Configurational subdiffusion of peptides: A network study

Thomas Neusius,[1] Isabella Daidone,[2] Igor M. Sokolov,[3] and Jeremy C. Smith[1,4]
[1]*Computational Molecular Biophysics, Universität Heidelberg, Im Neuenheimer Feld 368, D-69120 Heidelberg, Germany*
[2]*Dipartimento di Chimica, Università dell'Aquila, Via Vetoio (Coppito 1), I-67010 L'Aquila, Italy*
[3]*Institut für Physik, Humboldt-Universität zu Berlin, Newtonstraße 15, D-12489 Berlin, Germany*
[4]*Center for Molecular Biophysics, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee, 37831-6164, USA*
(Received 29 April 2010; revised manuscript received 10 December 2010; published 2 February 2011)

Molecular dynamics (MD) simulation of linear peptides reveals configurational subdiffusion at equilibrium extending from 10^{-12} to 10^{-8} s. Rouse chain and continuous-time random walk models of the subdiffusion are critically discussed. Network approaches to analyzing MD simulations are shown to reproduce the time dependence of the subdiffusive mean squared displacement, which is found to arise from the fractal-like geometry of the accessible volume in the configuration space. Convergence properties of the simulation pertaining to the subdiffusive dynamics are characterized and the effect on the subdiffusive properties of representing the solvent explicitly or implicitly is compared. Non-Markovianity and other factors limiting the range of applicability of the network models are examined.

DOI: 10.1103/PhysRevE.83.021902 PACS number(s): 87.15.H−, 87.14.ef, 87.15.Ya, 87.15.ap

I. INTRODUCTION

The internal fluctuations of proteins and peptides are essential for biological function and occur over a wide range of time scales [1]. One of the intriguing features of protein internal dynamics is the presence of subdiffusive fluctuations; that is, the mean squared displacements (MSD) of the atomic random walks in configuration space exhibit a sublinear time dependence, and this has been observed over many decades of time. On the ps-ns time scales, subdiffusion has been observed in molecular dynamics (MD) simulations [2–4] and in neutron spin-echo experiments [5]. Further, single molecule fluorescence resonance energy transfer spectroscopy has allowed projections of the MSD on internal distance fluctuations to be determined, revealing subdiffusive dynamics occurring also on the time scales of milliseconds to seconds [6–9].

In previous work [4] the question was addressed of whether anomalous diffusion at equilibrium observed in proteins pertains only to large and complex molecules showing at least secondary structure, or is a generic property of a larger class of heteropolymers, including, for example, peptides, the building blocks of proteins. It was found that the diffusional anomalies in the behavior of the end-to-end distance and in single principal components of the motion are observed even for relatively small and flexible peptides lacking secondary structure.

The main questions addressed in the present work are

1. What is the source of the anomalous diffusion? Is it mostly determined by the intramolecular interactions, or by interactions with solvent molecules (e.g., hydrodynamic interactions)?

2. Which theoretical model is the most promising for the description of the anomalous internal diffusion?

As to the first question, we find here that the phenomenon of subdiffusion—although not fully insensitive to the model of solvent chosen—is satisfactorily reproduced even using an implicit solvent model and Langevin dynamics. Therefore, intramolecular interactions determine the subdiffusivity of the internal dynamics.

To answer the second question, we first analyze common models of anomalous diffusion. Two popular classes of models used to describe the internal anomalous diffusion of proteins are generalized Gaussian models (including the Rouse model for a linear polymer chain) and random energy models (REM), both of which have been used to describe nonequilibrium properties of proteins, such as are observed in the folding process [11–13]. We find here highly nonlinear interactions, inconsistent with Gaussian models, and equilibrium subdiffusive dynamics in conflict with REM. The Gaussian models and REM are ruled out and instead, the diffusion process is confined to a fractal cluster representing the accessible volume of the configuration space.

Diffusion on a fractal substrate, although popular in other fields, has not often been applied to polymer dynamics. While formally taking place in the extremely high-dimensional configurational space comprising the spatial coordinates of all atoms, the equilibrium fluctuations of peptides are essentially confined to a low-dimensional subset. Crucial for the emergence of subdiffusion is the fractal geometry of this dynamically relevant subset. The diffusion on a fractal set (as exemplified, e.g., by the diffusion on a percolation cluster) is known to exhibit an MSD sublinear in the lag time [14].

Conceptually, the energy landscape may be thought of as a deep valley with a flat bottom surrounded by high mountain ridges, inaccessible to the equilibrium dynamics. The mountains confine the dynamics to the valley at equilibrium, and are explored only in nonequilibrium processes, such as folding. The ruggedness of the folding energy landscape justifies the REM approach in the case of folding. In contrast, the floor of the valley is flat on the $k_B\mathcal{T}$ scale. The entirety of energetically accessible valleys forms a fractal volume, which is modeled here as a network. The edges of the network represent the valleys and the nodes correspond to regions of the configurational space where multiple valleys merge. This network picture of the energy landscape is the main conceptual result of the present investigation.

TABLE I. Simulation lengths in units of microseconds (μs).

n	2	3	5	7
$(GS)_nW$	0.8	1.0	1.9	2.5

II. METHODS

Here we analyze the dynamics of peptide chains of the $(GS)_nW$ type (with G = glycine, S = serine, W = tryptophan, and $n = 2,3,5,7$); these peptide systems are small enough that their configuration space is well sampled while still exhibiting a subdiffusive MSD over more than three orders of magnitude in time.

