



EspcTM: Kinetic Transition Network Based on Trajectory Mapping in Effective Energy Rescaling Space

Zhenyu Wang¹, Xin Zhou² and Guanghong Zuo^{1*}

¹ T-Life Research Center, State Key Laboratory of Surface Physics, Department of Physics, Fudan University, Shanghai, China, ² School of Physical Sciences, University of Chinese Academy of Sciences, Beijing, China

OPEN ACCESS

Edited by:

Wenfei Li,
Nanjing University, China

Reviewed by:

Naoto Hori,
University of Nottingham,
United Kingdom

Xingcheng Lin,
Massachusetts Institute
of Technology, United States

Zhang Zhi Yong,
University of Science and Technology
of China, China

*Correspondence:

Guanghong Zuo
ghzuo@fudan.edu.cn

Specialty section:

This article was submitted to
Biological Modelling and Simulation,
a section of the journal
Frontiers in Molecular Biosciences

Received: 31 July 2020

Accepted: 24 September 2020

Published: 27 October 2020

Citation:

Wang Z, Zhou X and Zuo G
(2020) EspcTM: Kinetic Transition
Network Based on Trajectory
Mapping in Effective Energy Rescaling
Space. *Front. Mol. Biosci.* 7:589718.
doi: 10.3389/fmolb.2020.589718

The transition network provides a key to reveal the thermodynamic and kinetic properties of biomolecular systems. In this paper, we introduce a new method, named effective energy rescaling space trajectory mapping (EspecTM), to detect metastable states and construct transition networks based on the simulation trajectories of the complex biomolecular system. It mapped simulation trajectories into an orthogonal function space, whose bases were rescaled by effective energy, and clustered the interrelation between these trajectories to locate metastable states. By using the EspcTM method, we identified the metastable states and elucidated interstate transition kinetics of a Brownian particle and a dodecapeptide. It was found that the scaling parameters of effective energy also provided a clue to the dominating factors in dynamics. We believe that the EspcTM method is a useful tool for the studies of dynamics of the complex system and may provide new insight into the understanding of thermodynamics and kinetics of biomolecular systems.

Keywords: effective energy, molecular dynamics, trajectory mapping, Markov models, alanine dodecapeptide, transition network

INTRODUCTION

The biomolecules are fundamentally dynamic in nature (Chodera et al., 2007). Protein folding, for example, involves the conformation change from polypeptide chain to a particular tertiary topology over microseconds to seconds, a process that can go awry and lead to misfolding and cause disease (Chiti and Dobson, 2006; Gregersen et al., 2006; Chodera et al., 2007; Guo et al., 2012; Wei et al., 2016; Zhou et al., 2019). Allosteric enzyme catalysis involves transitions between multiple conformational substates, only a few of which may allow substate access or catalysis (Eisenmesser et al., 2002; Boehr et al., 2006; Buch et al., 2011). Protein–ligand binding may alter the transition kinetics among multiple conformational states; for example, intrinsically disordered protein may have structured and unstructured binding pathways (Ithuralde et al., 2016; Paul et al., 2017; Li et al., 2019; Pan et al., 2019; Weng and Wang, 2020). Understanding of biomolecular dynamics is pivotal to reveal the function of biomolecules. Computer simulations of biomolecules, which made the biomolecular dynamics visible *in silico*, provide valuable insight for understanding how the dynamics of biomolecules drives biology processes (Cheatham and Kollman, 2000;

Mirny and Shakhnovich, 2001; Norberg and Nilsson, 2002; Moraitakis et al., 2003; Levy et al., 2004; Zhou et al., 2004; Gao et al., 2005; Zuo et al., 2006, 2009; Li et al., 2008, 2013; Miyashita et al., 2009; Yang et al., 2014; Yan and Wang, 2019; Wu et al., 2020). In particular, molecular dynamics (MD) simulations can provide atomic-level details that are not always accessible in experiments and make this technique inevitable (Karplus and McCammon, 2002; Adcock and McCammon, 2006; Wang et al., 2009; Zuo et al., 2013). However, too many details will disguise the meaningful information. In most cases, the functional processes of biomolecules, the most interesting or important processes, correspond to slow dynamical processes. To extract these processes from numerous MD simulation trajectories, much effort has been involved in the development of methods for massive high-dimensional simulation data analysis. It was now well established from a variety of studies that an intelligible picture of the dynamics of biomolecules can be described as a transition network between several metastable states based on the simulation trajectories (Zwanzig, 1983; Kampen, 2007).

Markov state model (MSM) provides a powerful framework for analyzing dynamics of biosystems, such as MD simulations, to construct a transition network of metastable states. It has gained widespread use over the past several decades (Chodera et al., 2007; Gfeller et al., 2007; Noe et al., 2007; Bowman and Pande, 2010; Pande et al., 2010; Rao and Karplus, 2010; Bowman et al., 2013; Deng et al., 2013; Weber et al., 2013; Husic and Pande, 2018; Wang et al., 2018; Sengupta et al., 2019). In the analyzing process of MSM, the simulation conformations were first classified into thousands of small groups, named as microstates, by a geometric clustering method wherein these conformations were similar in geometry (Bowman et al., 2009; Pande et al., 2010). These microstates would be further clustered into several macrostates by standard spectral clustering method based on their transition frequency (Deuffhard and Weber, 2005; Chodera et al., 2007; Gfeller et al., 2007; Noe et al., 2007; Noe, 2008; Bowman and Pande, 2010; Pande et al., 2010; Rao and Karplus, 2010; Zuo et al., 2010; Bowman et al., 2013; Deng et al., 2013; Roblitz and Weber, 2013; Weber et al., 2013; Husic and Pande, 2018; Wang et al., 2018; Sengupta et al., 2019). Then, the transition network between the macrostates was reconstructed accordingly (Jayachandran et al., 2006; Buchete and Hummer, 2008; Prinz et al., 2011). Gong and Zhou (2010) presented the trajectory mapping (TM) method to construct a kinetic transition network of metastable states. Compared with MSM, TM grouped simulation trajectory pieces rather than individual conformations. They mapped the averaged conformation of each MD trajectory segment as a vector and calculate the principal components (PCs) of the trajectory-mapped vectors by the principal component analysis (PCA). The similar trajectory-mapped vectors were then grouped as metastable states by spectral clustering method, and transition events in simulation trajectories were further identified (Gong et al., 2015; Zhang et al., 2017; Zhang et al., 2019a; Zhang et al., 2019b).

In both MSM and TM methods, the discretization of MD trajectories, i.e., clustering of structures, plays a vital role in the analysis of MD trajectories. To make clustering of structures as accurate as possible, a variety of structural

metrics and their functions were employed in analysis, for example, the torsion angles of backbone, the proportion of native contacts, root mean square deviation, and solvated energy (Gong et al., 2015). These analyses can be effective when all input coordinates are sufficient and irrelevant to each other. Thus, PCA was used to find orthogonal collective coordinates, which are linear combinations of the input coordinates and covered most of variances with only the first several eigenvectors (Lever et al., 2017). However, as mentioned above, the slow dynamical process is the concerned part in most cases. It is not always true that the high variance directions correspond to the kinetically slow-motion mode. Thus, some methods have been developed to obtain slow-motion directions. In the MSM, time-structure based Independent Correlation Analysis (tICA) was used (Naritomi and Fuchigami, 2011, 2013; Perez-Hernandez et al., 2013; Schwantes and Pande, 2013). It finds the slow collective coordinates by eigen-decomposition of a Δt -interval autocorrelation matrix. In the TM, the averaged conformation of every τ -length MD trajectory segment was mapped as a vector in feature space to compose samples for the PCA method. It was argued that fast conformational fluctuations were suppressed after the segment averaging, and the PCs mainly involve slow motions (Zhang et al., 2017). In both tICA-MSM and TM methods, a hyper-parameter, Δt for tICA-MSM and τ for TM, is required. It is difficult for inexperienced users. It is possible to obtain the optimized model by an automated process instead of a process of trial and error. For example, one might consider weighting the input coordination by an order parameter relevant to the functional processes of biomolecules, so that the input coordinates with high correlation contribute the most to the distance calculation and make the clustering effective and efficient to catch the functional processes, i.e., slow-motion patterns of the biomolecular system.

In this paper, we will present a new method, named effective energy rescaling space trajectory mapping (EspcTM), for detecting metastable states and constructing transition networks. It is a parameter-free analysis framework based on the previous TM method. In the EspcTM method, every snapshot of the trajectories was described by a high-dimensional vector and mapped into an orthogonal functional space. Different from the TM method, the features were rescaled by the effective energy of the dynamics to make the space effective to describe the slow processes of the system, and no hyperparameter was required. Here, the effective energy, which was filtered from the total potential energy of simulation trajectories by fast Fourier transform (FFT) and multiple linear regression, is an efficacious order parameter to describe the slow conformational change of complex system. The PCA method was also employed for dimensionality reduction and orthogonalization of the functional space. The metastable states were assigned by a spectral clustering method based on projections of the trajectories in this feature space. Then, the Markov transition matrix is constructed based on the transitions between these metastable states. We show application of this method by the movement of a Brownian particle and conformational dynamics of an alanine dodecapeptide (Ala₁₂). It revealed their metastable states and

kinetic transition network, as well as provided additional insight into the dynamics of these two systems.

