


Origin of subdiffusions in proteins: Insight from peptide systemsChenliang Xia[✉], Xuefeng He, Jun Wang^{✉,*}, and Wei Wang^{✉,†}*School of Physics, Nanjing University, Nanjing 210093, People's Republic of China
and National Laboratory of Solid State Microstructure, and Collaborative Innovation Center of Advanced Microstructures,
Nanjing University, Nanjing 210093, People's Republic of China* (Received 29 August 2020; revised 9 November 2020; accepted 30 November 2020; published 28 December 2020)

Subdiffusive kinetics are popular in proteins and peptides as observed in experiments and simulations. For protein systems with diverse interactions, are there multiple mechanisms to produce the common subdiffusion behavior? To approach this problem, long trajectories of two model peptides are simulated to study the mechanism of subdiffusion and the relations with their interactions. The free-energy profiles and the subdiffusive kinetics are observed for these two peptides. A hierarchical plateau analysis is employed to extract the features of the landscape from the mean square of displacement. The mechanism of subdiffusions can be postulated by comparing the exponents by simulations with those based on various models. The results indicate that the mechanisms of these two peptides are different and are related to the characteristics of their energy landscapes. The subdiffusion of the flexible peptide is mainly caused by depth distribution of traps on the energy landscape, while the subdiffusion of the helical peptide is attributed to the fractal topology of local minima on the landscape. The emergence of these different mechanisms reflects different kinetic scenarios in peptide systems though the peptides behave in a similar way of diffusion. To confirm these ideas, the transition networks between various conformations of these peptides are generated. Based on the network description, the controlled kinetics based only on the topology of the networks are calculated and compared with the results based on simulations. For the flexible peptide, the feature of controlled diffusion is distinct from that of simulation, and for the helical peptide, two kinds of kinetics have a similar exponent of subdiffusion. These results further exemplify the importance of the landscape topology in the kinetics of structural proteins and the effect of depth distribution of traps for the subdiffusion of disordered peptides.

DOI: [10.1103/PhysRevE.102.062424](https://doi.org/10.1103/PhysRevE.102.062424)**I. INTRODUCTION**

Subdiffusions are typical kinetic behaviors in many complex systems, indicating a slow relaxation and long memory in the kinetics and reflecting the existence of multiple timescales related to the competition of multiple energy scales. This kind of kinetics has attracted much attention in various disciplines [1–4]. Proteins are typical systems with multiscale subdiffusion kinetic behaviors. For example, the slow relaxations and long-time correlations are observed in experiments for many different protein systems based on various techniques [5–9], including IR spectrum [5], fluorescence fluctuations [6], neutron scattering [7], single-molecular electron transfer experiments [8], and so on, which indicates that the subdiffusion kinetics are prevalent in protein systems. The concerned timescales span from microseconds to seconds. This kind of kinetic behavior is observed more frequently in simulations [10–17] by investigating the diffusions along the reaction coordinates [10] or principal components [11,12]. Even a larger range of timescales are observed from picoseconds to microseconds. The popularity of the subdiffusive kinetics in protein systems suggests that there are some intrinsic features of the interactions or structural organizations in protein

systems, and stimulated much research on the dynamics of proteins.

There are many complex interactions in protein systems. What kinds of physical ingredients in protein systems produce the subdiffusive behaviors in such a wide range of timescales? There are long-lasting discussions on this topic. First, these kinds of kinetic features are ascribed to the characteristics of solvent or solvent-protein interactions [18,19]. Some models are proposed with fractional Gaussian noises or are described with fractional Brownian dynamics [7,9,16]. Even a simple harmonic potential is involved to depict the interactions in proteins, the memory effect, and nonexponential (typically power-law) relaxation would be produced [6]. This kind of model can phenomenologically describe the single-molecular experiments about the equilibrium fluctuation of the distance between an electron transfer donor and acceptor pair within an enzyme [6]. The source from noise implies that the subdiffusion is driven by the solvent or through the interactions between the solvent and the fractional surface and/or structure of a protein. Since the aqueous environment is generally involved in protein systems, the water-slaved picture is a plausible source of subdiffusions in proteins.

However, the studies based on ultrafast spectroscopy indicate the internal interactions in a protein system have essential contributions to their dynamics [20,21]. Besides, the simulations with implicit solvents (where the kinetics of solvent molecules are discarded) also demonstrate the

*wangj@nju.edu.cn

†wangwei@nju.edu.cn

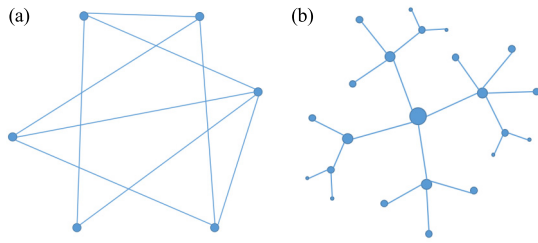


FIG. 1. Illustrations of (a) random network and (b) fractal network. For proteins, the vertices represent the conformations and the edges indicate the jumps between different conformations.

existence of subdiffusions [22]. To investigate the effects of the interactions in proteins is also important. To consider the heterogeneity of the protein interactions, the trapping model [6,11,13,14,23,24] is frequently used by considering the resemblance between proteins and glasses, that is, various local equilibrium states are randomly distributed on the landscape, which produce various traps with diverse depths [as shown in Fig. 1(a)]. In these models, multiple timescales come from the wide distribution of diverse escaping times. Continuous time random walk (CTRW) [25–31] is often used to characterize the kinetics of these models, and the waiting times between two jumps are picked from a distribution of the escaping times. This kind of model is more convenient to characterize the behavior on the high-dimensional landscape and provides the flexibility to describe the wide temporal correlations in subdiffusive kinetics, which gives a reasonable description for experiments and simulations. For instance, a long relaxation time can be achieved in the trapping model, while the typical relaxation time for harmonic models is about $10^{-8} \sim 10^{-9}$ s [32].

Indeed, the model of random traps is not a perfect description for the interactions in proteins [12]. In addition, the trapping model cannot be used to depict the subdiffusion at equilibrium, and the CTRW is not time invariant and indicates a kinetics far from equilibrium [25–31]. Physically, the native structure of a protein is generally selected by natural evolution and the energy landscape is largely determined by the native structure based on the funnel concept. The traps are not random but physically correlated. It is possible to postulate that the organization of the traps may provide additional mechanisms to affect the kinetics. Some traps may have more neighbors, while some others have fewer ones. Fractal energy landscapes with hierarchical structures [10,12,33] are postulated [as shown in Fig. 1(b)]. The topology of the network may produce different diffusive abilities on the landscape, since the system would like to wander around the central hubs which slow down the diffusion in conformational space. This kind of effect is related to the pathway entropy and is coupled with the contributions of the energy depths. This kind of picture has also been described recently by Meroz *et al.* [15]. These results demonstrate rich physics in the subdiffusive kinetics of protein systems.

During recent years, more complexities have been observed for protein systems. There are many proteins without stable native structures in natural condition. They are named intrinsically disordered proteins (IDPs) [34]. The interactions

in these proteins are largely different from typical globular proteins. Yet the subdiffusions are still observed for these protein systems [12,35]. From the view of relaxation kinetics, proteins with different structural features behave similarly. Are the mechanisms the same for two kinds of peptides? Indeed, the local interactions between the building blocks (amino acids) of these proteins are similar but there are clear differences of landscape architecture between the two kinds of proteins. The globular proteins generally have a funnel-shaped energy landscape due to the cooperative interactions, as suggested by Anfinsen principle [36]. Meanwhile, the IDP chains are only restricted by the peptide connectivity, and flat but rugged energy landscapes are more popular in IDP systems. It is widely known that the kinetics of proteins are determined by their energetic landscape. The long-time asymptotical behaviors are generally controlled by the large-scale features of the energy landscapes. This knowledge implies that there are different mechanisms related to the subdiffusive kinetics of the globular proteins and IDPs. What are the connections between the features of energy landscape and their subdiffusive kinetics? What are the dominant mechanisms of subdiffusions in different protein systems? The answers to these questions would complete our understanding on the subdiffusions in proteins and help to disclose the structure-dynamics relations of proteins, which are the tasks of the present paper.

In this paper, the mechanisms of subdiffusions are investigated based on the molecular simulations. Two peptides with different structural features and energy landscapes, GS₄ and A₈, are picked as the model systems for IDP and globular proteins. Their kinetics are simulated and the interactions and the energy landscape of these peptides are characterized. The relationship between mean square of displacement (MSD) and the lag time are determined based on the trajectories. The exponents describing the feature of subdiffusion are determined by fitting the MSD curve. From the MSD curves, the features of the landscape (including the geometric size and energetic height of various valleys) are determined based on the diffusive properties at different scales of time. According to these features of landscapes, several exponents can be calculated for the diffusions based on various kinds of assumptions. The comparison of these exponents with that from simulations suggests the contribution of certain physical ingredients. The results show that the depth distribution of traps on the landscape is the main reason to produce subdiffusion of the peptide GS₄, and the diffusive kinetics of the peptide A₈ is largely related to the fractal topology of the accessible landscape. To further confirm the connection between the landscape and the mechanism of kinetics, a network description for the peptide kinetics is generated based on the conformational clusters. The controlled kinetics are created based on the topology of the network. The comparison between the controlled kinetics and the realistic one supports the view on the mechanisms of two kinds of peptides.

