special case of MSRMA with $t_0 = 0$.

Because $\delta_i(t_0/2;\, Q)$ in Eq. 66 reduces the contributions of faster modes in $\delta_i(Q)$, the solutions of the generalized eigenvalue problem (68) provides better approximations to the slow relaxation modes and rates as t_0 becomes larger. Therefore, the relaxation times $\bar{\tau}_p$ obtained by the Markov state model are expected to be improved by solving Eq. 68 with $t_0 > 0$ rather than Eq. 64.

Application of RMA to a system with large conformational changes

In this section, we apply RMA to a protein system simulation to show the effectiveness of RMA. The selection of order parameters in simulations is important to analyze the trajectory. PCA, which is a static analysis method, extracts large structural fluctuations from simulations, and the obtained PC mode is used to obtain the order parameters. Moreover, it has now become possible to perform long simulations such as those of unfolded and folded protein structures, and when the simulation involves large structural changes, the difference between local minimum-energy states is relatively small compared with that between the folded and unfolded states. In this case, it is difficult for PCA to extract the effective modes or order parameters to accurately identify the local minimum-energy states. By contrast, RMA extracts slow relaxation modes. It is thought that the local minimum-energy states are usually stable so that the system remains in this state for a long time during a simulation. The order parameters with slow relaxation may correspond to the directions between local minimum-energy states. Thus, slow relaxation modes may be suitable order parameters to identify local minimum-energy states and the transitions between them. To validate this concept, we applied RMA to the 10-residue peptide, chignolin in water near its folding transition temperature.

The detailed results are described in Mitsutake et al. (2011). Chignolin consists of a 10-amino acid sequence, GYDPETGTWG and adopts a β-hairpin turn structure (Honda et al. 2004). Several simulations of chignolin have been reported to date (Satoh et al. 2006; Suenaga et al. 2007; Harada and Kitao 2011; Kührova et al. 2012; Okumura 2012). Previous research has shown that chignolin has a stable (native) and a misfolded state, which are both found as hairpin-like structures (see Fig. 5c). These two states have a common turn structure from Asp3 to Glu5 but slightly different hydrogen bond patterns. RMA requires a relatively high level of statistical precision for the time correlation matrices and therefore requires a long simulation where many transitions between local minimum-energy states occur. In addition, we sought to analyze the system with large conformational changes. Thus, we performed a 750-ns molecular dynamics simulation of chignolin in aqueous solution near the transition temperature from an extended structure (Case et al. 2014). We observed many transitions among structures, including the native, misfolded, and unfolded states, by performing the simulation at 450 K. We used the coordinates of C_α atoms on the backbone as coordinates so that the degrees of freedom were 30. After removing the translational and rotational motions from the coordinates of C_α atoms, PCA and RMA were carried out on the coordinates of C_α atoms (see Fig. 1). For RMA, we set t_0 and τ to 10.0 and 20.0 ps, respectively.

Figure 5 shows the free-energy surfaces obtained from PCA (a) and RMA (b). From the free-energy surface of PCA, the native and misfolded states were not distinguished because the conformational difference between them is much smaller than the conformational fluctuations of the system (the third PC mode distinguished the native and