A. Simulation details

The $(GS)_nW$ molecules are polypeptide chains formed of n (here $n = 2,3,5,7$) repeated GS segments (G = glycine, S = serine) with a terminal tryptophan (W) residue. The $(GS)_nW$ peptides do not fold into specific secondary structures [15].

The simulations of the $(GS)_nW$ peptides were performed with the simulation engine GROMACS [16] and the GROMOS96 force field [17]. The canonical ($NV\mathcal{T}$) ensemble was used, with the number of particles (N), volume (V), and temperature ($\mathcal{T}$) held constant. The constant temperature, $\mathcal{T} = 293$ K, was established by the isokinetic thermostat [18]. The peptide was placed in a rhombic dodecahedral box. The boundary of the box was at least $\approx$1.0 nm from all the atoms of the peptide. The bond lengths were kept fixed with the LINCS algorithm [19] and a time step of 2 fs for numerical integration of the equations of motion was used. The total simulation times are given in Table I. In what follows, we concentrate particularly on the $(GS)_5W$ peptide. In the explicit solvent simulations the solvent was modeled by the explicit simple point charge water model [20]. The liquid density was 55.32 mol/l ($\approx$1 g/cm^3). Periodic boundary conditions were used and long-range interactions treated by the particle mesh Ewald method [21] with a grid spacing of 0.12 nm and fourth-order B-spline interpolation. The real space cut-off distance was set to 0.9 nm. In the case of $(GS)_5W$ the simulation box had a volume of 34.81 nm^3 and contained in total 3569 atoms; the number of water molecules was 1543.

To examine the role of the solvent dynamics, an additional simulation was performed, apart from the simulations in explicit aqueous solution: the $(GS)_5W$ peptide with implicit solvent using the generalized Born/surface area (GB/SA) approach [22,23] together with an effective Langevin process imitating the solvent dynamics [16]. The Born radii were calculated using the fast asymptotic pairwise summation of Ref. [24]. The relevant parameters used, together with the GROMOS96 force field, can be found in Ref. [25]. To increase the efficiency of the surface area calculation, a mimic based on the Born radii was used [26]. Further details of a similar simulation study of a β hairpin can be found in Ref. [27].

B. Analysis tools

Multidimensional data analysis provides schemes that, depending on the quantities of interest, allow the amount of data to be efficiently reduced. One such scheme common in MD simulation is principal component analysis (PCA) [28–31].

The time-averaged value of the ith component of the mass-weighted position vector $\boldsymbol{x}(t)$ is denoted by $\bar{x}_i$. The PCA starts with the covariance matrix $\mathbf{C}$, whose components are given as

$$C_{ij} = \overline{[x_i(\tau) - \bar{x}_i]\,[x_j(\tau) - \bar{x}_j]}, \tag{1}$$

where the time average is performed over the variable τ.

The covariance matrix is symmetric by construction and is diagonalized by the orthogonal matrix $\mathbf{W}$, i.e.,

$$\mathbf{W}^T\mathbf{C}\mathbf{W} = \mathbf{\Lambda}, \tag{2}$$

where $\mathbf{\Lambda}$ has the components $\Lambda_{ij} = \delta_{ij}\lambda_i$. The matrix $\mathbf{W}$ represents a transition from one orthonormal basis set to another orthonormal basis set. The eigenvalues of the covariance matrix, λ_i, are the fluctuations along the eigenvectors. The eigenvalues are usually sorted in descending order, λ_1 being the largest eigenvalue. The coordinates $\boldsymbol{q} = \mathbf{W}^T\boldsymbol{x}$, in which the covariance matrix is diagonal, are termed the principal components (PC). The components of $\boldsymbol{q}(t) = \mathbf{W}^T\boldsymbol{x}(t)$ are referred to as PC modes. By construction, the PCs are uncorrelated coordinates (but they are, in general, not statistically independent). As in the case of normal modes, the low PCs are global coordinates involving many particles, while the high PCs are more localized. From the trajectory PC mode $q_i(t)$ the configuration space density is obtained as a normalized histogram, $\rho(q_i)$. Assuming equilibrium, the corresponding free energy—also referred to as potential of mean force (PMF)—reads $F(q_i) = -k_B\mathcal{T}\,\ln\,\rho(q_i)$.

III. RESULTS

A. Thermodynamics and kinetics

A detailed analysis of the PCs demonstrates that the free energy profiles are strongly anharmonic for the lowest PCs, and that the lowest PCs are strongly delocalized and involve a large number of atoms (see a detailed discussion in the supplemental material [10]).

We now examine kinetic quantities in the simulations. The MSD along PC i for the discrete data $q_\kappa = q_{i,\kappa} = q_i(\kappa\,\Delta t)$—we skip the q subscripts i for simplicity where appropriate—with Δt being the resolution of the trajectory, is obtained as

$$\langle \Delta x^2(t)\rangle = \langle \Delta x^2(\kappa\,\Delta t)\rangle \tag{3}$$

$$= \frac{1}{K-\kappa}\sum_{k=1}^{K-\kappa}(q_{k+\kappa} - q_k)^2, \tag{4}$$

where K is the total number of frames of the trajectory, the length of the simulation being $T = K\,\Delta t$. The time-averaging procedure in Eq. (3) is valid for $t \ll T$. For t values too close to T the time average in Eq. (3) is statistically not significant.