II. MODEL AND METHODS

A. Model

There are a large number of proteins with different sizes, diversified compositions, and complex interactions. Since

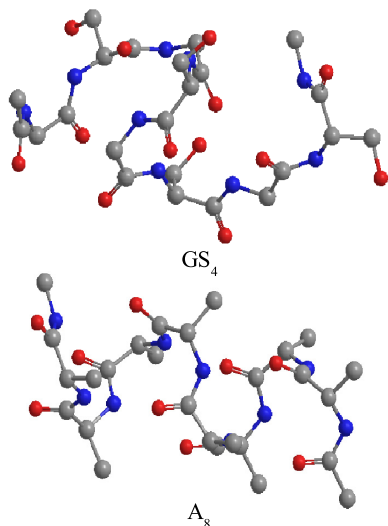


FIG. 2. The snapshots of the peptides GS_4 and A_8 . The carbon, oxygen, and nitrogen atoms are represented as grey, red, and blue spheres, respectively. The N and C termini of these two peptides are capped with acetyl and N -methylamide, respectively.

their dynamics largely depend on their structural features in the native state, the size scaling of various properties is not explicit and the comparisons between different proteins is not a trivial task. It is necessary to select some specific systems to suppress some unexpected factors. In addition, the characterization of the long-time relaxation of proteins requires long simulations to sample throughout the essential parts of the energy landscape. Presently, the simulations of proteins with all-atom models are still computationally demanding tasks for a generic protein. Some model systems are expected to surmount the practical difficulties. Based on these considerations, we pick two peptides as the objects in our investigations. These peptides have the same lengths but different structural features (as shown in Fig. 2). One is composed of four glycines and four serines (named GS_4) which are arranged in an interlaced manner. The other contains eight alanines, which is a typical helix former. The N and C termini of these two peptides are capped with acetyl and N -methylamide, respectively, to remove the effects of the terminal charges. This is important for short peptides. Based on the knowledge of proteins, these peptides have different structural features. The peptide GS_4 is rather flexible because of the small and polar side chains (even only a hydrogen atom as the side chain for glycine). Thus, this peptide generally takes random and fluctuating conformations and is frequently used as the linker loop in molecular engineering. Here, we use this peptide to mimic the IDP based on its structural features. Differently, the peptide A_8 has an energetically stable conformation in the helical shape and is often used as the typical system to investigate the helix-coil transition. This peptide can be considered as a minimal model of proteins with a specific native structure. We believe that, through the analysis of these peptides, the mechanisms of various proteins with different interactions and structural features can be determined.

B. Simulation setup

To simulate the dynamics of these concerned peptides, GROMACS 5.1.4 molecular simulation package [37] is employed. Amber ff03 force field [38] and SPC/E water model [39] are used. Initially, cubic boxes filled with water molecules are set up for both peptides. The size of box edge is 3.18 nm with 998 water molecules for GS_4 and 3 nm with 857 water molecules for A_8 , respectively. We carry out the simulations with the NPT ensemble. The temperature is set as 300 K and the pressure is set as 1 bar. The temperature controlling is implemented through the coupling with a thermostat through rescaling of velocity [40] and the pressure is held by a Parrinello-Rahman barostat [41]. The bulk compressibility is $4.5 \times 10^{-5} \text{ bar}^{-1}$. The particle Mesh Ewald method is used for long-range electrostatics, and the LINCS algorithm is used for all constraints of bonds. The cutoffs of the short-range neighbor list, the electrostatic interaction, and the van der Waals interaction are all 0.9 nm and the neighbor list is updated every 20 steps. The simulation time step is picked as 2 fs. The trajectories are up to 6 μs (namely, 3×10^9 steps) and we store the snapshots of the trajectories every 5000 steps.

C. Coordinates of landscapes

To characterize the free-energy landscape and the conformational dynamics, a projection to some certain subspaces may be helpful to illustrate the key features of the original conformational space. For example, as for polymers, the distance between the C_α atoms of the first and the last residues (named end-to-end distance Q_{ee}) can be used to characterize the overall size of the concerned polypeptide. For the peptide GS_4 , the structures are generally largely fluctuating and the geometrical shape would be statistically isotropic. This kind of feature looks like that for an athermal polymer. The characterization with Q_{ee} is reasonable due to its self-averaging feature. Meanwhile, for peptide A_8 , the helical structure is apparently not spherical. The end-to-end distance can characterize some part of the information about its shape. Yet some additional information is still necessary, which reflects the necessity of multiple dimensions to describe helical peptides. Due to the complexity of the landscape, the choice of suitable coordinates is not a trivial task.

Generally, the proper coordinates can be determined based on the data from sufficient sampling of the landscape. A typical method is the principal component analysis (PCA) based on a large set of conformations for our peptide systems. Practically, the PCA operations can be carried out based on the Cartesian coordinates (named xPCA) or dihedral angles (named dPCA) of the peptides. For the xPCA, to characterize the feature of the internal structures, the translational and rotational operations should be applied at first to remove global motions, which is realized by minimizing the root mean square of distances with respect to a certain reference state. Practically, the first frame along the simulation trajectory is selected as the reference state. Other choices of the reference states are also tested. The results are similar (data not shown). After this kind of structural alignment, the covariance C_{ij}

between the i th and j th monomers can be determined as

$$C_{ij} = \frac{1}{Q} \sum_{q=1}^Q (\Delta \vec{r}_{i,q} \cdot \Delta \vec{r}_{j,q}), \quad (1)$$

in which $\Delta \vec{r}_{i,q}$ is the displacement of i th residue in q th snapshot from the corresponding average position as

$$\Delta \vec{r}_{i,q} = \vec{r}_{i,q} - \frac{1}{Q} \sum_{q=1}^Q \vec{r}_{i,q}. \quad (2)$$

The eigenvectors of the corresponding covariance matrix $\{C_{ij}\}$ give a linear characterization of conformational fluctuations. Since we mainly focus on the backbone motions, it is possible to choose the C_α atoms of residues in analysis, similar to the analysis related to the Gaussian network model [42,43]. This kind of coarse graining can help us focus on the global structural variations. It is worth pointing out that the xPCA for flexible peptides may introduce remarkable local structural distortions in various modes [44,45], which is ascribed to the omissions for the bond and angle constraints in the related xPCA. The dPCA based on backbone dihedral angles is a better choice. Based on the backbone dihedrals (ϕ and ψ), the conformations can be represented as a sequence of dihedrals, ω_i with $0 \leq i < 2N$, where N is the number of residues, $\omega_{2k} = \phi_k$ and $\omega_{2k+1} = \psi_k$, and k is the sequential index of a residue. The conformations can be described as a series of vectors $\mathbf{u}_i = (\cos(\omega_i), \sin(\omega_i))$. Correspondingly, the covariance matrix can be calculated based on the vectors \mathbf{u}_i as

$$C_{i,j} = \frac{1}{Q} \sum_{q=1}^Q (\Delta \mathbf{u}_{i,q} \cdot \Delta \mathbf{u}_{j,q}), \quad (3)$$

$$\Delta \mathbf{u}_{i,q} = \mathbf{u}_{i,q} - \frac{1}{Q} \sum_{q=1}^Q \mathbf{u}_{i,q}. \quad (4)$$

Typically, for both kinds of PCA, the first several eigenvalues are apparently larger than the others. The corresponding eigencomponents can offer a reasonable approximation for the covariance matrix and construct a rational subspace to describe the conformations concisely. These components are called principal components. Practically, we pick the principal components based on the largest gap in the eigenvalue spectrum. In the subspace spanned with these components (eigenvectors), a reduced description for conformations can be obtained by projecting the conformations on these eigenvectors. These protocols give a set of self-consistent coordinates based on data and provide a rational way to characterize the essential features of conformations and kinetics.

D. Transition network

Considering the complex topology and interaction of proteins, the discovery of reaction coordinates and concise subspace is often not easy. The network language provides another way to describe the dynamics. That is, the conformations are clustered based on their neighborhood, and the clusters are defined as the vertices of the conformational network. With these vertices, the edges between them are determined based on their kinetic accessibility according to the trajectories. The vertices and edges build up a network picture for

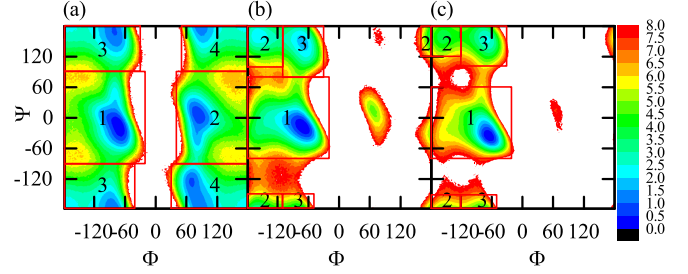


FIG. 3. Ramachandran map based on the backbone dihedral angle ϕ and ψ for (a) glycine, (b) serine, and (c) alanine. The unit of energy is $k_B T$. In each plot, several regions are marked with numbers. Each region represents a state of residue with similar structural features. For the residue glycines, four states are introduced, and for serines and alanines, three states are defined.

conformational space. In our work, the clustering is based on the internal dihedral coordinates. This is valid for present peptide systems with small sizes. Practically, based on the ϕ - ψ distribution (namely, the distribution on Ramachandran plot) of each kind of residue, the subspace spanned with ϕ and ψ can be divided into several regions. For the residue glycines, four states are introduced, and for serines or alanines, three states are defined accordingly, as shown in Fig. 3. Considering all the combinations, there are $4^4 \times 3^4 = 20736$ states for the peptide GS_4 , and $3^8 = 6561$ states for A_8 , respectively. Note that not all states can be visited in simulations, since the exclusive volume effect may prohibit some combinations of the dihedrals. To unify the transition time in the network, the lag time Δ_{lag} is set as 10 ps. These operations define the basic mapping from conformational space to the conformational cluster transition network (CCTN). In the visualization of the network, the populations of the clusters and the strengths of transition rates are represented by the sizes of the vertices and the thickness of the edges.

Quantitatively, on the network, the kinetics can be described with the corresponding transition matrix, whose elements record the transition probabilities between the i th and j th vertices (clusters), $T_{ij}(\Delta_{\text{lag}})$. The transition probability $T_{ij}(\Delta_{\text{lag}})$ can be calculated based on the number of the transition events as

$$T_{ij}(\Delta_{\text{lag}}) = \frac{N_{ij}(\Delta_{\text{lag}})}{\sum_k N_{ik}(\Delta_{\text{lag}})}, \quad (5)$$

where $N_{ij}(\Delta_{\text{lag}})$ is the number of transitions from i th vertice to j th vertice within the time Δ_{lag} . Note that the retention probability $T_{ii}(\Delta_{\text{lag}})$ can be determined based on the normalization condition $\sum_j T_{ij}(\Delta_{\text{lag}}) = 1$. With the transition matrix, the evolution of the conformation ensemble can be calculated based on the master equation as $\mathbf{p}(\tau + \Delta_{\text{lag}}) = T\mathbf{p}(\tau)$, where $\mathbf{p}(\tau)$ gives the distribution of conformations on the network.