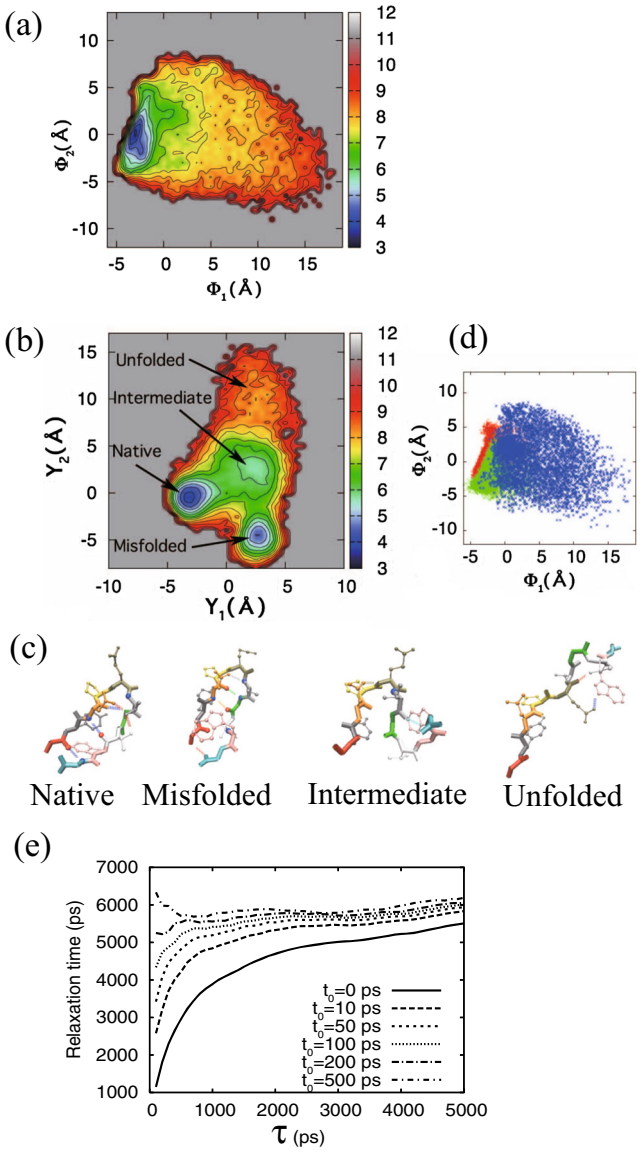

Fig. 5 The free-energy surfaces for **a** the first PC mode Φ_1 and the second PC mode Φ_2, and for **b** the first slowest RM and the second slowest RM in the case of $t_0 = 10.0$ ps and $\tau = 20.0$ ps. **c** Snapshots of the native, misfolded, intermediate, and unfolded states classified by RMA, and **d** distributions for the native (red), misfolded (green), and intermediate (blue) states on the free-energy surface of the first PC mode and the second PC mode. **e** Relaxation times of the second relaxation mode obtained by MSRMA as a function of the time interval τ. In **e**, the line of t_0 ps corresponds to the results of a simple Markov state model. The figure was reproduced from Mitsutake and Takano (2015)

misfolded states). By contrast, in RMA, the transition between the native and misfolded structures is slow, and the slowest relaxation mode was found to be the axis distinguishing them. This analysis showed that the slow relaxation mode is a good order parameter to distinguish the native and misfolded structure. Interestingly, we could also identify the intermediate structure. By extracting the structures in the center part of the free-energy surface shown in Fig. 5b, the cluster was formed with a turn structure common to the native and misfolded structures. Because the structures at both terminals fluctuate, a cluster of intermediate structures forming a turn is also obtained, while ignoring the fast relaxing movement of both terminals. The upper part of the free-energy surface shown in Fig. 5b corresponded to the extended structure. Figure 5c shows the characteristic structures for the four states. When plotting the points for the obtained intermediate structure on the free-energy surface of PCA in Fig 5d, the points were distributed widely because both terminals fluctuate. Thus, RMA can identify the characteristic structure, even when it is only partially formed. From the free-energy surface obtained by RMA, it is clarified that chignolin folds to the native or misfolded structures through the intermediate (turn) structure from the extended structures.

Because the structures were classified into a smaller number of states using the free-energy surface obtained by RMA, we then applied the Markov state model and MSRMA to analyze these four states: native, misfolded, intermediate, and unfolded states. Figure 5e shows the relaxation time $\tau_p = 1/\lambda_p$ obtained by MSRMA as a function of τ when $t_0 = 0, 10, 50, 100, 200$, and 500 ps. Because the first eigenvector corresponds to the steady state with infinite relaxation time $\tau_1 = \infty$, we show the second slowest relaxation times. The line of $t_0 = 0$ corresponds to the results of a simple Markov state model. In the case of $t_0 = 0$, the τ_p values slowly approach the appropriate time scale, i.e., the values for plateau regions or peak values of the solid lines, when τ is increased. For the lines of $t_0 > 0$, the values of τ_p quickly approach the appropriate time scale, i.e., those corresponding to the values for plateau regions or peak values. Thus, the slow relaxation times can be improved when applying MSRMA with $t_0 > 0$, which is introduced to reduce the relative weight of the faster modes.

Overall, RMA can be used to effectively analyze long simulations at room temperature and is also useful for investigating systems with large conformational changes, such as intrinsically disordered proteins and protein folding.