THEORY AND METHOD

The EspcTM method is an analysis framework to identify metastable states from simulation data in the effective energy rescaling space and construct the transition network between the states based on the theory of Markov chain. In the EspcTM, an ordered parameter, named effective energy, was introduced to rescale feature space of the system. The simulation trajectories were mapped into the space and discretized to obtain the kinetic transition network of the system based on Markov chain theory. **Figure 1** shows the flow chart of the EspcTM method, and details of the key steps are followed.

Feature Extraction

In our study, there were N_t frames in every trajectory. They were mapped into a space consisting of N_b basis functions

$\{\hat{A}^\mu(\vec{q})\}_{\mu=1,\dots,N_b}$. To eliminate the effect of various units of basis function, normalization was performed on every dimension. Then, every trajectory was described as an $N_t \times N_b$ -dimension matrix in the feature space, i.e., feature matrix

$$V = \left(\hat{A}^1(\vec{q}), \hat{A}^2(\vec{q}), \hat{A}^3(\vec{q}), \dots, \hat{A}^{N_b}(\vec{q}) \right) \quad (1)$$

where \vec{q} denotes the structural metrics, such as the torsion angle of backbone in peptide. Here, the basis functions $\{\hat{A}^\mu(\vec{q})\}_{\mu=1,\dots,N_b}$ should be chosen to identify typical conformational motions of systems. In this work, we used the sine and cosine of structural metrics as the feature space (Gong and Zhou, 2010; Gong et al., 2015).

Noise Reduction

It is obvious that every basis possesses different weight on describing the dynamics of complex system. It was argued that dynamics of complex systems, such as protein folding, can resemble a diffusive process on a rugged landscape of free energy (Onuchic et al., 1997). Thus, energy is an appropriate measure to rescale their coordinates. Most studies of complex system focus on the dynamics of a part of the system, and the rest of the system was regarded as the environment of the study object. For example, studies on protein folding focus on protein molecules. The conformational change of protein in protein folding is the interesting part, instead of the fluctuation of water molecules. However, the atoms of the system interacted with each other in a complicated way. The energy variation caused by the dynamics of the studied object is coupled with the energy caused by the fluctuation of the remaining part. It is difficult to isolate the meaningful energy in a frame without additional hypotheses. On the other hand, as mentioned above, the kinetic slowness is the main character of the interesting processes. Therefore, the dynamics of the important processes can be separated from the fluctuation in the frequency domain, where slow motion is treated as low-frequency signal and fluctuation can be filtered out as high-frequency noise.

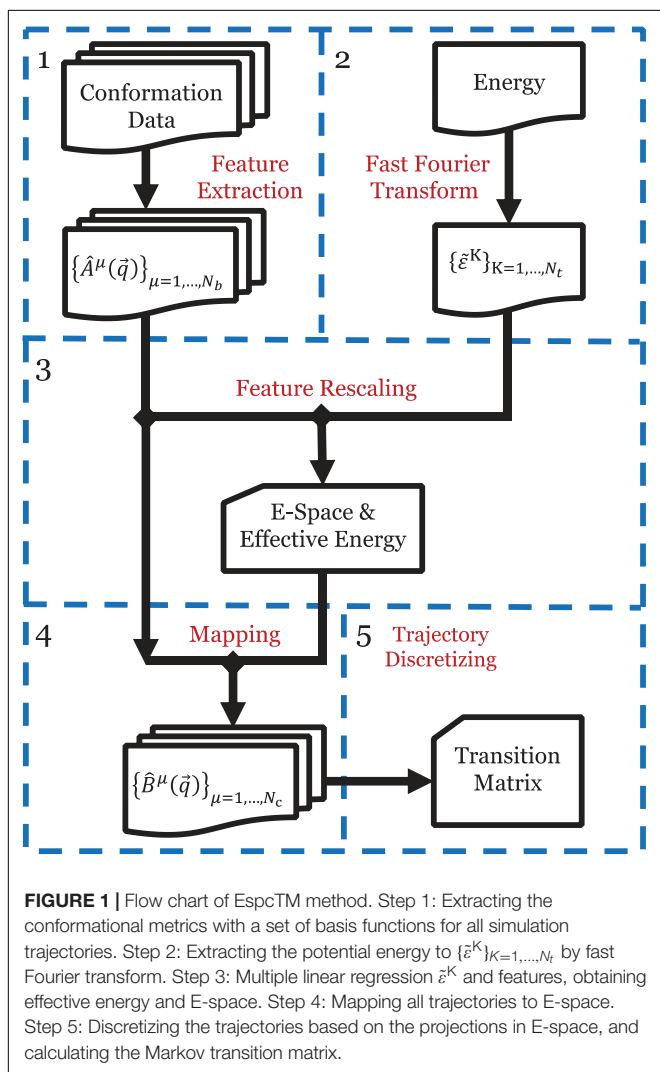
In this work, FFT (Cochran et al., 1967) was applied to transform the energy of trajectories into frequency space. For every trajectory, the coefficients of frequencies were obtained by

$$\tilde{\omega}_k = \sum_{n=0}^{N_t-1} \varepsilon_n \cdot e^{-in\omega_k} \quad (2)$$

Here, $i = \sqrt{-1}$ is the imaginary mark, n is the index of frames for the trajectory, ε_n is the total potential energy of the n th frame obtained from the simulation data, N_t is the number of frames of a trajectory, and $\omega_k = 2\pi k/N_t$ corresponds to a frequency. To reduce the false edge, even extension was used before FFT for every trajectory. Then, a reverse FFT was performed on the first K frequencies for every trajectory to obtain the $\tilde{\varepsilon}^K$ of every frame:

$$\tilde{\varepsilon}_n^K = \sum_{k=0}^{K-1} \tilde{\omega}_k \cdot e^{in\omega_k} \quad (3)$$

The fluctuation whose $\omega \geq \omega_K$ was excluded in $\tilde{\varepsilon}^K$. To determine the number K , we performed multiple linear regression



(Schneider et al., 2010) between K-energy vector $\tilde{\epsilon}^K$ and feature matrix V for all trajectories:

$$\tilde{\epsilon}^K = a_0^K + V \cdot \hat{a}^K + \epsilon^K \quad (4)$$

Here, a_0^K (scalar) and \hat{a}^K (N_b -dimensional vector) are the fitting parameters, and ϵ^K is the error for the multiple linear regression. The effective energy $\tilde{\epsilon} = \tilde{\epsilon}^{K^*} - \epsilon^{K^*}$ with the $K^* = \arg \max r(K)$. Here, $r(K) = \sqrt{1 - (\epsilon^K)^2 / (\sigma^K)^2}$ is the multiple correlation coefficient, $(\sigma^K)^2$ is the variance of $\tilde{\epsilon}^K$, and $r = 0$ for the case $(\sigma^K)^2 = (\epsilon^K)^2 = 0$. For multiple trajectories, the FFT was performed on every trajectory separately. Due to same length and time interval of all trajectories in our study, all trajectories were mapped into the same frequency space $\{\omega_k\}_{k=1, \dots, N_t}$. Thus, in the revised FFT, the K-energies of all trajectories are the summary of the same frequencies for every K. Before multiple linear regression, K-energy vectors $\tilde{\epsilon}^K$ and feature matrixes V of all trajectories were joined into a vector and a feature matrix for equation (4).

Feature Rescaling and Mapping

The regression coefficients \hat{a}^K were used as the weight factors on features. Every trajectory was described as a new $N_t \times N_b$ -dimension matrix:

$$\tilde{V} = V \cdot \text{diag}(\hat{a}^K) \quad (5)$$

Here, $\text{diag}(\hat{a}^K)$ is an $N_b \times N_b$ diagonal matrix with the elements of \hat{a}^K on its main diagonal. A PCA (Sims et al., 2005) was applied to reduce the dimension and orthogonalize the components of all trajectories \tilde{V} . Descending according to eigenvalues, the first N_c eigenvectors were selected to consist of an $N_b \times N_c$ matrix M . Here, $N_c \ll N_b$, and M is the mapping operator, which reduced the N_b - dimension vectors into N_c - dimension, given top N_c eigenvalues whose sum has over 90% fraction of the sum of all eigenvalues. Here, we named this N_c - dimension space as E-space since its input coordinates were weighted by the regression coefficients. By using the mapping operator M , we mapped all original feature matrixes V_j into the E-space. Therefore, every frame of the trajectories was described as an N_c - dimension vector $\{\hat{B}^\mu(\vec{q})\}_{\mu=1, \dots, N_c}$.