E. Analysis on hierarchical plateaus on the landscape

Besides the direct mapping of the landscape, the features of the landscape can also be analyzed through the dynamic fluctuations during the evolution. The method by Meroz *et al.* [15] is along this direction by extracting information from the local features of the MSD curve. In this method, the kinetics

is fitted based on a model of hierarchical energy landscape. It is assumed that there is a hierarchy of energy traps on the landscape. In this hierarchy, the i th tier contains the valleys (characteristic size as L_i) with energy barriers (characteristic energy E_i). In these valleys, there are smaller valleys belonging to the tier $i - 1$. The sizes and energy barriers satisfy the relations $L_{i-1} < L_i$ and $E_{i-1} < E_i$. There are multiple tiers in peptide systems. The set $\{L_i, E_i\}$ gives a characterization for the landscape of the peptide. With this model, the kinetics of the system can be determined by these features of the landscape. For example, there are several characteristic times for the tier- i valleys. After the time τ_i^c , the system can travel across the valley. After the time τ_i^e , the system can overcome the energetic barrier and escape from the present valley. The ratio between these timescales can be estimated based on the characteristic energy E_i as $\tau_i^c/\tau_i^e \propto \exp(-E_i/k_B T)$, where k_B is Boltzmann constant and T is the temperature. During the time τ ($\tau_i^c < \tau < \tau_i^e$), the diffusion of the system would be limited in the valley and the MSD would be estimated by the size of the valley as $\langle r^2(\tau) \rangle \propto L_i^2$. Based on these relations, the features L_i and E_i can be estimated based on the characteristic times.

How can we extract these characteristic times? It is suggested that the MSD curves encode the concerned information. The local plateaus in the MSD curves indicate the local equilibrium in certain valleys on the landscape and are tightly related to the characteristic times of the relaxation. Mathematically, the timescales corresponding to the plateaus can be determined based on the minima of the derivatives α_{local} of the MSD, $\alpha_{\text{local}}(\tau) = d\langle r^2 \rangle / d\tau$, which can be considered as an exponent of local diffusion. Practically, the derivative α_{local} is calculated numerically. Due to the numerical fluctuations of the MSD curve, there are some fluctuations for the exponent α_{local} . This may produce some artificial minima of α_{local} , which do not imply the plateaus of the MSD curve. These minima are generally observed within a short timescale due to the fast oscillations of the fluctuations. Therefore, the minima which happen within 100 ps are discarded to suppress the fluctuations. When a plateau is identified, the onset time $\hat{\tau}_i^c$ and the end time $\hat{\tau}_i^e$ of the plateau can be defined when the second derivative of the MSD curve is smaller (or larger) than a threshold ϵ . Physically, the time τ_i^c of tier i can be estimated based on the relation by considering the effect of the motions in the valley of tier- $(i - 1)$ as

$$\tau_i^c \propto \hat{\tau}_i^c \frac{\tau_{i-1}^c}{\tau_{i-1}^e}. \quad (6)$$

Especially, at tier-1, $\tau_1^c = \hat{\tau}_1^c$. In addition, the escaping time $\tau_i^e = \hat{\tau}_i^e$. With these relations, we can determine the characteristic timescales and the features of landscapes.

To produce diffusive kinetics based on the features of the landscape, additional assumptions are needed [15]. If the energy landscape has a fractal topology, the diffusive behavior attributed to this kind of feature can be estimated based on the crossing times and the size of valleys as

$$\tau_i^c \propto L_i^{1/\alpha_f}. \quad (7)$$

If the landscape is not fractal, this component is $\alpha_f = 1$. More frequently, there is a logarithm relationship between the

energy depth and the size of the valleys as

$$E_i = E_0 + \gamma \ln(L_i/L_0), \quad (8)$$

γ is a constant. On such a landscape, the diffusive kinetics may satisfy the relation

$$\langle r^2(\tau) \rangle \propto \tau^{2/(\gamma+1/\alpha_f)} \equiv \tau^\alpha, \quad (9)$$

where the exponent $\alpha = 2/(\gamma + \alpha_f^{-1})$, which describes integrative effects from the landscape topology and trap depth. If the landscape of the system is not fractal, the kinetics would be determined only by the trapping effect and the corresponding exponent α_t can be simplified as

$$\alpha_t = \frac{2}{1 + \gamma}. \quad (10)$$

Therefore, based on various assumptions of landscape, the diffusive exponent can be estimated. The comparison between these theoretical estimations and the actual value from the MSD curve can help to check the validity of the assumptions.

III. RESULTS AND DISCUSSIONS

A. Equilibrium of two peptides

To analyze the diffusive features of these two peptides, it is necessary to generate a continuous trajectory to carry out temporal correlation analysis. Considering the complexity of the landscape, a sufficient long trajectory is expected to cover the essential parts of the landscape. Facing these requirements, long continuous simulations are carried out for these two peptides, up to 6 μ s. Parts of these trajectories are also checked, and the similar subdiffusions are observed.

The evolutions of the end-to-end distance Q_{ee} for these peptides are shown in Figs. 4(a) and 4(b). As a remark, the structural fluctuations of the peptide GS₄ are much more frequent than those of peptide A₈. To better illustrate the characteristics of the variations of the peptides, different lengths of trajectories are shown in the Fig. 4 (100 ns for GS₄ and 6 μ s for A₈, respectively). This kind of consideration is also used in other presentations about the evolutions of the peptides. All other calculations are not limited to the part of the trajectories, and the total trajectories with the length of 6 μ s are used. In these trajectories, the distance Q_{ee} fluctuates around a certain value. For the peptide GS₄, a large Q_{ee} corresponds to extended conformations, while the conformations with small Q_{ee} may have the shape of a loop with its C and N termini close to each other. Some typical snapshots are given accordingly. The fluctuations of the Q_{ee} represent the conformational variations of the peptides. Based on the trajectory, the typical timescale of the conformational conversion from extend coil to compact loop is about several ns. During the whole trajectory, there are thousands of rounds of conversions, which ensure the equilibrium of the peptide. For peptide A₈, there is a similar fluctuating picture for the distance Q_{ee} , which represents the repeated wandering on the landscape. Meanwhile, peptide A₈ is somehow different from GS₄. Peptide A₈ has a more remarkable population corresponding to the middle Q_{ee} . Based on the snapshots, the most probable conformation is in a helical form. This is consistent with common knowledge for peptide A₈. Besides this kind of popular structure, the conformations with large Q_{ee} or small Q_{ee} are all deformations

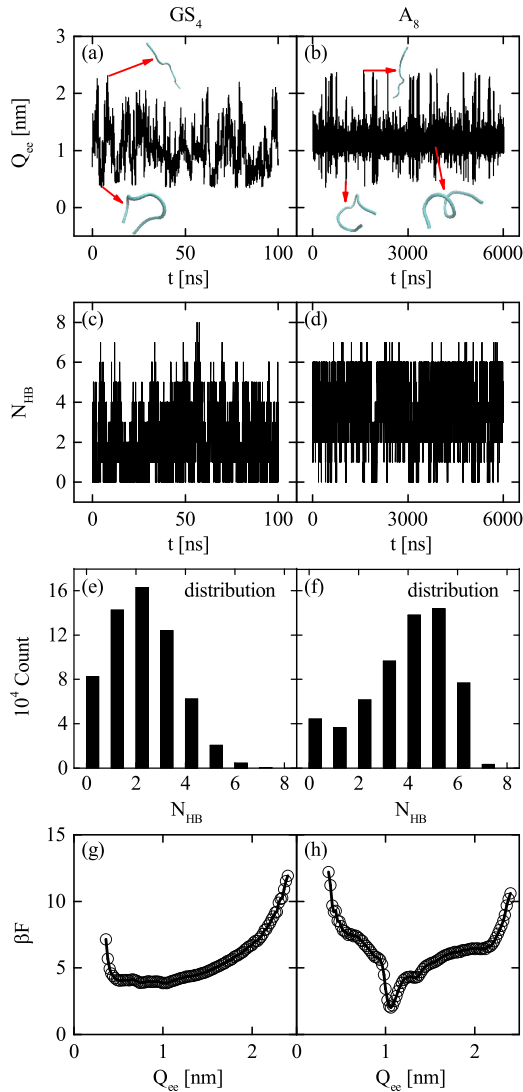


FIG. 4. The figure gives the evolutions of the end-to-end distance Q_{ee} for (a) GS₄ and (b) A₈, the evolution of the number N_{HB} of intrapeptide hydrogen bonds for (c) GS₄ and (d) A₈, the histogram of the number N_{HB} during the simulation for (e) GS₄ and (f) A₈ and the free-energy profiles of the peptides (g) GS₄ and (h) A₈ with the end-to-end distance as the reaction coordinate. Some typical snapshots of the peptides at certain times are shown (for GS₄, 100 ns of trajectories are used for clarity of presentation).

of the helix structure. That is, the ones with large Q_{ee} are extended and can be considered as the results by stretching the helix. Those conformations with small Q_{ee} are compact, corresponding to the bending of the helix so the head and tail of the helix meet each other. These kinds of shape variations are also observed in literature [46–48]. Another difference between peptides A₈ and GS₄ is that the typical timescale related to structural fluctuations for A₈ is much longer than that for GS₄. From Fig. 4(b), the timescale for A₈ is more than tens of nanoseconds. In fact, peptides A₈ and GS₄ have the same length. The longer characteristic time of fluctuations is ascribed to the long dwell time in the state with helical structure, which reflects the additional stability of the helical

structure. This demonstrates that there are clear differences of physics for peptides A₈ and GS₄.

Based on these trajectories, effective free-energy profiles can be calculated along the coordinate Q_{ee} , as shown in Figs. 4(g) and 4(h). For peptide GS₄, the free-energy profile has a flat shape in the range from about 0.4 nm to about 2 nm, which covers the essential part of the landscape and includes about 90% populations. This indicates that there is no (or weak) dependence of free energy on the shape of the peptide GS₄. The quick rise at small Q_{ee} originates from the exclusive volume effect of the terminal residues, while the gradual increase of free energy for large Q_{ee} reflects the entropic elasticity of the peptide chain. This feature also implies that the intrapeptide interactions are not strong and the chain behaves like an athermal polymer chain. For peptide A₈, the free-energy profile has a different shape from that for GS₄. A clear dip is observed around 1 nm, corresponding to the helical structure. Around this minimum, there is a quick increase of free energy, which forms a funnel in the range between 0.96 nm and 1.2 nm. About 72% of the population of the simulation is located in the funnel. Outside of the funnel, the profile is relatively flat, which is similar to those for GS₄. This is a reflection of the polymeric features of peptide chains. This kind of landscape is an example of the landscape of globular protein. This observation is consistent with helix-coil transition for A₈ [49]. With these features, it is reasonable to conclude that GS₄ and A₈ can be used as examples for IDP and globular proteins.