Conclusions

In this paper, we have reviewed the method and application of RMA, a dynamic analysis method for protein simulations. We described the definition of relaxation modes and rates, which correspond to the left eigenfunctions and eigenvalues of the time evolution operator of the master equation of the system, respectively. After providing the definition, we explained how to estimate the slow relaxation modes and rates from simulation data. We also summarized several new RMAs proposed, including RMA with multiple

evolution times, PCRMA, two-step RMA, and MSRMA. Finally, to demonstrate the effectiveness of RMA, we briefly presented the analysis results of the unfolding/folding simulation of the 10-residue peptide chignolin detected near the transition temperature. The simulation results showed that the relaxation mode is a good order parameter for not only extracting the transition between the native state and misfolded state but also for identifying the intermediate state, which is partially folded. This suggests that RMA is suitable to investigate a system with large structural changes and naturally denatured protein systems. Although RMA is efficient for a longer simulation than the longest relaxation time of the system, it can also extract rare events in a finite-time simulation such as that conducted at the microsecond scale. By examining the extent to which the correlation function can be reconstructed, we can clarify the information that can be obtained on dynamics using the obtained relaxation modes and rates. Theoretical studies to compare data of the Markov state model with experimental data from nuclear magnetic resonance and neutron scattering analyses have emerged recently (Xia et al. 2013; Lindner et al. 2013; Zheng et al. 2013; Bowman et al. 2014). In the future, it will also be important to interpret the theoretical relationships in light of experimental data.

Acknowledgements The authors would like to thank Mr. Toshiki Nagai, Mr. Taku Yamamoto, Mr. Yuta Koizumi, Mr. Satoshi Natori, and Mr. Naoyuki Karasawa at Keio University for fruitful discussions.

Compliance with ethical standards

Funding information This work was supported by JST PRESTO (JPMJPR13LB). This work was also partially supported by a Grant-in-Aid for Scientific Research (C) (No. 24540441) from the Japan Society for the Promotion of Science.

Conflict of Interests Ayori Mitsutake declares that he has no conflicts of interest. Hiroshi Takano declares that he has no conflicts of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any og the authors.

References

Abagyan R, Argos P (1992) Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. J Mol Biol 225:519–532

Amadei A, Linssen ABM, Berendsen HJC (1993) Essential dynamics of proteins. Proteins Struct Funct Genet 17:412–425

Baher I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 2:173

Bowman GR, Pande VS, Noé F (eds) (2014) An introduction to Markov state models and their application to long timescale molecular simulation. Springer, Dordrecht

Brooks B, Karplus M (1983) Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci USA 80:6571

Buchete N, Hummer G (2008) Coarse master equations for peptide folding dynamics. J Phys Chem B 112:6057

Case DA, Babin V, Betz RM, Cai Q, Cerutti DS, Cheatham IIITE, Darden TA, Duke RE, Gohlke H, Götz AW, Gusarov S, Homeyer N, Janowski P, Kaus J, Kolossváry I, Kovalenko A, Lee TS, Le Grand S, Luchko T, Luo R, Madej B, Merz KM, Paesani F, Roe DR, Roitberg A, Sagui C, Salomon-Ferrer R, Seabra G, Simmerling CL, Smith W, Swails J, Walker RC, Wang J, Wolf RM, Wu X, Kollman PA (2014) AMBER 14. University of California, San Francisco

Chodera JD, Swope WC, Pitera JW, Dill KA (2006) Long-time protein folding dynamics from short-time molecular dynamics simulations. Multiscale Model Simul 5:1214

Chodera JD, Singhal N, Vande VS, Dill KA, Swope WC (2007) Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. J Chem Phys 126:155101

Chodera JD, Noé F (2014) Markov state models of biomolecular conformational dynamics. Curr Opin Struct Biol 25:135

Cui Q, Bahar I (eds) (2005) Normal mode analysis: theory and applications to biological and chemical systems. Chapman & Hall/CRC, London

Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE (2012) Biomolecular simulation: a computational microscope for molecular biology. Annu Rev Biophys 41:429

de Gennes PG (1984) Scaling concepts in polymer physics. Cornell University Press, Ithaca

Doi M, Edwards SF (1986) The theory of polymer dynamics. Oxford University Press, Oxford

Eckart C (1935) Some studies concerning rotating axes and polyatomic molecules. Phys Rev 47:552–558

Freddolino PL, Harrison CB, Liu Y, Schulten K (2010) Challenges in protein folding simulations: timescale, representation, and analysis. Nat Phys 6:751

Garcia AE (1992) Large-amplitude nonlinear motions in proteins. Phys Rev Lett 68:2696–2699

Go N, Noguti T, Nishikawa T (1983) Dynamics of a small globular protein in terms of low-frequency vibrational modes. Proc Natl Acad Sci USA 80:3696

Hagita K, Takano H (2002) Relaxation mode analysis of a single polymer chain in a melt. J Phys Soc Jpn 71:673–676

Harada R, Kitao A (2011) Exploring the folding free energy landscape of a β-hairpin miniprotein, chignolin, using multiscale free energy landscape calculation method. J Phys Chem B 115:8806