Figure 1 illustrates the time-averaged MSD of the individual PC modes of the $(GS)_5W$ peptide simulations with the explicit [Fig. 1(a)] and implicit solvent model [Fig. 1(b)]. For the explicit water simulation, PCs 1, 2, and 3 exhibit a power-law behavior extending from 1 ps up to 10 ns. The exponent of the power law is $\approx$0.5 for all PCs. PCs 10, 20, and 30 also exhibit power-law behavior for short time scales but reach saturation more quickly. The implicit solvent simulation reproduces

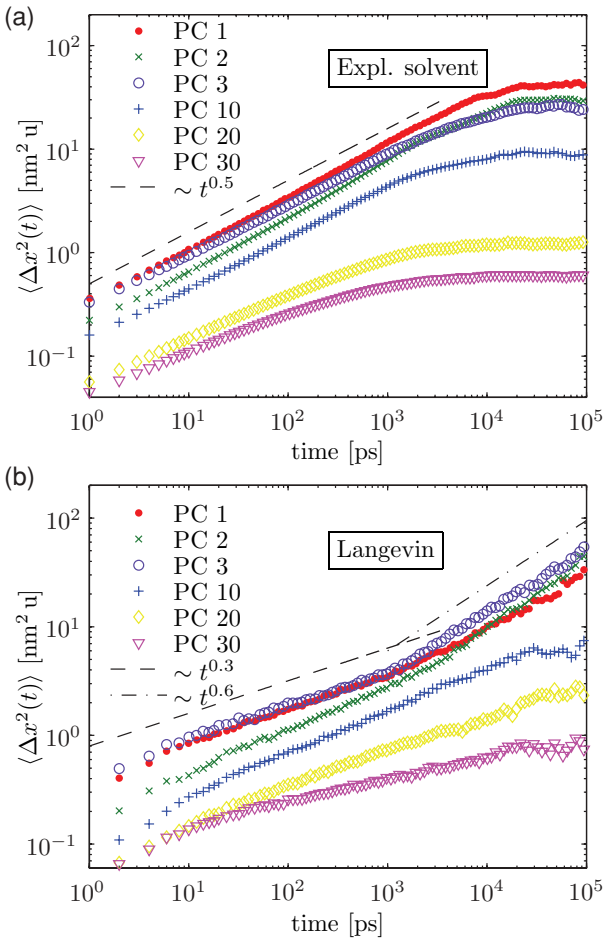

FIG. 1. (Color online) The time-averaged MSD of the $(GS)_5W$ peptide for various PCs.

the subdiffusive MSD. However, the exponent $\alpha \approx 0.3$ is considerably lower than in the explicit solvent simulation and for lag times longer than 1 ns the lowest PCs exhibit a second subdiffusive regime with an exponent $\alpha \approx 0.6$.

For lower lag times, the Langevin simulation exhibits an MSD exponent α lower than the one seen in the explicit solvent simulation, i.e., in the Langevin system subdiffusion is enhanced. Therefore, the solvent dynamics are neither the origin nor do they contribute substantially to the subdiffusivity seen in the subspace of the molecule coordinates.

B. Modeling subdiffusion

The presence of subdiffusive kinetics in the internal coordinates of biomolecules raises the question as to the underlying mechanism. Several models can account for subdiffusion, some of which have been discussed as candidates for biomolecular fluctuations, and are reviewed in the following.

1. The Rouse chain

As peptides are chains of amino-acid residues, chain models, including the Rouse model, can be useful in describing them [11,12,32,33]. The Rouse chain is a common model in polymer physics [34,35].

A Rouse chain consists of N beads of mass m coupled together by Hookean springs with typical frequency $\tilde{\omega}$. Coupling to a heat bath transfers an average energy to each degree of freedom, according to the equipartition theorem, and leads to a frictional force characterized by the friction constant ζ. It can be shown that, given a sufficiently long chain, the autocorrelation of the distance between two beads not too close together along the chain decays as a power law with exponent 1/2 [36]. However, this power law is bounded by the intrinsic time scale of the chain, the so-called Rouse time

$$\tau_R = \frac{\zeta (N-1)^2}{\pi^2 m \tilde{\omega}^2}. \tag{5}$$

Beyond τ_R the autocorrelations decay exponentially.

In order to account for the secondary structure of proteins the linear Rouse chain model has been generalized to bead-spring clusters [12,36], thereby skipping the condition that the springs form a linear chain and introducing cross-links. The geometry of the bead-spring cluster is described as a fractal in terms of its fractal exponents [36,37]. The vibrational eigenmodes, i.e., the normal modes, were calculated and found to be strongly localized [36]. Distance autocorrelation functions decay as power laws with various exponents up to a time scale that is approximately given by the inverse friction ζ^{-1} or, if smaller, by the inverse minimal eigenfrequency of the harmonic cluster ω_{min}^{-1} [36].