Physically, the variations of the structure and the free energy are driven by the intrapeptide interactions. Considering the compositions of the peptides, the number of the intrapeptide hydrogen bonds, N_{HB} , is a good indicator for the concerned interactions. Practically, the hydrogen bonds are measured based on the local geometry of interacting atoms. That is, the hydrogen bond is defined when the angle donor-hydrogen-acceptor is smaller than 30° and the distance between the donor and the acceptor is not larger than 0.35 nm. It is observed that the variations of the number N_{HB} share similar behaviors as those for Q_{ee} . In detail, for peptide GS₄, the states with fewer hydrogen bonds are generally extended, while those with more hydrogen bonds are often compact. The variation of N_{HB} is concurrent with the fluctuation of the shape. Yet peptide A₈ is a little different. The states with more hydrogen bonds correspond to the helical state and the extended and compact structures are all with fewer hydrogen bonds. The fluctuation of N_{HB} looks more frequent compared to that for Q_{ee} . Quantitatively, the distribution of N_{HB} can be determined based on the statistics on the trajectories in Figs. 4(c) and 4(d), similar to the analysis for free energy. For GS₄, the distribution of N_{HB} has a single peak, which indicates that there is only a single thermodynamic state. For this distribution, the most probable number is small, which indicates that the interactions are weak and cannot stabilize any specific structures. This is consistent with the picture of athermal polymer, and is similar to the behavior of IDP. Differently, for A₈, there is a small plateau around the region with small N_{HB} besides the peak around $N_{HB} = 5$. The most probable number (namely, $N_{HB} = 5$) is the same as the number of hydrogen bonds in the helix structure. This suggests the dominance of the helical structure for A₈. The existence of the small plateau

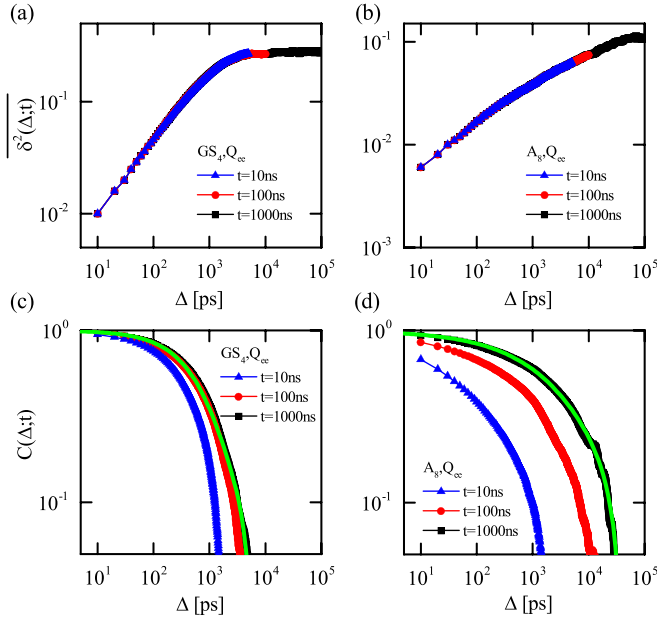


FIG. 5. This figure gives the TA-MSD (upper row) and autocorrelation functions (lower row) for peptides GS_4 (left column) and A_8 (right column). Green lines in (c) and (d) are the fittings for 1000 ns trajectories with Eq. (11).

reflects the tendency toward disordered structures. This kind of distribution implies a two-state feature for the peptide and is consistent with the picture from the corresponding free-energy profile. Note that this kind of characteristic is not strong due to the small size of A_8 . It is clear that a more comprehensive understanding can be achieved through the analysis of the interactions.

It is widely suggested that the subdiffusions of peptides are related to the nonergodic kinetics. To exam the aging effect in the kinetics of these two peptides, the time-averaged MSD (TA-MSD) and observation-time-dependent autocorrelation functions are calculated based on the trajectories with different lengths. Similar actions were carried out in literature [10]. Here, the MSD is calculated based on the end-to-end distance of the peptides. In Figs. 5(a) and 5(b), the slowing down of the diffusion is demonstrated. The slope becomes smaller following the increasing observation time (namely, the longer trajectory). This indicates an observation-time-dependent kinetics, and implies the aging in the kinetics. In Figs. 5(c) and 5(d), the correlation functions shift toward longer lag times with increasing observation time. This is another phenomenon of aging behavior. As a remark, for the peptide GS_4 , the difference of the correlation functions between the 100 ns and 1000 ns trajectories are relatively smaller, which implies that the peptide is approaching the thermodynamic equilibrium. This is ascribed to the small size of the concerned peptide and is consistent with the behaviors suggested from TA-MSD. Furthermore, based on the TA-MSD and autocorrelation function, the correlation functions can be well fitted with the equation

$$C(\Delta; t) = c_1 \exp[-(\Delta/\tau)^\beta] + c_2 B(\Delta/t, \alpha, 1 - \alpha) + c_3, \quad (11)$$

TABLE I. Eigenvectors corresponding to the dominated eigenvalues for GS_4 and A_8 .

Eigenvalue	1	2	3	4	5	6	7	8
0.19 (GS_4)	(-0.73	-0.21	0.32	0.31	-0.06	-0.18	0.13	0.41)
0.07 (A_8)	(-0.52	-0.37	-0.30	-0.12	0.14	0.29	0.38	0.50)
0.05 (A_8)	(0.40	0.22	-0.11	-0.54	-0.50	-0.10	0.19	0.45)

where Δ is the lag time of correlation time, t is the length of the concerned trajectory, $B(x, \alpha, 1 - \alpha)$ is the incomplete beta function, and $1 - \alpha$ is the exponent of subdiffusive TA-MSD, and the quantities τ , β , and $c_{1,2,3}$ are parameters to be determined based on fitting [as the green lines shown in Figs. 5(c) and 5(d) for the data related to 1000 ns trajectories]. This relation is derived based on the CTRW model in literature [10]. This indicates that the kinetics of peptide can be described by a noisy CTRW model and suggests the possible nonergodicity during the subdiffusive kinetics.

B. Projected kinetics of peptides based on PCA

For the peptides investigated here, the kinetics are not apparently all or none. This implies that multiple dimensions are necessary to depict the behaviors of the peptides. The coordinate Q_{ee} might not be a good coordinate since it does not capture the features of the native states of the peptides. It is necessary to propose a rational subspace to describe the conformations and their diffusive kinetics.

Here we employ the PCA method to generate the rational subspace and characterize the diffusive features accordingly based on the throughout sampling on the landscape. It is already observed in literature that the evolution in the subspace spanned with the dominant principal components contains most of the information of the kinetics [50–52]. As described in the Methods section, the C_α atoms are considered only in the xPCA. Thus, a 8×8 covariance matrix $\{C_{ij}\}$ is generated. In detail, the eigenvalue spectra based on xPCA for these two peptides are shown in Figs. 6(a) and 6(b). These eigenvalues correspond to various kinetic modes of the peptides. For the peptide GS_4 , there is a large gap between the largest eigenvalue and the others. This indicates that it is possible to describe the whole covariance matrix only with the component corresponding to the largest eigenvalue. A one-dimensional subspace can be deduced based on this principal component. Differently, for peptide A_8 , the distribution of eigenvalues is a little different from that for GS_4 . There are no dominant eigencomponents. The largest gap between eigenvalues is between the second largest eigenvalue and the rest of the smaller ones. This moderate gap implies a reduced two-dimensional description of the conformational space based on the first two eigencomponents. Practically, with these principal components, the elements of the covariance matrix can be reproduced with the error smaller than 30%. The features of these eigenvalue spectra outline the rationality of the xPCA representation.

It is worth pointing out that these essential components encode the kinetic characteristics of the peptides. As shown in Table I, the eigenvectors based on xPCA corresponding to these essential components are given. Based on the meaning

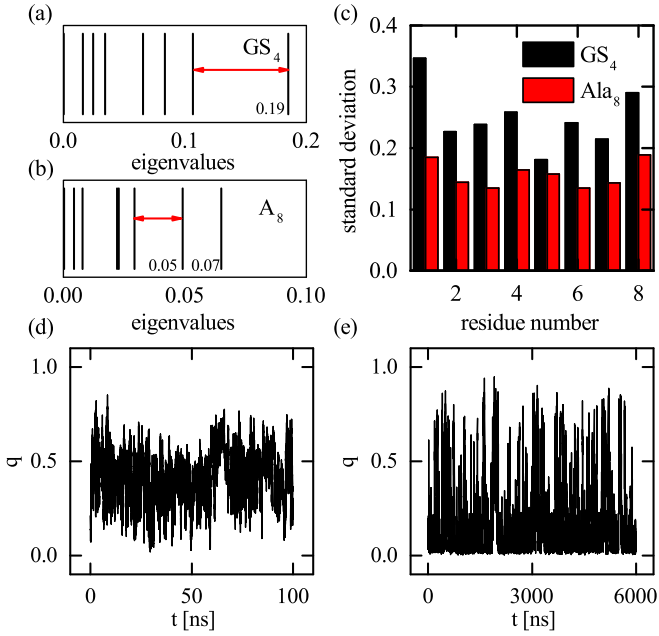


FIG. 6. This figure gives the spectra of eigenvalues of covariance matrix for the peptides (a) GS_4 and (b) A_8 , (c) the root mean square of fluctuations of residues in these two peptides (black for GS_4 and red for A_8), and the evolution in the subspaces spanned with essential eigencomponents for (d) GS_4 and (e) A_8 (for GS_4 , only 100ns of trajectories are used for clarity of presentation). The largest gaps of eigenvalues for these peptides are marked with red arrows in (a) and (b).

of the essential components, the elements of these eigenvectors reflect the fluctuations of the corresponding residues. For peptide GS_4 , the absolute values of the first and eighth elements are apparently larger than the others. It indicates that the first and eighth residues have stronger fluctuations than the others, which is confirmed by the fluctuations based on simulation data [as the black bar shown in Fig. 6(c)]. This kinetic behavior is physically reasonable. For a flexible chain such as GS_4 , the head and tail units are less constrained compared to the middle ones and large fluctuations are produced for the head and tail units, as suggested by the essential component of GS_4 . For peptide A_8 , the situation is a little bit complex. The feature of the eigenvector corresponding to the largest eigenvalue (0.07) is similar to that for the essential component of GS_4 . Specifically, these kinds of fluctuations imply the uncoiling of the helix from both termini. This reflects the basic characteristics of the peptide chain. In addition, for the eigenvector corresponding to the second largest eigenvalue (0.05), the large elements are related to the fourth and fifth residues. Based on the trajectory, the eigenvector corresponds to the bending motion of the helix. The fluctuations of the peptide A_8 would be the combination of these two kinds of motions. As expected, large fluctuations are observed for both head and tail residues and the middle residues [as the red bar in Fig. 6(c)]. This observation is consistent with the results by Sancho *et al.* on the peptide composed of five alanines [49]. These all exemplify the value of the essential components and ensure the validity of the xPCA. In the subspaces spanned by the essential components, the kinetics are demonstrated by

TABLE II. Eigenvalues for GS_4 and A_8 , corresponding to the dihedral angle principal component analysis (dPCA).