Trajectory Discretizing

The clustering of conformations was performed in the E-space, i.e., based on the analysis of the projection vectors $\{\hat{B}^\mu(\vec{q})\}_{\mu=1, \dots, N_c}$. Similar to the TM method (Gong and Zhou, 2010; Zhang et al., 2017), every trajectory was divided into a lot of isometric pieces, and the similarity between each two pieces was defined by their average vectors:

$$S(t, t') = \frac{\sum_i [B_i(t)B_i(t') + 1]}{\sqrt{\sum_i [B_i(t)B_i(t) + 1]} \times \sqrt{\sum_i [B_i(t')B_i(t') + 1]}} \quad (6)$$

Here, we replaced the vectors of frames by the average vectors of trajectory pieces. It reduced the size of the similarity matrix and cost of computation resource. In practice, the length of the trajectory pieces can be varied in a reasonable range. The Robust

Perron Cluster Analysis (PCCA+) method (Roblitz and Weber, 2013), implemented in pyEMMA (Scherer et al., 2015), was used to classify all pieces into N_s states based on the similarity matrix. Here, the number of states N_s was determined by the distribution of the eigenvalues of the similarity matrix (Roblitz and Weber, 2013). The Markov transition matrix P was obtained based on the discretized trajectories (Prinz et al., 2011). Since P is a row stochastic matrix, its largest left eigenvalue is 1. If there is a unique stationary distribution, it is true for our case, then the largest eigenvalue and the corresponding eigenvector is unique too. As the theory of stochastic process, the stationary distribution of the Markov process corresponds to the distribution of equilibrium state. More interestingly, the Markov transition matrix can also be used to reveal the dynamics of the system in non-equilibration conditions (Reuter et al., 2018).

Brownian Dynamic Simulation

For Brownian dynamic simulation, Brownian particles in the presence of a potential, U , are described by the Langevin equation

$$m \frac{dv(t)}{dt} = -\nabla U(x) - \gamma v(t) + \zeta(t) \quad (7)$$

where $\zeta(t)$ is a delta-correlated stationary Gaussian process with zero-mean. A two-dimensional Brownian particle was simulated on the surface with three potential wells in the toy model (see **Figure 2A**). Here, the potential $U(x)$ was defined as:

$$U(x) = -\varepsilon \left\{ \cos(x) + \sin(x) + \frac{1}{2} \cos(y) + 2 \cos(3x) \right. \\ \left. + 2 \exp \left[-20 \left(x + \frac{2}{3}\pi \right)^2 - 2y^2 \right] \right\} \quad (8)$$

with scaling parameter $\varepsilon = 40$. Multiple trajectories were generated from different initial sites randomly with extensive long simulations.

MD Simulation

In the MD simulation, the termini of Ala₁₂ were charged, which leads to versatile metastable structures (Noe et al., 2007). All atoms were modeled by using Amber03 force field. The molecule was solvated in a rhombic dodecahedral periodic box with the distance between the solutes and box boundary at least 10 Å. The SPC water model was used for solvation (see **Figure 3A**). The MD simulations were performed using the Gromacs package 4.6.5 (Hess et al., 2008). In the simulations, the covalent bonds involving H atoms were constrained by the LINCS algorithm, which allowed a time step of 2 fs. The long-range electrostatic interactions were treated with the particle-mesh Ewald method (Darden et al., 1993) with a grid spacing of 1.6 Å. The cutoff for the van der Waals interaction was set to 10 Å. The previous trajectory performed at high temperature was equilibrated by MD simulations for 100 ps at a constant pressure of 1 bar and a temperature of 500 K using Berendsen coupling (Berendsen et al., 1984). Then, the production simulations were performed in NVT ensemble at 500 K for 100 ns. All 50 systems extracted from high-temperature simulation had been iterated 100 ns in NVT

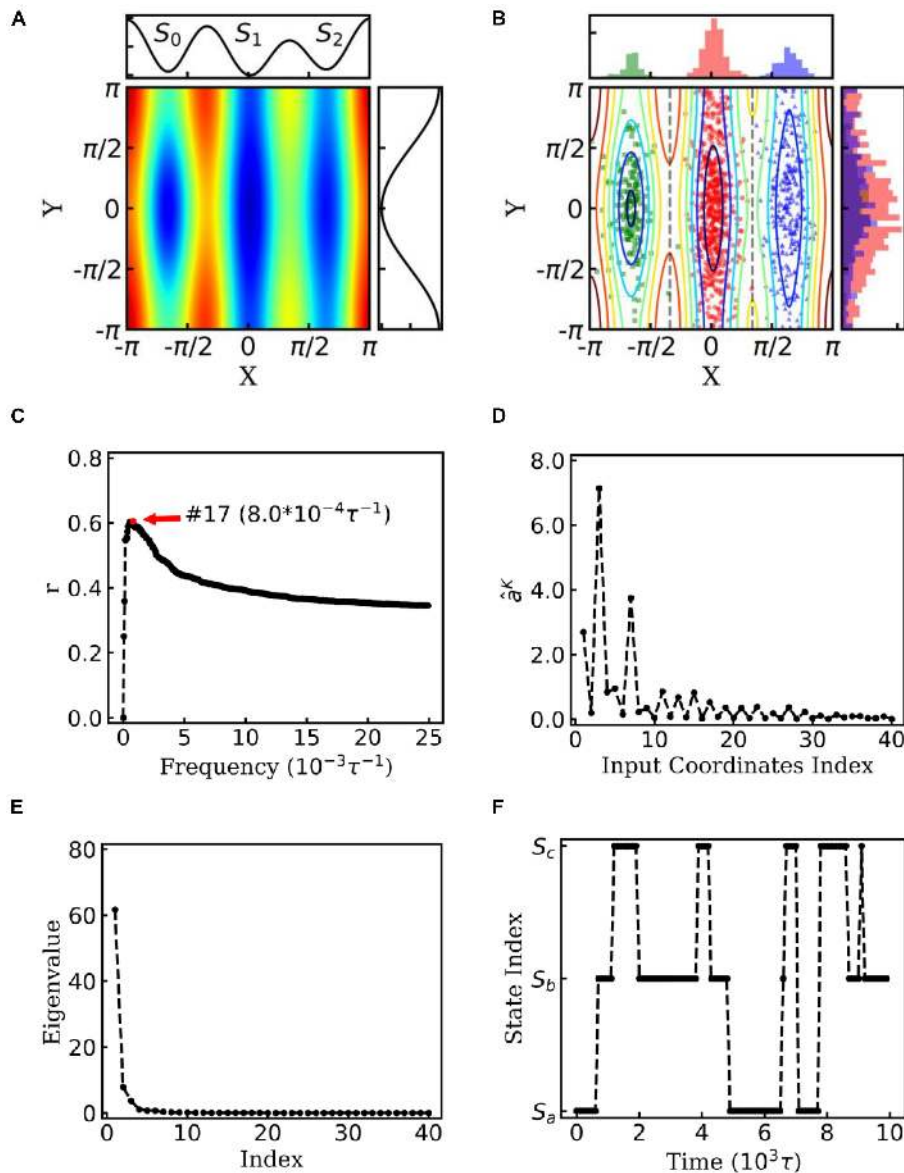


FIGURE 2 | EspcTM on dynamic of Brownian particle. **(A)** The energetic landscape of the toy model. Here, the potential function of the landscape was $-\varepsilon\{\cos(x) + \sin(x) + 2\cos(3x) + \frac{1}{2}\cos(y) + 2\exp[-20(x + \frac{2}{3}\pi)^2 - 2y^2]\}$. Three potential wells from left to right were S_0 , S_1 , and S_2 . The well of S_1 was deeper than that of the other two states, and the barrier between S_0 and S_1 was much higher than that between S_1 and S_2 . The black line on the top and right panel represents the potential along line $y = 0$ and $x = 0$, respectively. **(B)** Red, green, and blue dots represent three states of the snapshots of trajectories. The histograms of each state were shown on the top and right panel in different colors. **(C)** Multiple correlation coefficients of ε^K and all 40 conformational coordinates as a function of cutoff frequencies. Here, the maximum of the multiple correlation coefficient located at cutoff frequency equaling $8.0 \times 10^{-4}\tau^{-1}$. **(D)** The regression coefficients for all 40 features. The coordinates corresponding to basis functions $\sin(x)$, $\cos(x)$, and $\cos(2x)$ possessed large weights in the rescaling. **(E)** The eigenvalues in the PCA of trajectory-mapped vector. **(F)** A typical discretized trajectory.

ensemble at 300 K and recorded with time interval $\tau = 5$ ps. There are $20,000 \times 50$ frames in the analysis.