In the case of the biomolecules presented here, typical normal mode fluctuations are in the range of $\langle \Delta x^2 \rangle \approx 1$ nm^2. The corresponding frequencies are then $\tilde{\omega}^2 \approx 3k_B\mathcal{T}/m\langle \Delta x^2 \rangle$, which are in the range of $\tilde{\omega} \approx 1$ ps^{-1}. A typical friction value for a heavy atom at the surface of a biomolecule is $\gamma \approx 50$ ps^{-1} [38]. Equation (5) allows the Rouse time of the $(GS)_5W$ peptide ($N = 11$, the number of residues) to be estimated as $\tau_R \approx 40$ ps. In contrast, the subdiffusion found in the present simulation extends to 10 ns and beyond. Therefore, the Rouse chain model—irrespective of potential subdiffusive dynamics on short time scales—cannot provide a full understanding of the subdiffusive dynamics as found in the MD simulations presented above. Similar objections apply to subdiffusion found in single molecule fluorescence quenching experiments [8,9,11,39,40].

In accordance with the above findings, computer simulations revealed recently that linear chains with anharmonic single-minimum pair potentials do not exhibit subdiffusive PCs, whereas chains with multi-minimum pair potentials do [41].

2. Glassy properties and continuous time random walk

Non-exponential relaxation patterns, as found in the dynamics of proteins and peptides, are a common property of glassy materials. There are also other characteristics shared by glasses and proteins, such as non-Arrhenius dynamical temperature dependence, i.e., the logarithm of the reaction rate is not inversely proportional to the temperature [42,43] and the enhancement of fluctuations above the glass transition [44]. This similarity of proteins to glasses is a consequence of the large number of nearly isoenergetic potential minima (conformational substates) that both proteins and glasses can assume [42,45,46]. The energy landscape of glasses,

and the glassy dynamics that follows from it, have been successfully described as hopping between energy traps on a fully connected lattice, where the distribution of effective trap depths essentially determines the time evolution [47–50]. The essential properties of these trap models are equivalently reproduced by the continuous-time random walk (CTRW) model [51,52].

It has been demonstrated that the time-averaged MSD of a free, unbounded CTRW does not exhibit a sublinear lag-time behavior [4]. Although the time-averaged MSD of a confined CTRW may be found to exhibit a subdiffusive lag-time dependence, the time scales on which subdiffusivity is found are inconsistent with the simulation results [53] (see [10] for a more detailed discussion). Therefore, the distribution of traps or metastable states is incapable of accounting for the MSD found in the $(GS)_5W$ simulations.

C. Network representation of the configuration space

For peptides, a trap distribution cannot describe the properties of the random walk on the energy landscape, as the failure of CTRW to reproduce the diffusion process in the configuration space indicates. Therefore, geometrical aspects of the energy landscape must be considered. This can be achieved by a network representation of the configuration space, also known as a transition network [54].

To establish a transition network, the first step is to discretize the configurational volume explored in the course of the trajectory, i.e., to group the K discrete configurations of the trajectory into a smaller number of "states." There are multiple ways of doing this. Here, a number N_s of configurations was randomly chosen from the K configurations, $\{\boldsymbol{r}_1,...,\boldsymbol{r}_K\}$ visited in the course of the trajectory, $\boldsymbol{r}_i$ being the vector of coordinates of frame i. Then, each of the K configurations was assigned to that state, $\boldsymbol{r}_s \in \{\boldsymbol{r}_1,...,\boldsymbol{r}_{N_s}\}$, with the closest Euclidean distance in the configuration space. In the present context, alternative discretization schemes (e.g., with at least a minimum distance between every pair of configurations or a k-means approach) were found to have a very low impact on the results.

The discretization allows the time evolution of the system, as represented by the trajectory, to be mapped onto a series of integers $h_1,h_2,...,h_K$ with $h_i \in \{1,2,...,N_s\}$. The count matrix $\mathbf{Z}(\vartheta)$ is defined by its components

$$z_{ij}(\vartheta) = \sum_{k=1}^{K-\vartheta} \delta_{h_k,j}\delta_{h_{k+\vartheta},i}, \tag{6}$$

where δ_{ij} is the Kronecker delta. The system is found $z_{ij}(\vartheta)$ times in state i after being in state j exactly ϑ time steps earlier. At equilibrium, the matrix $\mathbf{Z}$ must obey the condition of detailed balance, i.e., $z_{ij}(\vartheta) = z_{ji}(\vartheta)$ [55].

The relative probability of being in state j exactly ϑ time steps after being in state i is denoted by s_{ij}. The matrix $\mathbf{S}$ is the so-called transition matrix of the system. Its components are obtained from the approximation

$$s_{ij}(\vartheta) \approx \tilde{s}_{ij}(\vartheta) = \frac{z_{ij}(\vartheta)}{\sum_{k=1}^{N_s} z_{kj}(\vartheta)}. \tag{7}$$

The approximate transition matrix $\tilde{\mathbf{S}}$ can be used as the input of a Markov model: to generate a pseudotrajectory starting at an arbitrary state $l_1 \in \{1,2,...,N_s\}$, the next state is drawn such that the probability of being in i is given by $s_{il_1}(\vartheta)$. To obtain a pseudotrajectory of length T, this procedure is repeated T/ϑ times. The time resolution of such a trajectory is bounded by ϑ. The discrete trajectory in state space can be mapped onto the configuration space, as each state can be identified with a configuration. That is, the network can be embedded into the configuration space. Here, state i is mapped onto the coordinates $\boldsymbol{r}_i$ used in the discretization scheme.