	GS_4	A_8
1	1.29	1.76
2	1.21	0.86
3	1.07	0.49
4	1.01	0.37
5	0.96	0.30
6	0.89	0.24
7	0.88	0.19
8	0.85	0.18
9	0.77	0.17
10	0.77	0.12
11	0.68	0.12
12	0.63	0.09
13	0.27	0.09
14	0.26	0.08
15	0.23	0.07
16	0.21	0.07

projecting the conformations to the concerned components, as shown in Figs. 6(d) and 6(e). Clearly, similar fluctuating behaviors are observed, which further confirms the correctness of the description with the subspace spanned with essential components determined by the xPCA.

Similar to the above xPCA, the dPCA based on dihedral angles can be carried out. The eigenvalues are given in Table II. Based on the criterion related to the largest gap of the eigenvalues, the first 12 components (for GS_4) and the first component (for A_8) are selected to build up the subspaces, respectively. These selections can produce similar errors of covariance matrix as those by xPCA. The diffusions of peptides in these subspaces can be characterized accordingly. The descriptions with dihedral angles consider the constraints of bonds and angles consistently. It is believed to bring a better characterization for the landscapes, though the meanings of the coordinates are not direct.

C. Subdiffusive kinetics and analysis on hierarchical plateau

With the different thermodynamic and kinetic properties of the peptides, it is expected that these two systems exhibit different diffusive features in the conformational space. To quantitatively characterize the conformational diffusion of the peptides, the MSD is calculated based on the projections of the peptide conformations in the selected subspaces. Here, the MSD $\langle r^2(\tau) \rangle$ is defined as

$$\langle r^2(\tau) \rangle = \frac{1}{T - \tau} \sum_{t=0}^{T-\tau} [r(t + \tau) - r(t)]^2, \quad (12)$$

where T is the total length of trajectory and τ is the lag time. r is the position of the system in the projected subspace. Here, the coordinate Q_{ee} and the subspace based on two kinds of PCAs are used to calculate the MSD. Note that the coordinates of the subspace based on dPCA are different from those based on end-to-end distance or xPCA. Thus, the MSD curves are presented in different subfigures. Generally, the lag time dependence of MSD may take the power-law form $\langle r^2(\tau) \rangle \sim \tau^\alpha$,

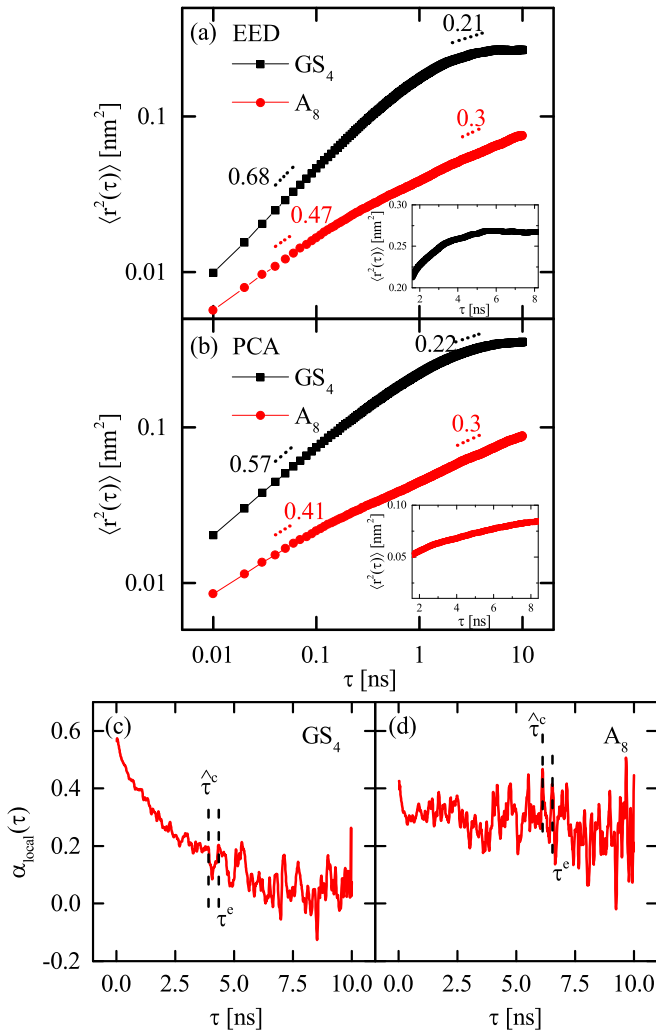


FIG. 7. This figure gives the mean square of displacement as a function of lag time in the subspace, (a) the coordinate of end-to-end distance (EED), and (b) spanned with essential components based on xPCA (the data are also shown in the insets with normal scale), and the exponents α_{local} related to local diffusion motions for the peptides (c) GS_4 and (d) A_8 (the onset time $\hat{\tau}^c$ and the escaping time τ^e for a certain valley are marked accordingly).

and the exponent α determines the characteristics of the diffusive kinetics, with $\alpha = 1$ corresponding to normal Brownian motion, and $\alpha \leq 1$ indicating a subdiffusion behavior. The corresponding relationships between MSD and the lag time τ are shown in Figs. 7(a), 7(b), and 8. The MSD curves in various subspaces are similar. For peptide GS_4 , the MSD curves first increase in a power-law manner and then slow down gradually with a small exponent, which demonstrates the process of equilibrium. The trend is independent of the selection of the coordinates to describe the motions. Quantitatively, the exponents in the short-time range are different. With the coordinate determined by xPCA, the exponent is 0.57 and, based on the dPCA, the exponent is 0.39, which are smaller than that (0.68) based on the coordinate Q_{ee} . This indicates that the coordinate based on PCA is more proper since it can describe the slower relaxation of the system. For the long-time range, similar exponents are achieved, 0.21 for

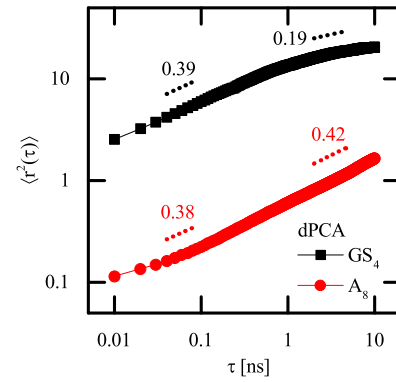


FIG. 8. This figure gives the mean square of displacement as a function of lag time in the subspace spanned with essential components based on dPCA.

Q_{ee} , 0.22 for the xPCA case, and 0.19 for the dPCA case. This exponent offers a depiction for the subdiffusion of the GS_4 . For peptide A_8 , the behaviors are similar. Yet the separation between short-time and long-time ranges is not as clear as that for GS_4 . This may be ascribed to the stronger internal friction in A_8 , since the peptide A_8 has stronger interactions compared to the flexible GS_4 . For the short-time range, the exponent is 0.41 for xPCA and 0.38 for dPCA, smaller than that (0.47) of Q_{ee} . The difference of the exponents is small but is still observable. For the long-time range, the exponent is the same for Q_{ee} and xPCA (both 0.30), and 0.42 for dPCA, which is different from that of GS_4 . Clearly, though the exponents are quantitatively different, these exponents are apparently smaller than 1, and the range of their values is consistent with those from other protein systems with various sizes and complexities (from 0.1 to 0.4) [15,53,54]. This reflects that our peptides are the rational models to investigate the subdiffusion in protein systems.

Based on these studies, peptides GS_4 and A_8 are different in their structures and energy landscapes but their diffusive kinetics are similar. Does the common diffusive feature mean that they share some common physical ingredient or is the similarity of the diffusion an accident by different properties? The analysis of the hierarchical landscape may help answer these questions. Based on the analysis on hierarchical plateaus [15] (as described in the Method section), the exponents α_f and α_t related to the size and depth of the multiscale valleys can be derived based on the MSD curves. In detail, the MSD curves for various peptides are not monotonic (as shown in the insets of Figs. 7(a) and 7(b)). Local fluctuations are evident. The derivatives α_{local} of the MSD in the subspace determined by xPCA for GS_4 and A_8 are exemplified in Figs. 7(c) and 7(d). The corresponding quantities, the onset time $\hat{\tau}^c$, and the escape time τ^e of a certain valley are marked in these figures. Based on these data for various tiers of valleys, the crossing time τ^c , the escape time τ^e , the size L , and the characteristic energy E can be determined based on the relations in the Method section. These characteristic quantities for each valley in the MSD curve based on the xPCA are shown in Table III and outline the features of the landscapes. It is observed that the characteristic sizes L_i for the peptide GS_4 are between 0.45 nm and 0.55 nm, and those for the peptide A_8 are between

TABLE III. Crossing time τ^c , escape time τ^e , characteristic size L , and characteristic energy E of each tier i via hierarchical plateau analysis for GS₄ and A₈.