Hayward S, Kitao A, Hirata F, Go N (1993) Effect of solvent on collective motions in globular protein. J Mol Biol 234:1207–1217

Hirao H, Koseki S, Takano H (1997) Molecular dynamics study of relaxation modes of a single polymer chain. J Phys Soc Jpn 66:3399–3405

Honda S, Yamasaki K, Sawada Y, Morii H (2004) 10 residue folded peptide designed by segment statistics. Structure 12:1507

Ichiye T, Karplus M (1991) Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Protein 11:205–217

Iwaoka N, Hagita K, Takano H (2015) Estimation of relaxation modulus of polymer melts by molecular dynamics simulations: application of relaxation mode analysis. J Phys Soc Jpn 84:044801. and references therein

Kamada M, Toda M, Sekijima M, Takata M, Joe J (2011) Analysis of motion features for molecular dynamics simulation of proteins. Chem Phys Lett 502:241

Karasawa N, Mitsutake A, Takano H (2017) Two-step relaxation mode analysis with multiple evolution times applied to all-atom molecular dynamics protein simulation. Phys Rev E 96:062408

Kitao A, Hirata F, Go N (1991) The effects of solvent on the conformation and the collective motions of protein: normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. Chem Phys 158:447

Kitao A, Go N (1999) Investigating protein dynamics in collective coordinate space. Curr Opin Struct Biol 9:164

Komatsuzaki T, Berry RS, Leitner DM (2011) Advancing theory for kinetics and dynamics of complex, many-dimensional systems. Wiley, Canada

Kottalam J, Case DA (1990) Langevin modes of macromolecules: applications to crambin and DNA hexamers. Biopolymers 29:1409

Koseki S, Hirao H, Takano H (1997) Monte Carlo study of relaxation modes of a single polymer chain. J Phys Soc Jpn 66:1631–1637

Kührova P, Simone AD, Otyepka M, Best RB (2012) Force-field dependence of chignolin folding and misfolding: comparison with experiment and redesign. Biophys J 102:1897

Lamm G, Szabo A (1986) Langevin modes of macromolecules. J Chem Phys 85:7334

Lane TJ, Shukla D, Beauchamp KA, Pande VS (2013) To milliseconds and beyond: challenges in the simulation of protein folding. Curr Opin 23:58

Lange OF, Grubmüller H (2007) Full correlation analysis of conformational protein dynamics. Proteins 70:1294

Levy RM, Srinivasan AR, Olson WK, McCammon JA (1984) Quasi-harmonic method for studying very low frequency modes in proteins. Biopolymers 23:1099–1112

Levitt M, Sander C, Stern PS (1985) Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. J Mol Biol 181:423

Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. Science 334:517

Lindorff-Larsen K, Margakis P, Piana S, Eastwood MP, Dror RO, Shaw DE (2012) Systematic validation of protein force fields against experimental data. PLos ONE 7:e32131

Lindner B, Yi Z, Prinz JH, Smith J, Noé F (2013) Dynamic neutron scattering from conformational dynamics I: theory and Markov models. J Chem Phys 139:175101

Matsunaga Y, Kidera A, Sugita Y (2015) Sequential data assimilation for single-molecule FRET photon-counting data. J Chem Phys 142:214115

McLachlan AD (1979) Gene duplications in the structural evolution of chymotrypsin. J Mol Biol 128:49–79

Mitsutake A, Iijima H, Takano H (2005) Principal component analysis and relaxation mode analysis of a peptide. Biophysics. 45: Supplement S214. Abstracts for the 43th Annual Meeting, The Biophysical Society of Japan. (in Japanese)

Mitsutake A, Iijima H, Takano H (2011) Relaxation mode analysis of a peptide system: comparison with principal component analysis. J Chem Phys 135:164102

Mitsutake A, Takano H (2015) Relaxation mode analysis and Markov state relaxation mode analysis for chignolin in aqueous solution near a transition temperature. J Chem Phys 143:124111

Miyashita O, Tama F (2008) Coarse-graining of condensed phase and biomolecular systems. CRC Press, Boca Raton, p 267

Moritsugu K, Koike R, Yamada K, Kato H, Kidera A (2015) Motion tree delineates hierarchical structure of protein dynamics observed in molecular dynamics simulation. PLoS ONE 10:e0131583

Mori T, Saito S (2015) Dynamic heterogeneity in the folding/unfolding transitions of FiP35. J Chem Phys 142:135101

Mori T, Saito S (2016) Molecular mechanism behind the fast folding/unfolding transitions of villin headpiece subdomain: hierarchy and heterogeneity. J Phys Chem B 120:11683