By definition, the above algorithm takes into account only the present state when determining the next transition, and is therefore inherently Markovian. Even in the statistical sense, the pseudotrajectories are expected to deviate from the random walk in the configuration space obtained from the full MD simulation.

There are four possible primary sources of error:

1. The dynamics may be non-Markovian on the relevant time scale represented by the lag time ϑ.

2. The discretization may not be appropriate for the dynamics present: stable configurations may not be resolved and mapped on various, dynamically different states.

3. The statistics may be insufficient and $\tilde{\mathbf{S}}$ a poor approximation.

4. There may be limited spatial resolution, i.e., too few discrete states.

The first point is due to the projection onto low-dimensional subspaces, and can be alleviated by taking more dimensions into account. The second point emerges from the discretization scheme, which in the present case was not designed to find dynamically stable configurations. Approaches based on dynamical quantities usually need very accurate statistics and lead therefore to a relatively low number of states [56–59]. However, the trade-off between statistical significance (point 3) and spatial resolution (point 4) renders most of these schemes inappropriate for the modeling of configurational subdiffusion.

Transition networks of the $(GS)_nW$ peptides were calculated from the MD trajectories based on a discretization scheme with 10 000 states, unless stated otherwise. The simulation lengths of the individual MD trajectories are given in Table I. As the network analysis in the full configuration space is numerically cumbersome, the analysis was performed on the subspace spanned by the first ten PCs. In each $(GS)_nW$ peptide, these ten strongly delocalized modes account for more than 75% of the overall fluctuations. However, considering Zwanzig's argument that projection may lead to correlations in time [60], the appearance of memory effects possibly pertaining to the time scale of the MD resolution, 1 ps, cannot be ruled out.

The fractal dimension of the transition network $\tilde{\mathbf{S}}(\vartheta)$ was obtained using volume-length scaling [14], i.e., in a sphere of radius R typically N edges are enclosed with

$$N \sim R^{d_f}. \tag{8}$$

The function $N(R)$ was calculated counting the edges inside spheres of various radii R for 1000 central nodes randomly

TABLE II. Fractal dimension d_f of the networks representing the configuration space $\tilde{\mathbf{S}}(\vartheta)$. None of the networks fill the full ten-dimensional embedding space, and all networks exhibit fractal behavior.

	d_f			
ϑ (ps)	1	10	100	1000
$(GS)_2W$	6.1	6.8	7.4	7.9
$(GS)_3W$	6.2	6.8	7.6	8.7
$(GS)_5W$	6.4	6.6	7.0	7.7
$(GS)_7W$	6.1	6.3	6.4	7.3

TABLE III. Random walk on the networks representing the configuration space $\tilde{\mathbf{S}}(\vartheta)$. The random walk dimension d_w was calculated from the MSD using $d_w = 2/\alpha$ [14]. The MSD exponent α used throughout the present paper to quantify the subdiffusivity, is given in the right panel for comparison.

	d_w				α			
ϑ (ps)	1	10	100	1000	1	10	100	1000
$(GS)_2W$	2.6	3.0	4.3	11.3	0.8	0.7	0.5	0.2
$(GS)_3W$	2.4	2.7	3.8	9.6	0.8	0.7	0.5	0.2
$(GS)_5W$	2.4	2.7	3.2	4.4	0.8	0.7	0.6	0.5
$(GS)_7W$	2.3	2.5	3.0	4.4	0.9	0.8	0.7	0.5

chosen, followed by an average over these 1000 individual $N(R)$. Equation (8) then yielded the fractal dimension d_f via a least-squares power-law fit to the data (Table II).

Networks corresponding to various ϑ were then used as Markov models. The random walk dimension d_w was obtained from the MSD exponent α as $d_w = 2/\alpha$, and both are listed in Table III. The network representation enables us to study how the energy landscape brings about the subdiffusive dynamics of the molecule.

The network diffusion reproduces the effect of subdiffusivity, insofar as the MSD exponent α is in the subdiffusive regime (Table III). However, the networks for $\vartheta = 1$ and 10 ps exhibit MSDs clearly larger than for the original trajectory, given as the red continuous lines in Fig. 2. The reason for the difference may be the presence of memory effects on the shorter time scales, which may possibly arise from the projection of the first ten PCs onto the ten-dimensional subspace. By construction, these memory effects cannot be reproduced by the Markovian network dynamics, and therefore, the subdiffusivity is underestimated, such that the network MSD exceeds the MSD found in the MD simulation data. Also, the limited spatial resolution of a network with 10 000 vertices affects the MSD, introducing additional noise, in particular, on the shorter time scales. For $\vartheta = 100$ ps, the network MSD and the MD MSD are closely similar. The