Tier i	GS ₄				A ₈			
	τ_i^c (ps)	τ_i^e (ps)	L_i (nm)	E_i ($k_B T$)	τ_i^c (ps)	τ_i^e (ps)	L_i (nm)	E_i ($k_B T$)
1	70.0	2240	0.468	3.47	20.0	830	0.180	3.73
2	71.9	2900	0.488	3.70	21.2	1110	0.188	3.96
3	74.8	3460	0.500	3.83	27.3	1690	0.201	4.13
4	84.6	4320	0.516	3.93	31.7	2200	0.210	4.24
5	85.6	5320	0.529	4.13	41.8	3230	0.223	4.35
6	85.9	6050	0.536	4.26	44.9	3730	0.227	4.42
7	88.0	6500	0.540	4.30	45.0	4010	0.230	4.49
8	88.1	6730	0.542	4.34	47.0	4480	0.234	4.56
9	89.5	7090	0.545	4.37	51.7	5510	0.241	4.67
10	90.1	7710	0.549	4.45	57.3	6520	0.247	4.73
11					57.4	6950	0.249	4.80
12					59.1	7580	0.253	4.85

0.18 nm and 0.26 nm. That is, the valleys for the peptide GS₄ are significantly larger than those for peptide A₈. This phenomenon is attributed to the flexibility of glycine residue. Without the side chain, the glycine residue can access a larger conformational space compared to other kinds of residues. Thus, the valleys are broader when the glycine is involved. At the same time, the crossing times τ_i^c for the peptide GS₄ are between 70 ps and 90 ps, which are apparently larger than those for the peptide A₈ (which is between 20 ps and 60 ps). This is consistent with the observation that the peptide GS₄ has wider valleys, since longer times are necessary to cross the wider valleys. Different from the size and the crossing time, the energy and escape time for two peptides are in a similar range. For a same tier, the characteristic energy E_i for GS₄ is a little smaller than that for A₈. This is reasonable since the valleys for peptide A₈ are often related to multiple interactions and the interactions in peptide GS₄ are usually irrelevant. Meanwhile, the escape time τ_i^e for peptide GS₄ is relatively larger than that for peptide A₈. This contradicts the intuition based on the relation of characteristic energies and may be ascribed to the entropic contribution from the width of the valleys. Note that these geometric and energetic sizes and various times are physically reasonable. For example, the timescales and characteristic energies are quantitatively consistent with the scales for hydrogen bonds (the main interactions in these peptide systems). This supports the validity of the plateau analysis.

Similarly, the characteristic quantities for each valley in the MSD curve based on dPCA are also determined, as shown in Table IV. Here, the quantities L_i do not correspond to the spatial length. In addition, the energetic and temporal quantities have similar features as those from the xPCA. Though the different subspaces are introduced, the hierarchical feature of the landscape is still approved.

As expected, the power-law relations between these characteristic quantities are observed (as shown in Figs. 9 and 10). That is, there are two logarithmic relations [15] $\ln \tau_i^c = \ln \tau_0^c + (1/\alpha_f) \ln L_i$, and $E_i = E_0 + \gamma \ln(L_i/L_0)$, as Eqs. (7) and (8). These relations are valid for the results based on

TABLE IV. Crossing time τ^c , escape time τ^e , characteristic size L , and characteristic energy E of each tier i via hierarchical plateau analysis for GS₄ and A₈ in dihedral angle principal component analysis (dPCA).

Tier i	GS ₄				A ₈			
	τ_i^c (ps)	τ_i^e (ps)	L_i	E_i ($k_B T$)	τ_i^c (ps)	τ_i^e (ps)	L_i	E_i ($k_B T$)
1	40.0	1180	3.320	3.38	40.0	750	0.624	2.93
2	43.4	1570	3.458	3.59	59.7	1750	0.744	3.38
3	50.0	2610	3.704	3.95	64.2	2460	0.796	3.65
4	51.7	3050	3.775	4.08	81.6	3510	0.857	3.76
5	56.7	4010	3.899	4.26	84.7	4270	0.892	3.92
6	57.7	5760	4.042	4.60	93.6	5150	0.927	4.01
7	57.9	6910	4.110	4.78	105.8	6840	0.987	4.17
8	58.4	7530	4.142	4.86	110.7	8060	1.023	4.29

both xPCA and dPCA. Linear relations are evident with logarithmic coordinates. This confirms the self-similarity feature of the landscape and kinetics. Based on these relations, the exponents α_f , α_t , and α for both xPCA and dPCA can be determined, as shown in Table V. The exponents α_{fit} based on the fitting of the MSD are also given in Table V. The equality of the exponents α and α_{fit} supports the validity of the hierarchical analysis. In addition, the exponent α_f is the result of fitting the landscape with a fractal model, and the exponent α_t is the result of fitting the landscape with a model with only the contribution of traps. These two exponents are based on the assumptions of fractal landscape and of depth distribution of traps, and reflect two aspects of the systems. At the same time, the integrative effect of both factors gives exponent α , which describes the real subdiffusive behavior on such a landscape. The similarity between these exponents and that derived directly from the kinetics may help to judge which kind of assumption is suitable for the concerned systems. To determine which factor contributes essentially, we compare the exponent α with α_f (or α_t). If $\alpha \approx \alpha_f$, the fractal model is a reasonable model to describe the landscape, and if $\alpha \approx \alpha_t$, it is rational to conclude that the diffusion on the landscape is governed by the trap effect. Based on the data in Table V, the peptide GS₄ satisfies the relation $\alpha \approx \alpha_t$, indicating the essential contribution of the trap effect for the peptide GS₄. Differently, peptide A₈ has exponent α , which equals

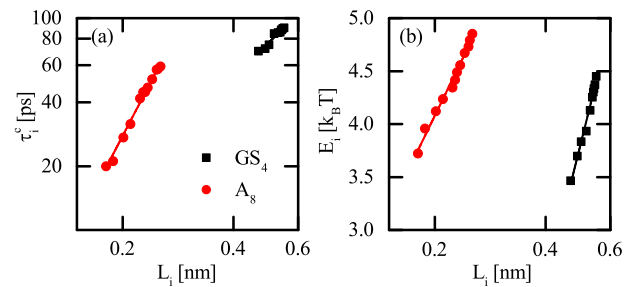


FIG. 9. The figure gives the relation (a) between crossing time τ^c and characteristic size L , and (b) between E_i and L_i for the peptides GS₄ (black solid square) and A₈ (red solid circle) in the subspace spanned with essential components based on xPCA.

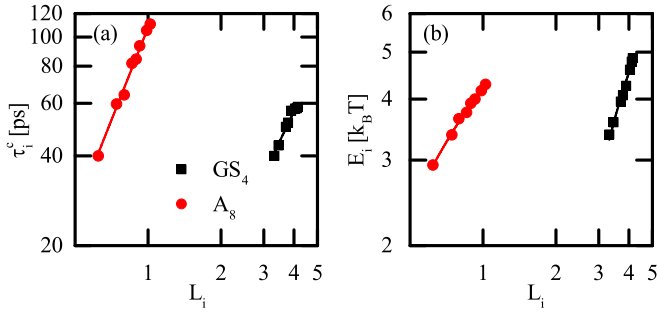


FIG. 10. The figure gives the relation (a) between crossing time τ^c and characteristic size L , and (b) between E_i and L_i for the peptides GS₄ (black solid square) and A₈ (red solid circle) in the subspace spanned with essential components based on dPCA.

approximately α_f , which implies that the fractal model is good for the concerned peptide. That is to say, the diffusive kinetics of the flexible peptide GS₄ is close to the behavior determined by the trap feature, while the behavior of the helical peptide A₈ is more likely to diffuse on the fractal landscape. These results demonstrate that two peptides behave differently based on this kind of judgment. It is worth pointing out that the relation between the exponents α and α_f (α_t) is valid for the results based on either xPCA or dPCA. This further supports the views about the mechanisms of the subdiffusive kinetics of peptides. Some data for larger protein systems are also listed. These proteins are similar to peptide A₈. This implies that A₈ can be used as the representation of structural proteins from their diffusive behavior.

At first sight, the difference of GS₄ and A₈ contradicts with the intuition: The diffusion of the flexible chain with weaker intrainteractions is governed by the energetic trap, and the landscape topology is important for the proteins with a remarkable energy funnel. In fact, a structural peptide should have much deeper energy traps compared to the nonstructural ones. This kind of energetic feature seems not to be consistent with the present conclusions about the mechanisms of the peptides. Actually, these diffusive behaviors reflect the key ingredients of the concerned systems. The subdiffusions of various peptides are induced by the intrachain interactions of peptides. The chains without any intrachain interactions would have a normal diffusion on a flat landscape. The emergence of the various interactions slows down the diffusions and produce the subdiffusion. For peptide GS₄, the flexibil-

TABLE V. The exponents α_f , α_t , and α based on hierarchical plateau analysis, and the exponent α_{fit} based on MSD curve for the peptides GS₄ and A₈, as well as for several proteins [15].

	α_t	α_f	α	α_{fit}
GS ₄ (xPCA)	0.28	0.58	0.25	0.22
A ₈ (xPCA)	0.49	0.30	0.31	0.30
GS ₄ (dPCA)	0.26	0.59	0.24	0.20
A ₈ (dPCA)	0.54	0.48	0.42	0.45
Protein G	0.58	0.26	0.31	0.30
Rab11a (based on PCA)	0.60	0.18	0.25	0.25
Fre-FAD	0.37	0.21	0.22	0.24

ity and the weak interactions make the overall shape of the landscape rather flat. The local interactions determine the key timescales of the diffusion. For peptide GS₄, the intrachain interactions (such as the hydrogen bonds) are often sparse and not correlated, which matches the model with uncorrelated traps. These interactions form random traps on the flat landscape. The diffusion depends on the distribution of single traps, similar to the typical trapping model. Differently, for peptide A₈, the intrachain interactions produce a global funnel governed by the specific native state on the landscape. The funnel topology indicates that there is a global architecture of various local energy minima, which indicates that the traps on the landscapes are highly correlated and connected specifically with a limited number of neighbors (not randomly with arbitrary traps). Under such a circumstance, the connectivity of traps or the topology of the landscape (rather than the depth of the traps) takes an essential role to determine the diffusion of the peptide. These traps are organized together and a network with hierarchical structures is formed. The kinetics is determined by the topology of landscape. The depth of the funnel controls the folding temperature and will not change the properties of the diffusion. The fractal topology is a reflection of the funnel, which is consistent with minimal frustrated principle. Note that the diffusion is related by the fluctuations on the energy landscape (rather than the energetic bias of the landscape). There are different features for flexible and structural peptides, which is not directly consistent with the intuition about the landscape. Based on these analysis, we can understand that there are different mechanisms of the similar subdiffusive kinetics in two peptides.