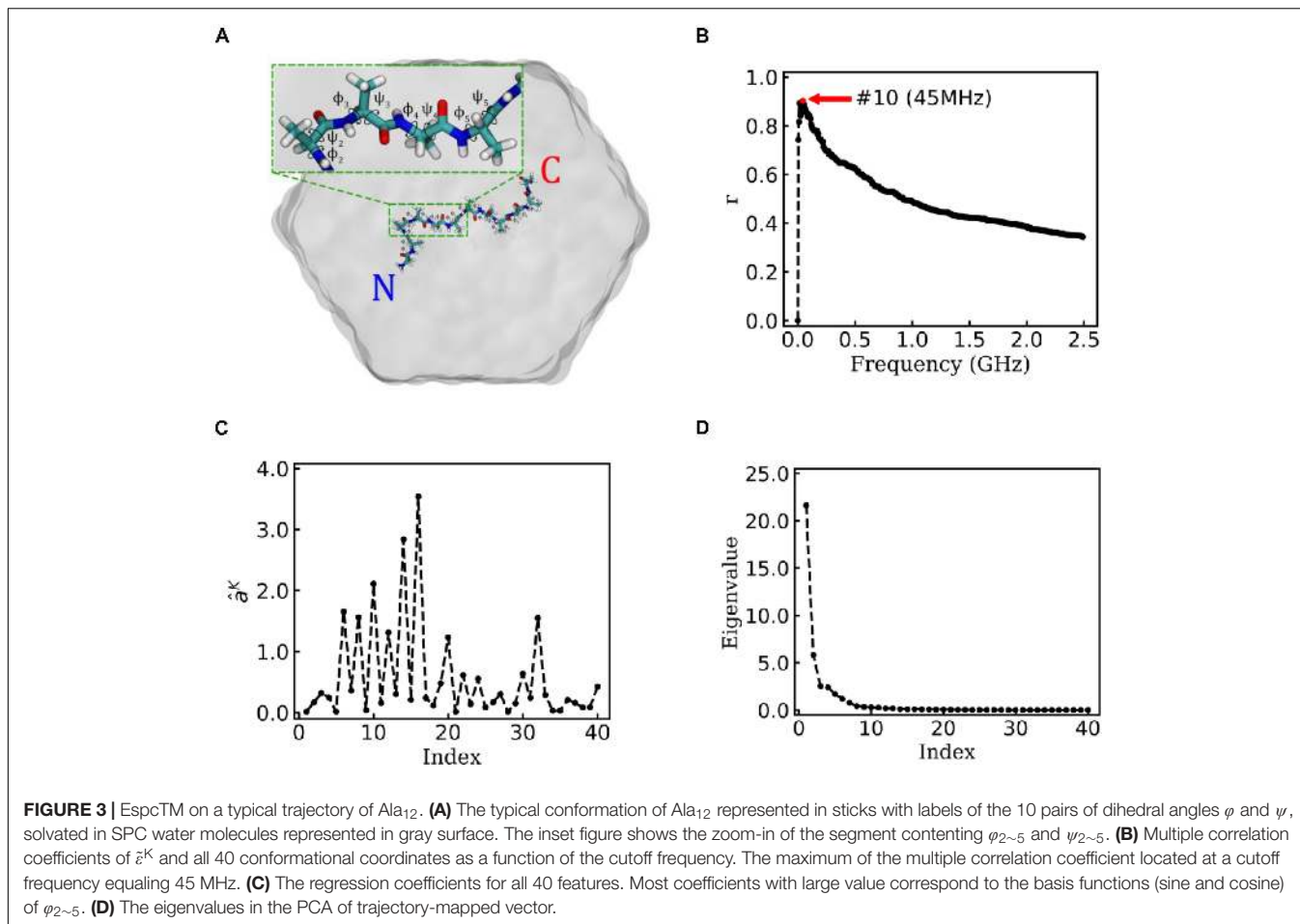
RESULTS AND DISCUSSION

The EspcTM method was first illustrated with a toy model, i.e., the dynamics of a Brownian particle on a two-dimensional

surface. Then, it was applied to investigate the conformational dynamics of alanine dodecapeptide (Ala_{12}), and a transition network between metastable states of Ala_{12} was constructed.

Toy Model

In the toy model, a two-dimensional Brownian particle was moving in the field with three potential wells (see Figure 2A). Ten extensive long simulations, which started from different sites



randomly, were performed to make the distribution of samples close to the theoretical values. **Figure 2B** shows the positions and distribution of the samples of these trajectories. In the analysis, $\sin(n\theta)$ and $\cos(n\theta)$ were selected as the basis functions. θ indicates the coordinate x or y , and $n = 1, \dots, 10$ for every coordinate in the EspcTM analysis of the toy model. Hence, the trajectories were mapped into a 40-dimensional functional space, e.g.,

$$\begin{aligned} & \sin(x), \sin(y), \cos(x), \cos(y), \dots \\ & \sin(10x), \sin(10y), \cos(10x), \cos(10y) \end{aligned} \quad (9)$$

All values of the trajectories were normalized in every dimension before they were fitted with $\tilde{\varepsilon}^K$.

Figure 2C shows the multiple correlation coefficient between $\tilde{\varepsilon}^K$ and the values of these 40 features as a function of the cutoff frequency. There was a maximum multiple correlation coefficient at $K = 17$, and $\tilde{\varepsilon} = \tilde{\varepsilon}^{17} - \varepsilon^{17}$ was selected as the effective energy. **Figure 2D** shows regression coefficients between the energy $\tilde{\varepsilon}^{17}$ and features. As shown in **Figure 2D**, the basis functions $\sin(x)$, $\cos(x)$, and $\cos(2x)$ possessed large weight in the rescaling. It should be noted that to consider the effect of the random force by solvation in Brown dynamics, additional energies with Gaussian distribution were added into the energies

of the Brownian particle, so that information of potential was mixed with white noise in linear regression. PCA was performed on these effective energy rescaled samples. **Figure 2E** shows the eigenvalues in descending order. It is obvious that apart from the first two eigenvalues, other eigenvalues were very small. The first two eigenvectors were selected to compose the E-space of the toy model, as well as the mapping operator. By using the mapping operator M , composed by these two eigenvectors, all samples were mapped into the E-space.

By using the PCCA+ algorithm, all samples had been grouped into three states (shown by colored dots in **Figure 2B**). As shown in **Figure 2B**, these three states corresponded to the three wells in the potential. A discretized trajectory who visited all three states is shown in **Figure 2F**. The Markov transition matrix P was obtained based on the discretized trajectories (see **Table 1**). The stationary distribution, which corresponds to the distribution of the thermodynamic equilibrium, was obtained by the eigen-decomposition of the Markov transition matrix and shown in **Table 1**. As a benchmark, the distribution of equilibrium state predicted by the theory of statistical physics is shown in **Table 1** as well. It is obvious that the result obtained by the EspcTM method is similar to the theoretical values. Furthermore, the Markov transition matrix contains kinetic information about the system as well. The lifetime of these states,

TABLE 1 | Transition matrix and stationary distribution of the Markov model, distribution obtained by theory of equilibrium statistical physics, and lifetime of states for the dynamics of a Brownian particle.

	Transition matrix P			Stationary distribution	Theory	Lifetime (100 τ)
	S0	S1	S2			
S0	0.882	0.069	0.049	0.184	0.186	8.5
S1	0.024	0.858	0.118	0.532	0.538	7.08
S2	0.032	0.221	0.747	0.284	0.276	3.99

which were calculated by the diagonal elements of the transition matrix, is also shown in **Table 1**. It was found that the state S_0 possessed the lowest occurring probability but the longest lifetime. This indicated that the kinetically stable state was not the thermodynamically stable state for this dynamic system.

Dynamics of Alanine Dodecapeptide

Alanine dodecapeptide (Ala₁₂), consisting of 12 alanine residues, is a typical model molecule for MD study (Noe et al., 2007). The MD trajectories of an Ala₁₂ was used as an example to test the EspcTM method. According to the previous study (Gong and Zhou, 2010; Gong et al., 2015), sine and cosine of backbone dihedral angles (φ , ψ) were used as basis functions in the analysis of the MD trajectories of Ala₁₂. Here, φ is defined as the backbone dihedral angle around the bond connecting C_α and N atoms and ψ is defined as the backbone dihedral angle around the bond connecting C_α and carbonyl carbon atoms (Hovmoller et al., 2002). There are 10 pairs of dihedral angles φ and ψ for Ala₁₂ (see **Figure 3A**), and 40 basis functions were finally included in the analysis, e.g.,

$$\sin(\psi_i), \sin(\varphi_i), \cos(\psi_i), \cos(\varphi_i) \quad (10)$$

Here, $i = 1, \dots, 10$ indicates the index of dihedrals of Ala₁₂ from N-terminal to C-terminal. Based on these basis functions, the EspcTM was first applied on a typical trajectory and then on all the 50 trajectories.

State Transition of a Typical Trajectory

Figure 3B shows the result of the multiple linear regression between $\tilde{\varepsilon}^K$ and functions of the dihedral angles of Ala₁₂ for a typical trajectory. There is a maximum of the multiple correlation coefficient, similar to the case of movement of Brownian particle, at 45 MHz (see **Figure 3B**). Therefore, the summary of the first 10 lowest frequencies of energy $\tilde{\varepsilon}^{10}$ was used in the analysis. The regression coefficients between the energy $\tilde{\varepsilon}^{10}$ and functions of dihedral angles are shown in **Figure 3C**. It was found that most factors with large weight corresponded to the basis function (sine and cosine) of $\varphi_{2\sim5}$ (see the inset figure of **Figure 3A**). This indicates that the structure change near N-terminal contributes more to large-scale conformational change than C-terminal in this typical simulation trajectory.