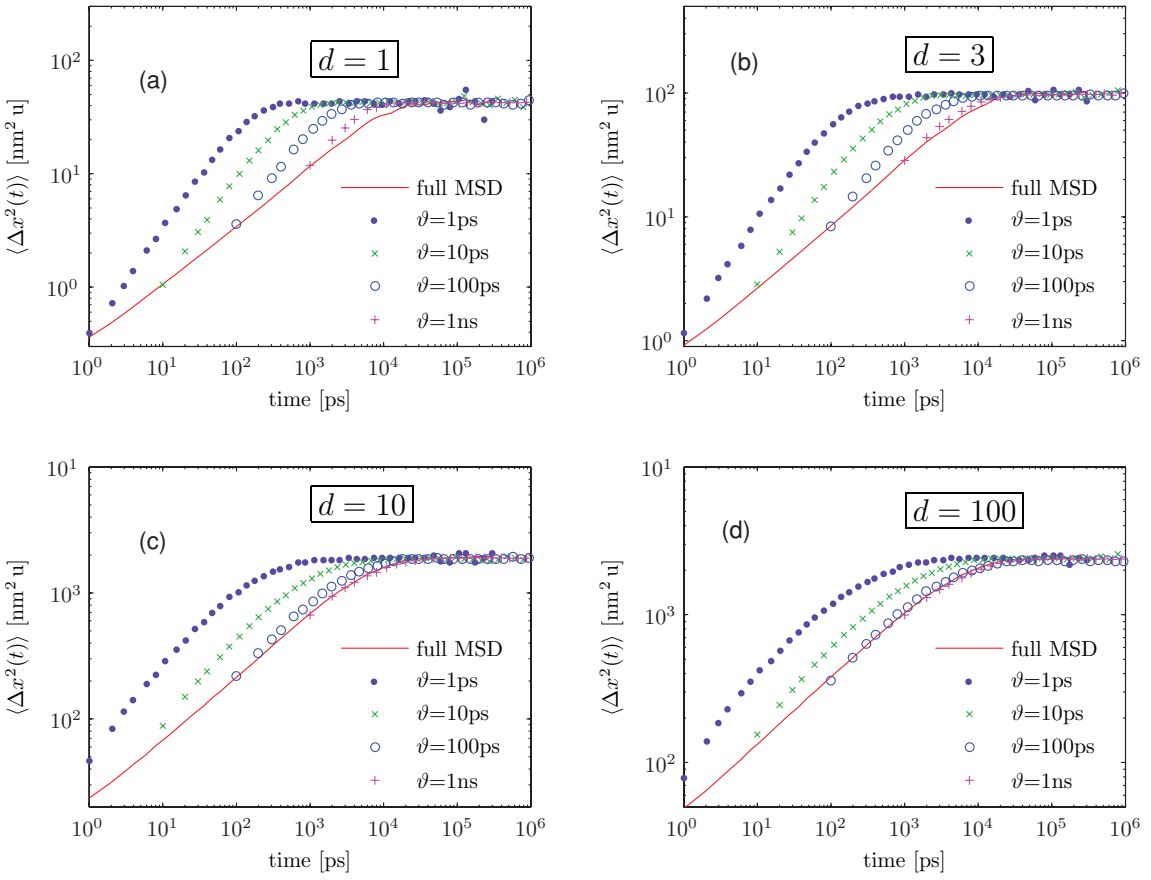

FIG. 2. (Color online) MSD arising from diffusion on a transition network representing the configuration space, for various lag times ($\vartheta = 1$, 10, and 100 ps and 1 ns) and projections onto d-dimensional subspaces, $d = 1$, 3, 10, and 100. The continuous red line represents the MD MSD in the projection with the corresponding d.

TABLE IV. Projection of $(GS)_5W$ onto configurational subspaces—fractal dimension d_f of transition networks. Subspaces with $d = 1$, 3, 10, and 100 with lag time $\vartheta = 1$, 10, and 100 ps, and 1 ns. The fractal dimension of the networks was obtained using Eq. (8). The MSD exponent α was calculated from the random walks illustrated in Fig. 2.

d	d_f				α			
ϑ (ps)	1	10	100	1000	1	10	100	1000
1	1.6	1.6	1.7	1.7	0.9	1.0	0.8	0.6
3	1.4	1.4	1.5	1.6	0.9	0.9	0.8	0.5
10	6.4	6.6	7.0	7.7	0.8	0.7	0.6	0.5
100	6.3	6.5	6.8	7.7	0.7	0.6	0.5	0.3

significance of the networks with $\vartheta = 1$ ns is limited by the fact that the time resolution is close to the time scale of saturation, on which the walker is affected by the finite accessible volume. The finite volume also accounts for the tendency of the MSD exponent α to be underestimated by the random walks on the 1 ns transition network, $\tilde{\mathbf{S}}(\vartheta = 1\text{ ns})$.

In the above analysis, the dynamics were projected onto the subspace spanned by the first ten PCs, i.e., the network was embedded in a d-dimensional space with $d = 10$. In order to characterize the influence of the projection, the dynamics of the $(GS)_5W$ peptide is now also analyzed for projections onto subspaces with $d = 1$, 3, and 100. The fractal dimensions of the resulting networks are given in Table IV. The dimension d_f of the networks obtained from the projections onto $d = 1$ and $d = 3$ dimensions are in the range 1.4–1.7. The d_f values obtained for $d = 10$ are 6.4–7.7, close to those found for $d = 100$. Hence, the fractal structure of the network $\tilde{\mathbf{S}}(\vartheta)$ is essentially developed in the subspace spanned by the first ten PCs. However, the MSD of the network obtained from the projection on $d = 100$ dimensions is considerably closer to the original MSD than that for $d = 10$. The fact that the