Nagai T, Mitsutake A, Takano H (2013) Principal component relaxation mode analysis of an all-atom molecular dynamics simulation of human lysozyme. J Phys Soc Jpn 82:023803

Nagai T, Mitsutake A, Takano H (2009) Relaxation mode analysis of a biopolymer system by molecular dynamics. Biophysics. 49 Supplement S75. (Abstracts for the 47th Annual Meeting, The Biophysical Society of Japan

Naritomi Y, Fuchigami S (2011) Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: the case of domain motions. J Chem Phys 134:065101

Naritomi Y, Fuchigami S (2013) Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis. J Chem Phys 139:215102

Natori S, Takano H (2017) Two-step relaxation mode analysis with multiple evolution times: application to a single [n]polycatenane. J Phys Soc Jpn 86:43003

Noé F, Horenko I, Schütte C, Smith JC (2007) Hierarchical analysis of conformational dynamics in biomolecules: transition networks of metastable states. J Chem Phys 126:155102

Noé F, Fischer S (2008) Transition networks for modeling the kinetics of conformational change in macromolecules. Curr Opin Struct Biol 18:154

Noé F, Clementi C (2017) Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods. Curr Opin Struct Biol 43:141

Okumura H (2012) Temperature and pressure denaturation of chignolin: folding and unfolding simulation by multibaric-multithermal molecular dynamics method. Proteins 80:2397

Pérez-Hernández G, Paul F, Giorgino TG, Fabritiis D, Noé F (2013) Identification of slow molecular order parameters for Markov model construction. J Chem Phys 139:015102

Prinz J, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C, Noé F (2011) Markov models of molecular kinetics: generation and validation. J Chem Phys 134:174105

Risken H (1989) The Fokker-Planck equation: methods of solution and applications 2nd Ed. Springer-Verlag. Berlin, Heidelberg

Zwanzig R (2001) Nonequilibrium statistical mechanics. Oxford university press, New York

Saka S, Takano H (2008) Relaxation of a single knotted ring polymer. J Phys Soc Jpn 77:034001

Sakuraba S, Joti Y, Kitao A (2010) Detecting coupled collective motions in protein by independent subspace analysis. J Chem Phys 133:185102

Satoh D, Shimizu K, Nakamura S, Terada T (2006) Folding free-energy landscape of a 10-residue mini-protein, chignolin. FEBS Letters 580:3422

Schütte C, Fischer A, Huisinga W, Deuflhard P (1999) A direct approach to conformational dynamics based on hybrid Monte Carlo. J Comput Phys 151:146

Schwantes CR, Pande VS (2013) Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. J Chem Theor Comput 9:2000

Schwantes CR, McGibbon RT, Pande VS (2014) Perspective: Markov models for long-timescale biomolecular dynamics. J Chem Phys 141:090901

Singhal N, Snow CD, Pande VS (2004) Using path sampling to build better Markovian state models: predicting the folding rate and

mechanism of a tryptophan zipper beta hairpin. J Chem Phys 121:415
Suenaga A, Narumi T, Futatsugi N, Yanai R, Ohno Y, Okimoto N, Taiji M (2007) Folding dynamics of 10-residue beta-hairpin peptide chignolin. Chem Asian J 2:591
Swope WC, Pitera JW, Suits F (2004) Describing protein folding kinetics by molecular dynamics simulations. 1. Theory J Phys Chem B 108:6571
Takano H, Miyashita S (1995) Relaxation modes in random spin systems. J Phys Soc Jpn 64:3688–3698
Tama F, Sanejouand YH (2001) Conformational change of proteins arising from normal mode calculations. Protein Engin 14:1
Tama F, Brooks III CL (2002) The mechanism and pathway of pH induced swelling in Cowpea chlorotic mottle virus. J Mol Biol 318:733
Tirion MM (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Lev Lett 77:1905
Wu H, Nüske F, Paul F, Klus S, Koltai P, Noé F (2017) Variational Koopman models: slow collective variables and molecular kinetics from short off-equilibrium simulations. J Chem Phys 146: 154104
Xia J, Deng JN, Levy RM (2013) NMR relaxation in proteins with fast internal motions and slow conformational exchange: model-free framework and Markov state simulations. J Phys Chem B 117:6625
Zheng Y, Lindner B, Prinz JH, Noé F, Smith J (2013) Dynamic neutron scattering from conformational dynamics II: application using molecular dynamics simulation and Markov modeling. J Chem Phys 139:175102
Zuckerman DM (2010) Statistical physics of biomolecules: an introduction. CRC Press, New York