Figure 3D shows the eigenvalues of weighted samples of this trajectory. As shown in **Figure 3D**, the following analysis on this trajectory was performed in the space made up of the first six eigenvectors. **Figure 4A** shows the similarity matrix and the

representative structure of the trajectory. It was obvious that there were four metastable states in the trajectory. The discretized trajectory is shown in the middle panel of **Figure 4B**. The secondary structure of the peptide was analyzed by DSSP (Kabsch and Sander, 1983; Touw et al., 2015) and shown in the top panel of **Figure 4B**. The simulation started from a structure with some of the N-terminal α -helix formed (also see the representative structure), i.e., the state S_b . This state was unstable and only existed about 6.4 ns in the 100-ns trajectory. The α -helix formed in this state acted as a nucleus that promoted the formation of the α -helix of the C-terminal of the Ala₁₂. Then, the trajectory transitioned to the S_a state, in which most of the residues of the peptide formed the α -helix structure. State S_a was more stable than state S_b . It appeared two times in this trajectory and existed about 58.0 ns in total. However, between the two occasions of the state S_a , the α -helix of two termini had been temporally uncoiled and interacted with the α -helix in the middle of the peptide, i.e., the state S_c . This state is unstable and existed only for 16.4 ns in this trajectory. After the state S_c , the peptide folded to the state S_a again. Finally, the peptide unfolded into a random coil, i.e., state S_d , with low structural similarity.

The bottom panel of **Figure 4B** shows effective energy as a function of time for this trajectory. It was calculated from the total energy of the whole biosystem, including the peptide and water molecules. Initially, the energy caused by the conformational change of the peptide was concealed by the noise of the dynamics of water molecules as well as the fluctuation of itself. It seemed that the total energy (shown in gray) varied randomly and dramatically. However, by using the FFT and regression, we obtained the effective energy (shown in red). It was synchronous with conformational change and state transition of the peptide. More interestingly, the effective energy of stable state, state S_a , was much lower than the other three states, in which most of the α -helix was formed. This implied that the stability of this state was supported by energy. On the other hand, the state S_d possessed the highest energy and large conformational variations. This implied that the unfolded coil structure was stabilized by the entropy.

Transition Network of Ala₁₂

To obtain statistically significant conclusions, we performed the analysis of EspcTM method on 50 MD trajectories. **Figure 5A** shows the result of the multiple linear regression between $\tilde{\varepsilon}^K$ and functions of the dihedral angles of Ala₁₂ for these 50 trajectories. The maximum of the multiple correlation coefficient was found at the frequency equal to 15 MHz. The summation of the first four lowest frequencies of energy $\tilde{\varepsilon}^4$ was used in the analysis. **Figure 5B** shows the regression coefficients between the energy $\tilde{\varepsilon}^4$ and features. It consistently showed that $\varphi_{2\sim5}$ played important roles in the dynamics of the Ala₁₂ though there was a phase shift on $\varphi_{2\sim5}$ caused small weights on the cosine of $\varphi_{2\sim5}$. This indicates that local structure changes near the N-terminal, especially the $\varphi_{2\sim5}$, were the major contributors to the slow conformational change of the Ala₁₂. According to the result of the PCA on the weighted feature space, the clustering algorithm was performed in the space made up of the first 10 eigenvectors, whose sum was over 90% sum of variation (see **Figure 5C**). Every

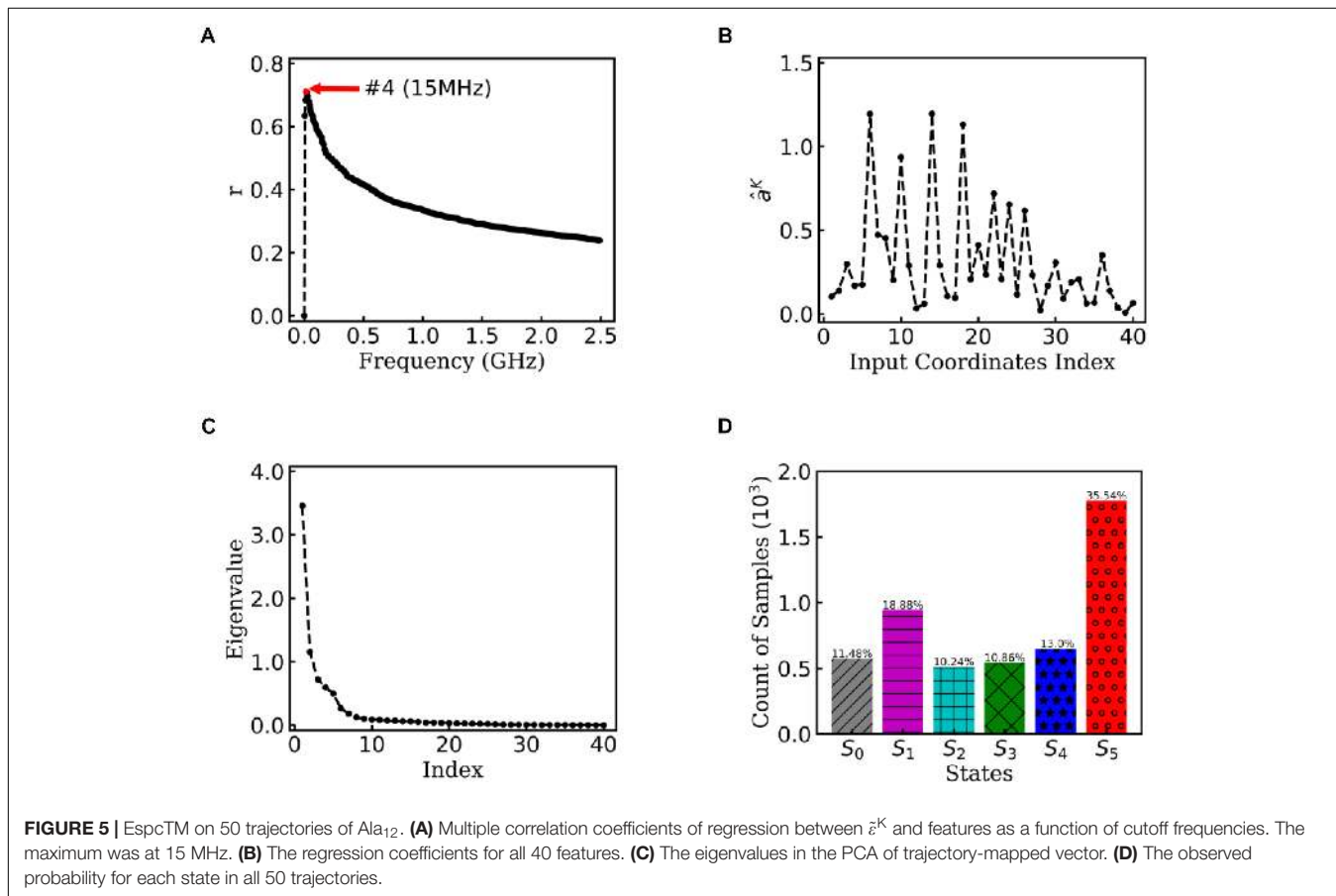
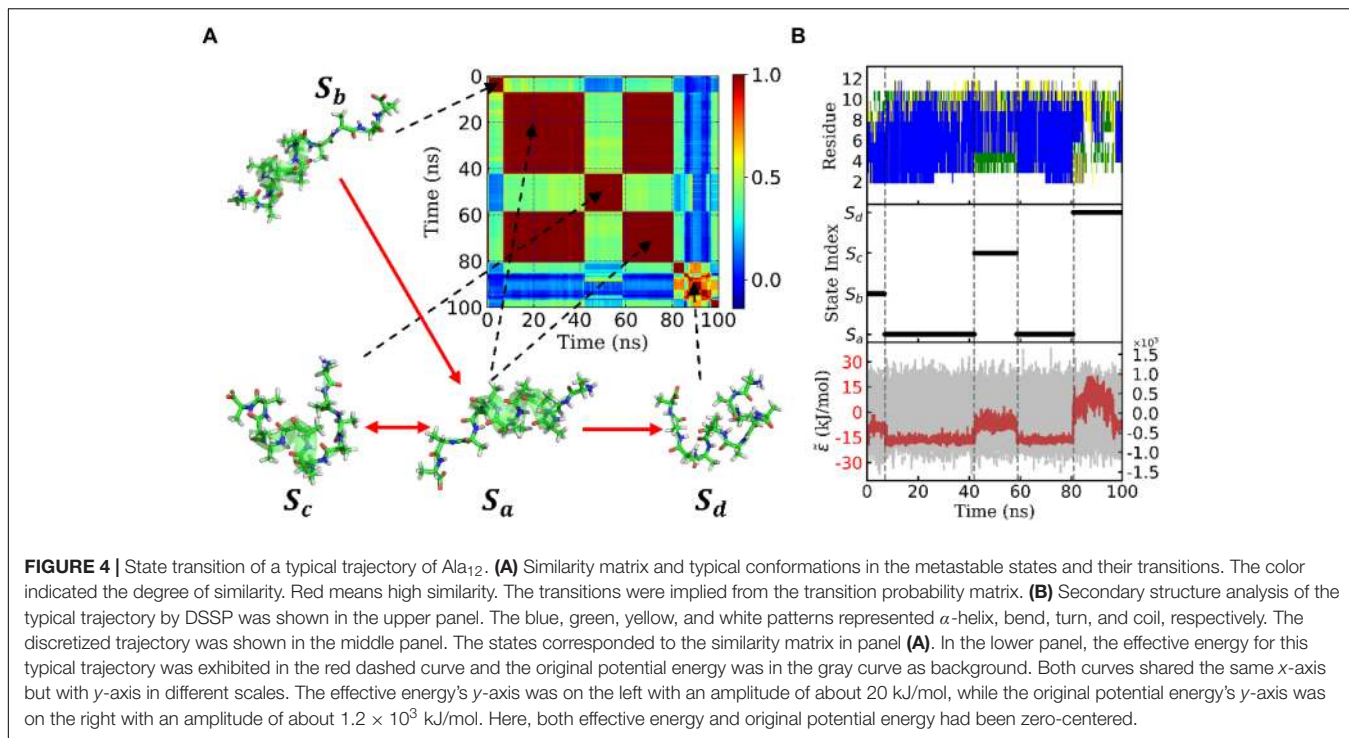