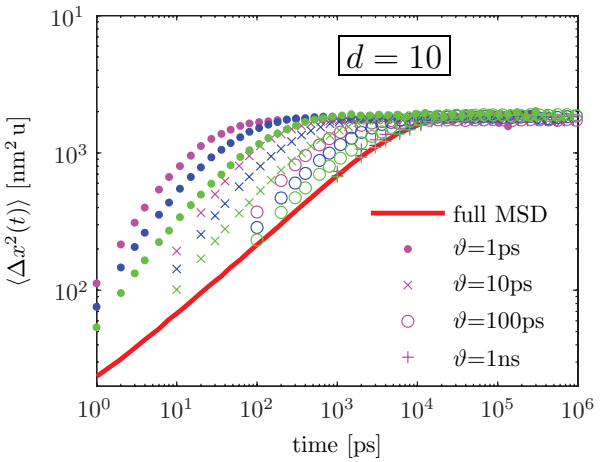

FIG. 3. (Color online) MSD arising from diffusion on a transition network representing the configuration space, for various lag times ($\vartheta = 1$,10, and 100 ps and 1 ns; symbols as given in the legend) and different numbers of states $N_s = 50$,500,5000 (top down) obtained after projections of the full trajectory onto a ten-dimensional subspace. The continuous red line represents the MD MSD in the corresponding projection.

$d = 10$ network exhibits the same fractal exponent as that with $d = 100$ but does not reproduce the kinetics as well demonstrates that, although the higher PC modes (here PC modes 11 to 100) make only a small contribution to the total MSD, they are kinetically significant.

As to the influence of the number of states, we analyzed network models with various numbers of discrete states. The MSD of network models with a low number of discrete states tends to underestimate the subdiffusivity (see Fig. 3). The low spatial resolution contributes a random noise to the trajectory. In particular, for short lag times, the random component dominates and obscures the sublinear part of the MSD.

IV. CONCLUSIONS

The present results demonstrate that the configurational space of a molecule can be represented by a network model, and that a Markov process on the network can reproduce the dynamics of the molecule. Similar results as those in the present paper were obtained for a β-hairpin protein motif (not shown).

Subdiffusive dynamics, as manifested by the sublinear time-dependence of the MSD, arises from a fractal-like geometry of the network. The validity and accuracy of this fractal-geometry representation of the dynamics is confined by two findings. First, memory effects due to the projection of the dynamics are present and lead to violation of the Markov property on short time scales. Second, in order to be numerically tractable and to have sufficient statistics, the number of discrete states is limited and this restricts the spatial resolution of the network in the configuration space. Furthermore, the discretization is not uniquely defined.

The comparison of simulations with explicit and implicit solvent shows that the memory effects due to the projection onto the coordinates of a $(GS)_5W$ peptide, i.e, by neglecting the solvent coordinates, do not have a significant effect on the dynamic behavior of the peptide, in particular, for the MSD, on the time scale of 2 ps (the time resolution of the MD simulation) and longer.

The results show that a Markov process based on a network representation of the configuration space allows the subdiffusive MSD of the MD simulation to be reproduced with high accuracy on time scales of 100 ps and above. The fractal-like nature of the transition network is characterized by the fractal dimension and identified as the essential mechanism that gives rise to subdiffusive dynamics.

The fractal structure of the network is found to be essentially developed in the subspace spanned by the first ten PCs. However, the MSD of the network obtained from the projection on 100 dimensions is considerably closer to the original MSD than that for ten dimensions demonstrating that, although the higher PC modes make only a small contribution to the total MSD, they are kinetically significant.

ACKNOWLEDGMENTS

J.C.S. thanks the United States Department of Energy for financial support under a Laboratory-Directed Research and Development grant to Oak Ridge National Laboratory. I.M.S. thankfully acknowledges financial support from the DFG joint collaborative grant (SFB 555).