D. Network picture for the diffusive kinetics

Based on the analysis on hierarchical landscapes, it is likely to conclude that the subdiffusion of peptide A₈ is mainly attributed to the fractal topology of energy landscape, while the behavior of peptide GS₄ is controlled by random trapping. Beside the fitting with hierarchical model, a network picture of the distribution and transition in the conformational space for these two peptides is valuable to approach the problem about the subdiffusion mechanism. A₈ is expected to show a remarkable feature of fractal topology, while the transitions for GS₄ look more random. We give the conformational cluster transition network [10] (CCTN) in Figs. 11(a) and 11(b). CCTN is a network that characterizes the transitions between different states in dynamics. Each vertice in CCTN represents a state classified based on dihedral angles. In Figs. 11(a) and 11(b), vertices of larger sizes indicate that they have more neighbors. The edges in CCTN are undirected. Thicker edges indicate that the transition frequencies of them are higher. Similar networks have often been used to describe the structural dynamics of complex systems [55,56]. Since there are too many nodes in the whole 6 μ s trajectory, for clarity, we only show the vertices with the top 4% of neighbor degrees. The edges of discarded vertices are not considered, resulting in a few isolated vertices in the CCTN figure. The degree distribution $P(d)$ is defined as the probability density function of finding a vertex connected directly with d neighbors. Degree distributions $P(d)$ for both peptides are shown in Figs. 11(c) and 11(d). $P(d)$ is normalized and it can be well fitted by a

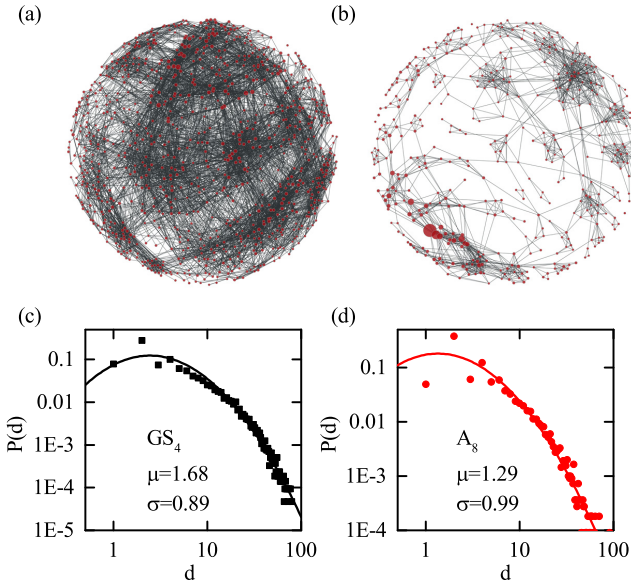


FIG. 11. Conformational cluster transition networks (CCTN) for (a) GS_4 with 859 vertices and 7691 edges, and (b) A_8 with 438 vertices and 1868 edges. The CCTN of GS_4 is a uniform network, while that of A_8 has some centralized clusters. The normalized degree distributions $P(d)$ of the CCTN for (c) GS_4 and (d) A_8 are shown. Lines are fitted by a log-normal distribution. The mean value μ and the standard deviation σ are fit parameters.

log-normal distribution as

$$P(d) = \frac{1}{\sqrt{2\pi}\sigma d} \exp\left\{-\frac{[\ln(d) - \mu]^2}{2\sigma^2}\right\}. \quad (13)$$

Based on our simulations, 859 vertices and 7691 edges are identified in the CCTN of GS_4 , while 438 vertices and 1868 edges are identified in the CCTN of A_8 , where some vertices with fewer neighbors are discarded. According to the log-normal fitting of the degree distribution, the mean value μ and the standard deviation σ for GS_4 are 1.68 and 0.89, while for A_8 are 1.29 and 0.99, respectively. GS_4 has an obviously larger μ than A_8 and their σ are very close. That means the CCTN of GS_4 has more linkers and the vertices of it have more neighbors to connect. The network of GS_4 looks more uniform compared to that of A_8 . At the same time, for GS_4 , the sizes of vertices are similar and there are no huge vertices. It means there are no dominant conformations in kinetics. This is consistent with GS_4 's trajectory and energy landscape in Fig. 4. GS_4 folds and unfolds very quickly in trajectories. Its Q_{ee} oscillates quickly, which means different conformations are easily connected by low-energy barriers. The flexible, unstructured property of GS_4 means there are not too many restrictions on the transitions between any two states. The energy topology of GS_4 is like the irregular network in Fig. 1(a). Larger μ of GS_4 indicates there are many vertices with lots of neighbors. These high-d vertices make transitions between different conformations simple and easy. Decentralized transitions weaken the fractal mechanism in the subdiffusion of GS_4 . This implies that trapping may play a more important role according to the hierarchical plateau analysis. On the other hand, the CCTN of A_8 is classified to some centralized

clusters. Instead of many high-d vertices in GS_4 , A_8 has a few big vertices and small vertices are connected by them. The energy topology of A_8 is like the fractal network in Fig. 1(b). The largest, centralized vertex in Fig. 1(b) corresponds to the complete helical conformation. Different unfolded states cannot be connected directly and they need a helical conformation as a hub between them. For example, an unfolded state with first HB broken cannot directly move to an unfolded state with third HB broken. It should fold to a complete helical structure at first and then unfold. A_8 usually takes tens or hundreds of nanoseconds to unfold and fold for one time. Unfolded states are separated by helical conformation in the trajectory of A_8 as shown in Fig. 4(b). This leads to some big clusters in the CCTN of A_8 . These larger vertices correspond to more complete helical structures. The peptide A_8 would like to evolve around these big vertices during most of the time, so the peptide remembers previous conformations for long time and slow relaxation is produced. The CCTN of A_8 is more like a scale-free network, which shows a strong property of fractal nature. This visualization gives us direct impressions about the organization of the conformations and the possible kinetics, which are qualitatively consistent with the above analysis on hierarchical landscape.

To further exemplify the importance of two kinds of ingredients in the diffusion, we carry out some controlled simulations based on the network picture. The subdiffusion driven by fractal topology may give different properties for GS_4 and A_8 , if we exclude the influence of trapping factor and keep fractal topology of energy landscape unchanged. Here we categorize the conformations of peptides into some finite states. The method of classification is the same as that of CCTN. Then a transition probability matrix $T_{ij}(\Delta t)$ is obtained according to the transition probability from state i to state j , after a lag time Δt . This matrix describes the transitions between any two states in dynamical process. To exclude the trapping factor, we average the nonzero matrix element (except diagonal matrix element) of each row in the transition probability matrix. It means that for a certain conformation, the next conformations to transit have the same probability and the same waiting time. The energy barriers for this conformation have the same escape time and the trapping factor is excluded. However, the reachable conformations are not changed, which means the fractal topology of energy landscape is not changed. With such a setup, the lifetimes of the traps are unchanged (since the lifetime is related to the total outgoing probability), that is, the thermodynamic feature is unchanged. Meanwhile, the local diffusions depend only on the local connectivity between traps (namely, the topology of the landscape). In our setup, some of the transition probabilities even increase compared to the natural case. This matrix is composed of fractal topology only. If the diffusion property obtained from the topology-only transition probability matrix is similar to that obtained from PCA, we can say that its source of subdiffusion is the fractal topology of the energy landscape. To confirm the different sources of subdiffusion for GS_4 and A_8 , we take the topology-only transition probability matrix of both peptides and apply them to a random starting state. After continuously applying the matrix, we get trajectories and the lag time dependence of MSD from it.

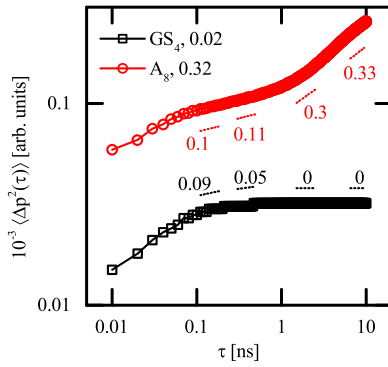


FIG. 12. The relation between mean square of displacement and the lag time calculated based on topology-only transition probability matrix for the peptides GS_4 (open black square) and A_8 (open red circle). Local exponents are shown near the curves.

The MSD of the controlled kinetics for these two peptides are given in Fig. 12, which is plotted with double logarithmic coordinates. The time that MSD of GS_4 reaches a stable state (about 500 ps) here is much shorter than that (about 5000 ps) in PCA (in Fig. 7). This demonstrates the importance of the trap distribution, and some deep traps contribute essentially to the slow diffusion (subdiffusion). Meanwhile, the MSD of A_8 increases with lag α , similar to its result of PCA. The subdiffusive exponent α obtained by topology-only transition probability matrix and PCA are compared in Table VI. We can see that the subdiffusive exponent α of GS_4 obtained by the topology-only transition probability matrix is 0.02, which is far less than that (0.22) obtained by PCA, while the α of A_8 obtained by topology-only transition probability matrix is 0.32, close to that (0.30) by PCA. Based on the above studies, the landscape of GS_4 has a weak fractal topology. After excluding the trapping factor, the GS_4 quickly reaches its thermodynamic equilibrium. Here the equilibrium time is about 500 ps, which is much smaller than the time for the peptide to go to equilibrium in realistic simulations. This implies the importance of the traps on the diffusive kinetics. More interestingly, after removing the variance of traps, the diffusion is largely different from the result based on direct simulations.

TABLE VI. Subdiffusive exponents α fitted by power law for GS_4 and A_8 , based on topology-only transition probability matrix and in the subspace determined with PCA.

	α of transition probability matrix	α of PCA
GS_4	0.02	0.22
A_8	0.32	0.30

This further demonstrates that the effect of traps is irreplaceable to describe the diffusion of the peptide GS_4 . For peptide A_8 , the situation is different. The long-time diffusion is almost the same as that in direct simulations. The removed trapping effect does not affect the long-time relaxation. This implies the importance of the landscape topology in the diffusion of peptide A_8 . These results further prove that the subdiffusion of A_8 is mainly attributed to the fractal topology of the energy landscape. On the other hand, the source of subdiffusion for GS_4 is not fractal topology. The comparison with the controlled kinetics clearly supports the different mechanisms for diffusions in various peptide systems.

IV. CONCLUSIONS

The subdiffusive dynamics of GS_4 and A_8 have been confirmed and, more importantly, their sources are different. The subdiffusion of GS_4 is mainly caused by the depth distribution of various traps, while the subdiffusion of A_8 is attributed to the fractal topology of the landscape. These provide some insight to understand the subdiffusive kinetics of globular proteins and IDPs. This may help us to understand how the subdiffusion emerges in systems with various interactions and the diverse kinetic behaviors of complex protein systems.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No. 11774157 and No. 11934008). The authors acknowledge HPCC of Nanjing University and of NLM for the computational support.