TABLE 2 | Transition matrix and stationary distribution of the Markov model and lifetime of states for the dynamics of Ala₁₂.

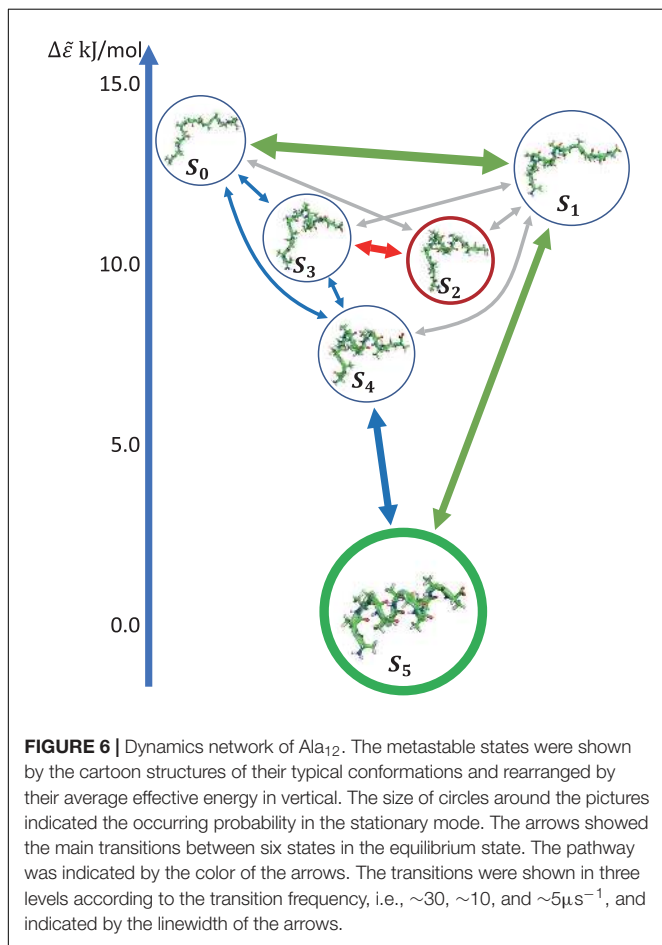
	Transition matrix						Stationary distribution	Lifetime (ns)
	S0	S1	S2	S3	S4	S5		
S0	0.610	0.276	0.038	0.031	0.036	0.009	0.114	2.61
S1	0.170	0.693	0.021	0.037	0.026	0.053	0.187	3.30
S2	0.042	0.037	0.758	0.115	0.030	0.018	0.103	4.16
S3	0.032	0.064	0.109	0.725	0.050	0.020	0.109	3.67
S4	0.031	0.038	0.024	0.042	0.798	0.067	0.131	4.97
S5	0.003	0.027	0.005	0.006	0.025	0.934	0.356	15.15

trajectory was divided into 100 pieces. Thus, there were 5,000 vectors, which represent 100×50 trajectory pieces. Six states were identified from these 50 trajectories.

Figure 5D shows the histogram of these six states. Here, the state transitions were obtained from the 50 trajectories with the lag time 1.0 ns. The transition matrix and stationary distribution are shown in **Table 2**. It was found that the stationary distribution obtained by the transition matrix was consistent with the histogram. The state S₅ had a much higher occurring probability than that of other states in the equilibrium state. **Figure 6** displays these six states, represented by their typical structures in cartoons, along with their average effective energy

in vertical. The unfolded states S₀, in which peptide unfolded into a random coil, possessed the highest energy and located at the top of the figure. The folded state S₅, in which the peptide folded into α -helices, possessed the lowest effective energy and located at the bottom of the figure. Between these two states, the peptide was half-folded. In the state S₁, a helix was formed in the N-terminal of the peptide. In states S₂, S₃, and S₄, some helices were formed in the C-terminal. A remarkable gap between the effective energy of state S₄ and state S₅ separated the folded state from the other five states. This implied that the energy is the reason for the stability of the folded state.

Furthermore, we obtain the dynamics and kinetics of the system based on the transition matrix. **Figure 6** shows the main transition between six states in lines with arrows. The most frequent transition, about $32 \mu\text{s}^{-1}$, occurred between the state S₀ and S₁ due to the high flexibility of the peptide in these two states. This high transition frequency made the lifetime of these two states lower than that of states S₂, S₃, and S₄, though the occurring probabilities of these two states were a little higher than the other three states. In the transition network, there were two main folding pathways from the unfolded state to the folded state. The fast folding pathway, which passed through state S₁ and was shown by green arrows, formed the α -helices from the N-terminal to the C-terminal directly. The slow folding pathway, which involved states S₂, S₃, and S₄, was shown by blue and red arrows and was more complex than the fast one. In this pathway, the α -helices formed from the C-terminal to the N-terminal, i.e., passed through states S₃ and S₄ sequentially. The misfolded state S₂ connected with state S₃. A detailed structural study showed that the structures of states S₂ and S₄ were very similar. However, some misfolded residues hindered the formation of the N-terminal helix in the state S₂. To reach the folded state, it must unfold into state S₃. These results indicated that the N-terminal helix plays a vital role in the folding of the peptide in kinetics. It is consistent with the aforementioned result of linear regression, that the $\varphi_{2\sim5}$ of the peptide possessed large rescaling factors, as well as the results by other experimental groups, that alanine-rich peptides folded into the α -helix in the N-terminal at first (Millhauser et al., 1997; Yoder et al., 1997). It must be noted that, as we mentioned before, the biomolecules are intrinsically dynamic (Chodera et al., 2007) and the unfolded states of the peptide were transferred to each other frequently. These two pathways only described the major folding process of Ala₁₂. Some minor branches in the folding pathways also existed.



CONCLUSION

In this work, we introduced our EspcTM method by applying it to investigate the movement of Brownian particle and conformational dynamics of Ala₁₂ in this work. In the study of Brownian particle, by using the EspcTM method, we obtained three states from simulation trajectories. The regions of the states given by EspcTM are in accordance with the potential wells of the landscape. In addition, the equilibrium distribution obtained by the kinetic transition network-based Markov chain theory was consistent with the theoretical result. In the study of Ala₁₂, a meaningful kinetic transition network was obtained to describe the folding behavior of Ala₁₂. The effective energy, which was filtered from the total potential energy of simulation trajectories by FFT and multiple linear regression, was shown to be an efficacious order parameter to describe the conformational change of Ala₁₂. We showed that the folding process of Ala₁₂ was synchronous with the change of effective energy. The folded state, in which most of the residues were in helices, possessed the lowest effective energy and was most stable in thermodynamics. Two major folding pathways were also found in the kinetic network. The N-terminal helix of the Ala₁₂ was found to play an important role in the folding of Ala₁₂ in both thermodynamics and kinetics. This is consistent with previous experimental result. Thus, the EspcTM is expected to be a powerful tool for studies of dynamics of complex systems and should be applied to studies of dynamics of large biomolecule systems to improve our understanding of the thermodynamics and kinetics of biomolecular systems.

Technically, the EspcTM method is an analysis framework based on the TM method. It identifies metastable states from simulation data and constructs the transition network between the states based on the theory of Markov chain. Different from the TM method, we provided a *de novo* solution to obtain an analysis space, named as E-space, to describe the slow processes in the EspcTM method. This solution is based on a parameter-free optimization approach. Thus, the EspcTM method is friendly to inexperienced users. The E-space is independent from the

TM method. It is convenient to use it in the MSM method. For the experienced users, especially those with knowledge on the dynamics of system, they can set cutoff frequency manually as well. Furthermore, as an extension of the EspcTM method, some new transfer functions, such as logistic function and ReLU, can also be used in the energy filter process. The wavelet analysis method can be used in transforming the energy between time domain and frequency domain.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

GZ and ZW designed the study. ZW collected data and carried out the calculation. ZW, XZ, and GZ wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 11474068) and the National Basic Research Program of China (973 Project; Grant No. 2013CB834100).