[1] K. Henzler-Wildman and D. Kern, Nature (London) **450**, 964 (2007).
[2] B. Hess, Phys. Rev. E **62**, 8438 (2000).
[3] G. R. Kneller and K. Hinsen, J. Chem. Phys. **121**, 10278 (2004).
[4] T. Neusius, I. Daidone, and I. M. Sokolov, Phys. Rev. Lett. **100**, 188103 (2008).
[5] G. R. Kneller, Phys. Chem. Chem. Phys. **7**, 2641 (2005).
[6] G. Luo, I. Andricioaei, X. Sunney Xie, and M. Karplus, J. Phys. Chem. B **110**, 9363 (2006).
[7] W. Min, G. Luo, B. J. Cherayil, S. C. Kou, and X. S. Xie, Phys. Rev. Lett. **94**, 198302 (2005).
[8] S. C. Kou and X. S. Xie, Phys. Rev. Lett. **93**, 180603 (2004).
[9] H. Yang and X. Sunney Xie, J. Chem. Phys. **117**, 10965 (2002).
[10] See supplemental material at [http://link.aps.org/supplemental/10.1103/PhysRevE.83.021902] for mathematical details of the present paper.
[11] H. Neuweiller, M. Löllmann, S. Doose, and M. Sauer, J. Mol. Biol. **365**, 856 (2007).
[12] J. Tang and R. A. Marcus, Phys. Rev. E **73**, 022102 (2006).
[13] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991).
[14] S. Havlin and D. Ben-Avraham, Adv. Phys. **36**, 695 (1987).
[15] I. Daidone, H. Neuweiler, S. Doose, M. Sauer, and J. C. Smith, PLoS Comput. Biol. **6**, e1000645 (2010).
[16] D. van der Spoel, E. Lindahl, and B. Hess, *GROMACS User Manual: Version 3.3*, The GROMACS development team, 2006.
[17] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. **91**, 43 (1995).
[18] D. Brown and J. H. R. Clarke, Mol. Phys. **51**, 1243 (1984).
[19] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, J. Comput. Chem. **18**, 1463 (1997).
[20] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[21] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[22] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, J. Am. Chem. Soc. **112**, 6127 (1990).
[23] D. Bashford and D. A. Case, Annu. Rev. Phys. Chem. **51**, 129 (2000).
[24] D. Qiu, P. S. Shenkin, F. P. Hollinger, and W. C. Still, J. Phys. Chem. A **101**, 3005 (1997).
[25] J. Zhu, Y. Shi, and H. Liu, J. Phys. Chem. B **106**, 4844 (2002).
[26] M. Schaefer, C. Bartels, and M. Karplus, J. Mol. Biol. **284**, 835 (1998).
[27] I. Daidone, M. B. Ulmschneider, A. Di Nola, A. Amadei, and J. C. Smith, Proc. Natl. Acad. Sci. U.S.A. **104**, 15230 (2007).
[28] R. M. Levy, A. R. Srinivasan, W. K. Olson, and J. A. McCammon, Biopolymers **23**, 1099 (1984).
[29] A. E. García, Phys. Rev. Lett. **68**, 2696 (1992).
[30] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen, Proteins: Struct., Funct., Genet. **17**, 412 (1993).
[31] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. (Springer, New York, 2002).
[32] K. S. Kostov and K. F. Freed, Biophys. J. **76**, 149 (1999).
[33] P. Debnath, W. Min, X. S. Xie, and B. J. Cherayil, J. Chem. Phys. **123**, 204903 (2005).
[34] P. E. Rouse, J. Chem. Phys. **21**, 1272 (1953).
[35] G. Strobl, *The Physics of Polymers* (Springer, Berlin, 1997).
[36] R. Granek and J. Klafter, Phys. Rev. Lett. **95**, 098106 (2005).
[37] S. Reuveni, R. Granek, and J. Klafter, Phys. Rev. Lett. **100**, 208101 (2008).
[38] R. M. Venable and R. W. Pastor, Biopolymers **27**, 1001 (1988).
[39] H. Yang, G. Luo, P. Karnchanaphanurach, T.-M. Louie, I. Rech, S. Cova, L. Xun, and X. Sunney Xie, Science **302**, 262 (2003).
[40] J. Tang and S.-H. Lin, Phys. Rev. E **73**, 061108 (2006).
[41] S. Fugmann and I. M. Sokolov, J. Chem. Phys. **131**, 235104 (2009).
[42] 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, Phys. Rev. Lett. **62**, 1916 (1989).
[43] P. W. Fenimore, H. Frauenfelder, B. H. McMahon, and R. D. Young, Proc. Natl. Acad. Sci. U.S.A. **101**, 14408 (2004).
[44] W. Doster, S. Cusack, and W. Petry, Nature (London) **337**, 754 (1989).
[45] R. H. Austin, K. Beeson, L. Eisenstein, H. Frauenfelder, I. C. Gunsalus, and V. P. Marshall, Phys. Rev. Lett. **32**, 403 (1974).
[46] A. Ansari, J. Berendzen, S. F. Bowne, H. Frauenfelder, I. E. T. Iben, T. B. Sauke, E. Shyamsunder, and R. D. Young, Proc. Natl. Acad. Sci. U.S.A. **82**, 5000 (1985).
[47] C. Monthus and J.-P. Bouchaud, J. Phys. A **29**, 3847 (1996).
[48] S. Burov and E. Barkai, Phys. Rev. Lett. **98**, 250601 (2007).
[49] A. Kitao, S. Hayward, and N. Go, Proteins: Struct., Funct., Genet. **33**, 496 (1998).
[50] J.-P. Bouchaud and A. Georges, Phys. Rep. **195**, 127 (1990).
[51] H. Scher and E. W. Montroll, Phys. Rev. B **12**, 2455 (1975).
[52] R. Metzler and J. Klafter, Phys. Rep. **339**, 1 (2000).
[53] T. Neusius, I. M. Sokolov, and J. C. Smith, Phys. Rev. E **80**, 011109 (2009).
[54] D. J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, 2003).
[55] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 5th ed. (Elsevier, Amsterdam, 2004).
[56] F. Noé, D. Krachtus, J. C. Smith, and S. Fischer, J. Chem. Theory Comput. **2**, 840 (2006).
[57] S. Auer, M. A. Miller, S. V. Krivov, C. M. Dobson, M. Karplus, and M. Vendruscolo, Phys. Rev. Lett. **99**, 178104 (2007).
[58] F. Noé, I. Horenko, C. Schütte, and J. C. Smith, J. Chem. Phys. **126**, 155102 (2007).
[59] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, J. Chem. Phys. **126**, 155101 (2007).
[60] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, New York, 2001).