- [1] Y. He, S. Burov, R. Metzler, and E. Barkai, Random Time-Scale Invariant Diffusion and Transport Coefficients, *Phys. Rev. Lett.* **101**, 058101 (2008).
- [2] J. H. Jeon, V. Tejedor, S. Burov, E. Barkai, C. Selhuber-Unkel, K. Berg-Sorensen, L. Oddershede, and R. Metzler, In Vivo Anomalous Diffusion and Weak Ergodicity Breaking of Lipid Granules, *Phys. Rev. Lett.* **106**, 048103 (2011).
- [3] K. Agarwal, S. Gopalakrishnan, M. Knap, M. Muller, and E. Demler, Anomalous Diffusion and Griffiths Effects Near the Many-Body Localization Transition, *Phys. Rev. Lett.* **114**, 160401 (2015).
- [4] E. Guardado-Sanchez, A. Morningstar, B. M. Spar, P. T. Brown, D. A. Huse, and W. S. Bakr, Subdiffusion and Heat Transport in a Tilted Two-Dimensional Fermi-Hubbard System, *Phys. Rev. X* **10**, 011042 (2020).
- [5] I. E. T. Iben, D. Braunstein, W. Doster, H. Frauenfelder, M. K. Hong, J. B. Johnson, S. Luck, P. Ormos, A. Schulte, P. J. Steinbach, A. H. Xie, and R. D. Young, Glassy Behavior of a Protein, *Phys. Rev. Lett.* **62**, 1916 (1989).
- [6] H. Yang, G. Luo, P. Karnchanaphanurach, T.-M. Louie, I. Rech, S. Cova, L. Xun, and X. S. Xie, Protein conformational dynamics probed by single-molecule electron transfer, *Science* **302**, 262 (2003).
- [7] G. R. Kneller, Quasielastic neutron scattering and relaxation processes in proteins: Analytical and simulation-based models, *Phys. Chem. Chem. Phys.* **7**, 2641 (2005).
- [8] W. Min, G. Luo, B. J. Cherayil, S. C. Kou, and X. S. Xie, Observation of a Power-Law Memory Kernel for Fluctuations Within a Single Protein Molecule, *Phys. Rev. Lett.* **94**, 198302 (2005).

- [9] S. C. Kou and X. S. Xie, Generalized Langevin Equation with Fractional Gaussian Noise: Subdiffusion Within a Single Protein Molecule, *Phys. Rev. Lett.* **93**, 180603 (2004).
- [10] X. Hu, L. Hong, M. D. Smith, T. Neusius, X. Cheng, and J. C. Smith, The dynamics of single protein molecules is non-equilibrium and self-similar over thirteen decades in time, *Nat. Phys.* **12**, 171 (2015).
- [11] G. Luo, I. Andricioaei, X. S. Xie, and M. Karplus, Dynamic distance disorder in proteins is caused by trapping, *J. Phys. Chem. B* **110**, 9363 (2006).
- [12] T. Neusius, I. Daidone, I. M. Sokolov, and J. C. Smith, Subdiffusion in Peptides Originates from the Fractal-Like Structure of Configuration Space, *Phys. Rev. Lett.* **100**, 188103 (2008).
- [13] F. Rao and M. Karplus, Protein dynamics investigated by inherent structure analysis, *Proc. Natl. Acad. Sci. USA* **107**, 9152 (2010).
- [14] E. M. Bertin and J.-P. Bouchaud, Subdiffusion and localization in the one-dimensional trap model, *Phys. Rev. E* **67**, 026128 (2003).
- [15] Y. Meroz, V. Ovchinnikov, and M. Karplus, Coexisting origins of subdiffusion in internal dynamics of proteins, *Phys. Rev. E* **95**, 062403 (2017).
- [16] G. R. Kneller and K. Hinsen, Fractional Brownian dynamics in proteins, *J. Chem. Phys.* **121**, 10278 (2004).
- [17] Y. Matsunaga, C. B. Li, and T. Komatsuzaki, Anomalous Diffusion in Folding Dynamics of Minimalist Protein Landscape, *Phys. Rev. Lett.* **99**, 238103 (2007).
- [18] H. Frauenfelder, P. W. Fenimore, and B. H. McMahon, Hydration, slaving and protein function, *Biophys. Chem.* **98**, 35 (2002).
- [19] P. W. Fenimore, H. Frauenfelder, B. H. McMahon, and F. G. Parak, Slaving: Solvent fluctuations dominate protein dynamics and functions, *Proc. Natl. Acad. Sci. USA* **99**, 16047 (2002).
- [20] E. T. J. Nibbering, H. Fidder, and E. Pines, Ultrafast chemistry: Using time-resolved vibrational spectroscopy for interrogation of structural dynamics, *Annu. Rev. Phys. Chem.* **56**, 337 (2005).
- [21] J. T. M. Kennis and M.-L. Groot, Ultrafast spectroscopy of biological photoreceptors, *Curr. Opin. Struct. Biol.* **17**, 623 (2007).
- [22] A. K. Sangha and T. Keyes, Proteins fold by subdiffusion of the order parameter, *J. Phys. Chem. B* **113**, 15886 (2009).
- [23] B. R. Gelin and M. Karplus, Role of structural flexibility in conformational calculations: Application to acetylcholine and beta-methylacetylcholine, *J. Am. Chem. Soc.* **97**, 6996 (1975).
- [24] C. Monthus and J.-P. Bouchaud, Models of traps and glass phenomenology, *J. Phys. A* **29**, 3847 (1996).
- [25] R. Metzler and J. Klafter, The random walk's guide to anomalous diffusion: A fractional dynamics approach, *Phys. Rep.* **339**, 1 (2000).
- [26] E. Barkai and Y. C. Cheng, Aging continuous time random walks, *J. Chem. Phys.* **118**, 6167 (2003).
- [27] J. W. Haus and K. W. Kehr, Diffusion in regular and disordered lattices, *Phys. Rep.* **150**, 263 (1987).
- [28] I. M. Sokolov, A. Blumen, and J. Klafter, Linear response in complex systems: CTRW and the fractional Fokker-Planck equations, *Physica (Amsterdam)* **302**, 268 (2001).
- [29] I. M. Sokolov, A. Blumen, and J. Klafter, Dynamics of annealed systems under external fields: CTRW and the fractional Fokker-Planck equations, *Europhys. Lett.* **56**, 175 (2001).
- [30] Y. Meroz and I. M. Sokolov, A toolbox for determining subdiffusive mechanisms, *Phys. Rep.* **573**, 1 (2015).
- [31] R. Metzler, J.-H. Jeon, A. G. Cherstvy, and E. Barkai, Anomalous diffusion models and their properties: Non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking, *Phys. Chem. Chem. Phys.* **16**, 24128 (2014).
- [32] J. Tang and S. H. Lin, Distance versus energy fluctuations and electron transfer in single protein molecules, *Phys. Rev. E* **73**, 061108 (2006).
- [33] R. Granek and J. Klafter, Fractons in Proteins: Can They Lead to Anomalous Decaying Time Autocorrelations? *Phys. Rev. Lett.* **95**, 098106 (2005).
- [34] J. Chen and R. W. Kriwacki, Intrinsically disordered proteins: Structure, function and therapeutics, *J. Mol. Biol.* **430**, 2275 (2018).
- [35] I. L. Morgan, R. Avinery, G. Rahamim, R. Beck, and O. A. Saleh, Glassy Dynamics and Memory Effects in an Intrinsically Disordered Protein Construct, *Phys. Rev. Lett.* **125**, 058001 (2020).
- [36] C. B. Anfinsen, Principles that govern the folding of protein chains, *Science* **181**, 223 (1973).
- [37] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, *J. Chem. Theory. Comput.* **4**, 435 (2008).
- [38] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, *J. Comput. Chem.* **24**, 1999 (2003).
- [39] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, The missing term in effective pair potentials, *J. Phys. Chem.* **91**, 6269 (1987).
- [40] G. Bussi, D. Donadio, and M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.* **126**, 014101 (2007).
- [41] M. Parrinello and A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, *J. Appl. Phys.* **52**, 7182 (1981).
- [42] T. Haliloglu, I. Bahar, and B. Erman, Gaussian Dynamics of Folded Proteins, *Phys. Rev. Lett.* **79**, 3090 (1997).
- [43] I. Bahar, A. R. Atilgan, M. C. Demirel, and B. Erman, Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability, *Phys. Rev. Lett.* **80**, 2733 (1998).
- [44] A. Altis, P. H. Nguyen, R. Hegger, and G. Stock, Dihedral angle principal component analysis of molecular dynamics simulations, *J. Chem. Phys.* **126**, 244111 (2007).
- [45] R. I. Cukier, Ferreting out correlations from trajectory data, *J. Chem. Phys.* **135**, 225103 (2011).
- [46] R. I. Dima and D. Thirumalai, Asymmetry in the shapes of folded and denatured states of proteins, *J. Phys. Chem. B* **108**, 6564 (2004).
- [47] A. Vitalis, X. Wang, and R. V. Pappu, Quantitative characterization of intrinsic disorder in polyglutamine: Insights from analysis based on polymer theories, *Biophys. J.* **93**, 1923 (2007).
- [48] H. T. Tran and R. V. Pappu, Toward an accurate theoretical framework for describing ensembles for proteins under strongly denaturing conditions, *Biophys. J.* **91**, 1868 (2006).
- [49] D. de Sancho, A. Sirur, and R. B. Best, Molecular origins of internal friction effects on protein-folding rates, *Nat. Commun.* **5**, 4307 (2014).

- [50] A. Amadei, A. B. Linssen, and H. J. Berendsen, Essential dynamics of proteins, *Proteins* **17**, 412 (1993).
- [51] A. L. Tournier and J. C. Smith, Principal Components of the Protein Dynamical Transition, *Phys. Rev. Lett.* **91**, 208106 (2003).
- [52] M. Post, S. Wolf, and G. Stock, Principal component analysis of nonequilibrium molecular dynamics simulations, *J. Chem. Phys.* **150**, 204110 (2019).
- [53] P. Senet, G. G. Maisuradze, C. Foulie, P. Delarue, and H. A. Scheraga, How main-chains of proteins explore the free-energy landscape in native states, *Proc. Natl. Acad. Sci. USA* **105**, 19708 (2008).
- [54] L. Milanesi, J. P. Waltho, C. A. Hunter, D. J. Shaw, G. S. Beddard, G. D. Reid, S. Dev, and M. Volk, Measurement of energy landscape roughness of folded and unfolded proteins, *Proc. Natl. Acad. Sci. USA* **109**, 19563 (2012).
- [55] F. Noé, I. Horenko, C. Schütte, and J. C. Smith, Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states, *J. Chem. Phys.* **126**, 155102 (2007).
- [56] T. Neusius, I. Daidone, I. M. Sokolov, and J. C. Smith, Configurational subdiffusion of peptides: A network study, *Phys. Rev. E* **83**, 021902 (2011).