ACKNOWLEDGMENTS

The authors thank the support of the State Key Laboratory of Applied Surface Physics and the Department of Physics, Fudan University, China.

REFERENCES

- Adcock, S. A., and McCammon, J. A. (2006). Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* 106, 1589–1615. doi: 10.1021/cr040426m
- Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984). Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi: 10.1063/1.448118
- Boehr, D. D., Mcelheny, D., Dyson, H. J., and Wright, P. E. (2006). The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* 313, 1638–1642. doi: 10.1126/science.1130258
- Bowman, G. R., Beauchamp, K. A., Boxer, G., and Pande, V. S. (2009). Progress and challenges in the automated construction of markov state models for full protein systems. *J. Chem. Phys.* 131:124101. doi: 10.1063/1.3216567
- Bowman, G. R., Meng, L. M., and Huang, X. H. (2013). Quantitative comparison of alternative methods for coarse-graining biological networks. *J. Chem. Phys.* 139:121905. doi: 10.1063/1.4812768
- Bowman, G. R., and Pande, V. S. (2010). Protein folded states are kinetic hubs. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10890–10895. doi: 10.1073/pnas.1003962107
- Buch, I., Giorgino, T., and De Fabritiis, G. (2011). Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10184–10189. doi: 10.1073/pnas.1103547108
- Buchete, N. V., and Hummer, G. (2008). Coarse master equations for peptide folding dynamics. *J. Phys. Chem. B* 112, 6057–6069. doi: 10.1021/jp0761665
- Cheatham, T. E., and Kollman, P. A. (2000). Molecular dynamics simulation of nucleic acids. *Annu. Rev. Phys. Chem.* 51, 435–471. doi: 10.1146/annurev.physchem.51.1.435
- Chiti, F., and Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* 75, 333–366. doi: 10.1146/annurev.biochem.75.101304.123901
- Chodera, J. D., Singhal, N., Pande, V. S., Dill, K. A., and Swope, W. C. (2007). Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* 126:155101. doi: 10.1063/1.2714538
- Cochran, W. T., Cooley, J. W., Favon, D. L., Helms, H. D., Kaenel, R. A., Lang, W. W., et al. (1967). What is the fast Fourier transform? *Proc. IEEE* 55, 1664–1674. doi: 10.1109/PROC.1967.5957

- Darden, T., York, D., and Pedersen, L. (1993). Particle mesh ewald - an N.Log(N) method for ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi: 10.1063/1.464397
- Deng, N. J., Dai, W., and Levy, R. M. (2013). How kinetics within the unfolded state affects protein folding: an analysis based on markov state models and an ultra-long MD trajectory. *J. Phys. Chem. B* 117, 12787–12799. doi: 10.1021/jp401962k
- Deuffhard, P., and Weber, M. (2005). Robust perron cluster analysis in conformation dynamics. *Linear Alg. Appl.* 398, 161–184. doi: 10.1016/j.laa.2004.10.026
- Eisenmesser, E. Z., Bosco, D. A., Akke, M., and Kern, D. (2002). Enzyme dynamics during catalysis. *Science* 295, 1520–1523. doi: 10.1126/science.1066176
- Gao, Y. Q., Yang, W., and Karplus, M. (2005). A structure-based model for the synthesis and hydrolysis of ATP by F-1-ATPase. *Cell* 123, 195–205. doi: 10.1016/j.cell.2005.10.001
- Gfeller, D., De Los Rios, P., Caflisch, A., and Rao, F. (2007). Complex network analysis of free-energy landscapes. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1817–1822. doi: 10.1073/pnas.0608099104
- Gong, L. C., and Zhou, X. (2010). Kinetic transition network based on trajectory mapping. *J. Phys. Chem. B* 114, 10266–10276. doi: 10.1021/jp100737g
- Gong, L. C., Zhou, X., and Ouyang, Z. C. (2015). Systematically constructing kinetic transition network in polypeptide from top to down: trajectory mapping. *PLoS One* 10:e0125932. doi: 10.1371/journal.pone.0125932
- Gregersen, N., Bross, P., Vang, S., and Christensen, J. H. (2006). Protein misfolding and human disease. *Annu. Rev. Genomics Hum. Genet.* 7, 103–124. doi: 10.1146/annurev.genom.7.080505.115737
- Guo, C., Luo, Y., Zhou, R. H., and Wei, G. H. (2012). Probing the self-assembly mechanism of diphenylalanine-based peptide nanovesicles and nanotubes. *ACS Nano* 6, 3907–3918. doi: 10.1021/nn300015g
- Hess, B., Kutzner, C., Van Der Spoel, D., and Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4, 435–447. doi: 10.1021/ct700301q
- Hovmöller, S., Zhou, T., and Ohlson, T. (2002). Conformations of amino acids in proteins. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 58, 768–776. doi: 10.1107/s0907444902003359
- Husic, B. E., and Pande, V. S. (2018). Markov state models: from an art to a science. *J. Am. Chem. Soc.* 140, 2386–2396. doi: 10.1021/jacs.7b12191
- Ithuralde, R. E., Roitberg, A. E., and Turjanski, A. G. (2016). Structured and unstructured binding of an intrinsically disordered protein as revealed by atomistic simulations. *J. Am. Chem. Soc.* 138, 8742–8751. doi: 10.1021/jacs.6b02016
- Jayachandran, G., Vishal, V., and Pande, V. S. (2006). Using massively parallel simulation and Markovian models to study protein folding: examining the dynamics of the villin headpiece. *J. Chem. Phys.* 124:164902. doi: 10.1063/1.2186317
- Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637. doi: 10.1002/bip.360221211
- Kampen, N. G. V. (2007). *Stochastic Processes in Physics and Chemistry*. Amsterdam: North Holland.
- Karplus, M., and McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9, 646–652. doi: 10.1038/nsb0902-646
- Lever, J., Krzywinski, M., and Atman, N. (2017). Principal component analysis. *Nat. Methods* 14, 641–642. doi: 10.1038/nmeth.4346
- Levy, Y., Wolynes, P. G., and Onuchic, J. N. (2004). Protein topology determines binding mechanism. *Proc. Natl. Acad. Sci. U.S.A.* 101, 511–516. doi: 10.1073/pnas.2534828100
- Li, C. X., Yang, D. M., Ma, P. A., Chen, Y. Y., Wu, Y., Hou, Z. Y., et al. (2013). Multifunctional upconversion mesoporous silica nanostructures for dual modal imaging and in vivo drug delivery. *Small* 9, 4150–4159. doi: 10.1002/smll.201301093
- Li, W. F., Wang, J., Zhang, J., Takada, S., and Wang, W. (2019). Overcoming the bottleneck of the enzymatic cycle by steric frustration. *Phys. Rev. Lett.* 122:238102. doi: 10.1103/PhysRevLett.122.238102
- Li, W. F., Zhang, J., Wang, J., and Wang, W. (2008). Metal-coupled folding of Cys(2)His(2) zinc-finger. *J. Am. Chem. Soc.* 130, 892–900. doi: 10.1021/ja075302g
- Millhauser, G. L., Stenland, C. J., Hanson, P., Bolin, K. A., and Vandeven, F. J. M. (1997). Estimating the relative populations of 3(10)-helix and alpha-helix in Ala-rich peptides: a hydrogen exchange and high field NMR study. *J. Mol. Biol.* 267, 963–974. doi: 10.1006/jmbi.1997.0923
- Mirny, L., and Shakhnovich, E. (2001). Protein folding theory: from lattice to all-atom models. *Annu. Rev. Biophys. Biomol. Struct.* 30, 361–396. doi: 10.1146/annurev.biophys.30.1.361
- Miyashita, N., Straub, J. E., and Thirumalai, D. (2009). Structures of beta-Amyloid Peptide 1-40, 1-42, and 1-55-the 672-726 Fragment of APP-in a Membrane Environment with Implications for Interactions with gamma-Secretase. *J. Am. Chem. Soc.* 131, 17843–17852. doi: 10.1021/ja905457d
- Moraitakis, G., Purkiss, A. G., and Goodfellow, J. M. (2003). Simulated dynamics and biological macromolecules. *Rep. Prog. Phys.* 66, 383–406. doi: 10.1088/0034-4885/66/3/203
- Naritomi, Y., and 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. doi: 10.1063/1.3554380
- Naritomi, Y., and 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. doi: 10.1063/1.4834695
- Noe, F. (2008). Probability distributions of molecular observables computed from Markov models. *J. Chem. Phys.* 128:244103. doi: 10.1063/1.2916718
- Noe, F., Horenko, I., Schütte, C., and Smith, J. C. (2007). Hierarchical analysis of conformational dynamics in biomolecules: transition networks of metastable states. *J. Chem. Phys.* 126:155102. doi: 10.1063/1.2714539
- Norberg, J., and Nilsson, L. (2002). Molecular dynamics applied to nucleic acids. *Accounts Chem. Res.* 35, 465–472. doi: 10.1021/ar010026a
- Onuchic, J. N., Lutheyschulten, Z., and Wolynes, P. G. (1997). Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48, 545–600. doi: 10.1146/annurev.physchem.48.1.545
- Pan, A. C., Jacobson, D., Yatsenko, K., Sriharan, D., Weinreich, T. M., and Shaw, D. E. (2019). Atomic-level characterization of protein-protein association. *Proc. Natl. Acad. Sci. U.S.A.* 116, 4244–4249. doi: 10.1073/pnas.1815431116
- Pande, V. S., Beauchamp, K., and Bowman, G. R. (2010). Everything you wanted to know about markov state models but were afraid to ask. *Methods* 52, 99–105. doi: 10.1016/j.ymeth.2010.06.002
- Paul, F., Wehmeyer, C., Abualrous, E. T., Wu, H., Crabtree, M. D., Schoneberg, J., et al. (2017). Protein-peptide association kinetics beyond the seconds timescale from atomistic simulations. *Nat. Commun.* 8:1095. doi: 10.1038/s41467-017-01163-6
- Perez-Hernandez, G., Paul, F., Giorgino, T., De Fabritiis, G., and Noe, F. (2013). Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* 139:015102. doi: 10.1063/1.4811489
- Prinz, J. H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., et al. (2011). Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.* 134:190401. doi: 10.1063/1.3565032
- Rao, F., and Karplus, M. (2010). Protein dynamics investigated by inherent structure analysis. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9152–9157. doi: 10.1073/pnas.0915087107
- Reuter, B., Weber, M., Fackeldey, K., Roblitz, S., and Garcia, M. E. (2018). Generalized markov state modeling method for nonequilibrium biomolecular dynamics: exemplified on amyloid beta conformational dynamics driven by an oscillating electric field. *J. Chem. Theory Comput.* 14, 3579–3594. doi: 10.1021/acs.jctc.8b00079
- Roblitz, S., and Weber, M. (2013). Fuzzy spectral clustering by PCCA plus : application to Markov state models and data classification. *Adv. Data Anal. Classif.* 7, 147–179. doi: 10.1007/s11634-013-0134-6
- Scherer, M. K., Trendelkamp-Schroer, B., Paul, F., Perez-Hernandez, G., Hoffmann, M., Plattner, N., et al. (2015). PyEMMA 2: a software package for estimation, validation, and analysis of markov models. *J. Chem. Theory Comput.* 11, 5525–5542. doi: 10.1021/acs.jctc.5b00743
- Schneider, A., Hommel, G., and Blettner, M. (2010). Linear regression analysis part 14 of a series on evaluation of scientific publications. *Dtsch. Arztebl. Int.* 107, 776–782. doi: 10.3238/arztebl.2010.0776
- Schwantes, C. R., and Pande, V. S. (2013). Improvements in markov state model construction reveal many non-native interactions in the folding of NTL9. *J. Chem. Theory Comput.* 9, 2000–2009. doi: 10.1021/ct300878a

- Sengupta, U., Carballo-Pacheco, M., and Strodel, B. (2019). Automated Markov state models for molecular dynamics simulations of aggregation and self-assembly. *J. Chem. Phys.* 150:115101. doi: 10.1063/1.5083915
- Sims, G. E., Choi, I. G., and Kim, S. H. (2005). Protein conformational space in higher order phi-psi maps. *Proc. Natl. Acad. Sci. U.S.A.* 102, 618–621. doi: 10.1073/pnas.0408746102
- Touw, W. G., Baakman, C., Black, J., Te Beek, T. A. H., Krieger, E., Joosten, R. P., et al. (2015). A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* 43, D364–D368. doi: 10.1093/nar/gku1028
- Wang, C. L., Lu, H. J., Wang, Z. G., Xiu, P., Zhou, B., Zuo, G. H., et al. (2009). Stable liquid water droplet on a water monolayer formed at room temperature on ionic model substrates. *Phys. Rev. Lett.* 103:137801. doi: 10.1103/PhysRevLett.103.137801
- Wang, W., Cao, S. Q., Zhu, L. Z., and Huang, X. H. (2018). Constructing markov state models to elucidate the functional conformational changes of complex biomolecules. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 8:e1343. doi: 10.1002/wcms.1343
- Weber, J. K., Jack, R. L., and Pande, V. S. (2013). Emergence of glass-like behavior in markov state models of protein folding dynamics. *J. Am. Chem. Soc.* 135, 5501–5504. doi: 10.1021/ja4002663
- Wei, G. H., Xi, W. H., Nussinov, R., and Ma, B. Y. (2016). Protein ensembles: how does nature harness thermodynamic fluctuations for life? the diverse functional roles of conformational ensembles in the cell. *Chem. Rev.* 116, 6516–6551. doi: 10.1021/acs.chemrev.5b00562
- Weng, J. W., and Wang, W. N. (2020). Dynamic multivalent interactions of intrinsically disordered proteins. *Curr. Opin. Struc. Biol.* 62, 9–13. doi: 10.1016/j.sbi.2019.11.001
- Wu, K., Xu, S., Wan, B., Xiu, P., and Zhou, X. (2020). A novel multiscale scheme to accelerate atomistic simulations of bio-macromolecules by adaptively driving coarse-grained coordinates. *J. Chem. Phys.* 152:114115. doi: 10.1063/1.5135309
- Yan, Z. Q., and Wang, J. (2019). Superfunneled energy landscape of protein evolution unifies the principles of protein evolution, folding, and design. *Phys. Rev. Lett.* 122:018103. doi: 10.1103/PhysRevLett.122.018103
- Yang, J. R., Shi, G. S., Tu, Y. S., and Fang, H. P. (2014). High correlation between oxidation loci on graphene Oxide. *Angew. Chem. Int. Ed.* 53, 10190–10194. doi: 10.1002/anie.201404144
- Yoder, G., Pancoska, P., and Keiderling, T. A. (1997). Characterization of alanine-rich peptides, Ac-(AAKAA)(n)-GY-NH₂ (n=1-4), using vibrational circular dichroism and Fourier transform infrared. Conformational determination and thermal unfolding. *Biochemistry* 36, 15123–15133. doi: 10.1021/bi971460g
- Zhang, C. B., Xu, S., and Zhou, X. (2019a). Identifying metastable states of biomolecules by trajectory mapping and density peak clustering. *Phys. Rev. E* 100:033301. doi: 10.1103/PhysRevE.100.033301
- Zhang, C. B., Ye, F. F., Li, M., and Zhou, X. (2019b). Enhanced sampling based on slow variables of trajectory mapping. *Sci. China Phys. Mech. Astron.* 62:62. doi: 10.1007/s11433-018-9313-1
- Zhang, C. B., Yu, J., and Zhou, X. (2017). Imaging metastable states and transitions in proteins by trajectory map. *J. Phys. Chem. B* 121, 4678–4686. doi: 10.1021/acs.jpcc.7b00664
- Zhou, H., Yang, Z. X., Tian, X., Chen, L., Lee, S., Huynh, T., et al. (2019). Lanosterol disrupts the aggregation of amyloid-beta peptides. *ACS Chem. Neurosci.* 10, 4051–4060. doi: 10.1021/acscchemneuro.9b00285
- Zhou, R. H., Huang, X. H., Margulis, C. J., and Berne, B. J. (2004). Hydrophobic collapse in multidomain protein folding. *Science* 305, 1605–1609. doi: 10.1126/science.1101176
- Zuo, G. H., Hu, J., and Fang, H. P. (2009). Effect of the ordered water on protein folding: an off-lattice Go-like model study. *Phys. Rev. E* 79(Pt 1):031925. doi: 10.1103/PhysRevE.79.031925
- Zuo, G. H., Kang, S. G., Xiu, P., Zhao, Y. L., and Zhou, R. H. (2013). Interactions between proteins and carbon-based nanoparticles: exploring the origin of nanotoxicity at the molecular level. *Small* 9, 1546–1556. doi: 10.1002/smll.201201381
- Zuo, G. H., Li, W. F., Zhang, J., Wang, J., and Wang, W. (2010). Folding of a small rna hairpin based on simulation with replica exchange molecular dynamics. *J. Phys. Chem. B* 114, 5835–5839. doi: 10.1021/jp904573r
- Zuo, G. H., Wang, J., and Wang, W. (2006). Folding with downhill behavior and low cooperativity of proteins. *Proteins Struct. Funct. Bioinform.* 63, 165–173. doi: 10.1002/prot.20857
- Zwanzig, R. (1983). From classical dynamics to continuous-time random-walks. *J. Stat. Phys.* 30, 255–262. doi: 10.1007/bf01012300

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Wang, Zhou and Zuo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.