Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supplemental Material for

Role of ergodicity, aging, and Gaussianity in resolving the origins of biomolecule subdiffusion

Jun Li*

School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China

*To whom correspondence should be addressed, E-mail: phyjun@sjtu.edu.cn

This PDF file includes:

METHODS

Details of All-atom molecular dynamics simulations

S1. ssDNA

The single-stranded DNA (ssDNA) studied here is polyadenylic acid with 8 monomers of thymine (poly dT_8), and the cartoon conformation presents in Fig. 1b (blue). The starting configuration for the MD study is generated by the x3dna module (http://web.x3dna.org). The GROMACS (version 2016.3) tools package [1] was used in conjunction with the AMBER99SB force field [2] to model ssDNA dynamics. Poly dT_8 was solvated in a periodic box of dimensions 5×5×5 nm^3. This contained approximately 3,950 TIP3P models [3] water molecules with the appropriate amount of Na^+ counterions to neutralize the negative phosphate charges, and the total system contains about 12,000 atoms. Periodic boundary conditions (PBC) were applied in all dimensions with long-range electrostatic interactions characterized by Particle Mesh Ewald (PME) method [4,5]. The pressure/temperature coupling was performed using the Parrinello-Rahman algorithm [6] and Nosé-Hoover methods [7], respectively. MD simulation was carried out using the leap-frog algorithm for integrating Newton's equation of motion for 100 ns at constant temperature (300 K) and pressure (1 bar). Van der Waals interaction was truncated at 1.2 nm, with the LJ potential switched to zero gradually at 1.0 nm [5]. The short-range electrostatic interactions was truncated at r_c = 12 Å [4,5]. Following that, a 100 ns long simulation was produced with a time step of 2 fs integration. The coordinates of the system coordinates were saved at every 1 ps, and this time resolution is commonly used for molecular dynamics simulations of proteins [8-13].

S2. dsDNA

Here, we carried out MD simulations on the 11 base-pairs B-form double-stranded DNA (dsDNA) with the sequential sequence of GCACTGCTAGG, and its radius of gyration (R_g) is ~1.2 nm. A three-dimensional cartoon structure of studied dsDNA visualized in Fig. 1b (green), and its initial configuration was generated by the x3dna module (http://web.x3dna.org). One hundred independent 100 ns long MD simulations were carried out by the simulation engine GROMACS (Version 2016.3) [1] and the AMBER99SB force field for nucleic acid [14]. A dsDNA was solvated in a periodic box of dimensions 6×6×6 nm^3. The simulated system comprised a small dsDNA and about 6,700 TIP3P water molecules [3], together with the appropriate amount of sodium counterions to balance the negative phosphate charges, leading to this total system size of about 20,000 atoms. The temperature was set to be 300K using the Nosé-Hoover methods [7]. The pressure coupling was kept

constant using the Parrinello-Rahman algorithm [6] to the reference pressure one atmospheric (1 atm). Van der Waals forces were evaluated using a 10-12 Å switching scheme. Long-range electrostatic forces were computed using the particle-Mesh Ewald (PME) methods, 1.5 Å Fourier-space grid, and a 12 Å cut-off for the real-space Coulomb interaction [4,5]. The system coordinates were written into the trajectory file every 1 ps and integrated every 2 fs. We used the end-to-end distance of dsDNA to characterize its internal dynamics.

S3. KRAS

Human KRAS (PDB Code: 3GFT) is a single domain GTPase protein involved in the RAS/MAPK signaling pathway and is a tumor suppressor in pharmaceutics. The molecular weight of KRAS is roughly 21 kD with 167 amino acids, and its R_g is ~1.5 nm. Here, KRAS is tested as a model for studying the structural dynamics of small, single-domain proteins. To characterize KRAS internal dynamics, we calculate the distance between two equal-size segments, i.e., residues 1-76 and 77-167. One hundred independent 100 ns simulations (using the same initial coordinates but different velocities) were produced in the NPT ensemble (constant number of particles, pressure, and temperature) using the MD engine GROMACS (version 2016.3) [1] with the CHARMM36 force field on the in-house supercomputing cluster. Volume of the simulated box is 7×7×7.5 nm^3. We used the Berendsen thermostat method [7] to maintain the simulation temperature at 300K and Parrinello-Rahman barostat [6] coupling to keep the pressure at reference (1 bar). Energy minimization of 50,000 steps was initially used, and then NVT and NPT ensembles were performed for 10ns each. The TIP3P explicit water model was chosen [3], and sodium ions were added to the system to neutralize the total charge. PME algorithms were used to calculate long-range electrostatic interactions [4,5], with a cutoff of 1 nm. The short-range van der Waals (VdW) interactions were treated using a cutoff of 10 Å. For all the 100 ns simulations, system coordinates were saved every 1 ps, and a time step of 2 fs was used for integration MD.

S4. PGK

A typical cartoon structure of yeast enzyme phosphoglycerate kinase (PGK) presents in Fig. 1b, which contains N- and C-terminal domains (residues 1-185, residues 200-389) linked by a helix hinge (residues 186-199 & 390-415). This three-domain protein's molecular weight is about 45 kD with 415 residues (PDB ID: 3PGK) [15]. One hundred 100 ns long MD simulations were performed to examine its dynamics. The simulation started from the initial crystal configuration, and the MD

software GROMACS (Version 2016.3) was used [1], as it is the most used tool in the study of biomolecular dynamics. The force field used for protein was the CHARMM27 force field on a local computing cluster [16,17]. PGK, which was placed at least 1.0 nm from the box edge, was simulated inside cubic cell center (8×7.5×10 nm^3) using periodic boundary conditions (PBC), leading to this total system size of about 101000 atoms, ~95000 TIP3P water molecules filled into the box [3]. In the isothermal−isobaric (NPT) ensemble using a Nosé-Hoover thermostat algorithm [7] at 300K and the pressure to 1 bar using Parrinello-Rahman barostat(12). The non-bonded Van der Waals (VdW) interaction was truncated at 1.2 nm, at which the VdW interactions reach zero with the LJ potential gradually at 1.0 nm. The Coulomb cut-off distance of short-range electrostatic interactions was r_c = 12 Å. For a distance beyond 1.2 nm, the Particle Mesh Ewald (PME) method [4,5] was used. The VdW interactions were treated using a cut-off of 1 nm. The simulation coordinates of the system were recorded every 1 ps with at least 100 ps equilibration time.

S5. SHP2

The crystal structure of non-receptor protein tyrosine phosphatase (PTP) SHP2 with E76A mutation can be found in the PDB data bank file 5XZR [18]. The molecular mass of this multi-domain phosphatase is about 64.24 kD with ~530 residues, and the corresponding R_g is about 2.6 nm. All SHP2 simulations were carried out using GROMACS (Version 2016.3) with the CHARMM27 force field for this enzyme [16,17]. The system was solvated in a rectangular water box (edge lengths 8.5×9.5×10.5 nm^3) with periodic boundary conditions (PBC), leading to a total system size of about 83,000 atoms with a single SHP2 protein molecule and ~25000 water molecules. 77 Na+ and Cl- ions were added to neutralize the total charge. All simulations were carried out using the TIP3P water model [3] in the NVT ensemble using a Nośe-Hoover thermostat at 300 K [19]. The pressure coupling was performed using the Parrinello-Rahman algorithm with a coupling time of τ = 2 ps [6]. Van der Waals interactions (VdW) were truncated at 1.0 nm, with the LJ potential switched to zero gradually from 1.0 nm to 1.2 nm. The short-range electrostatic interactions within the cut-off distance of r_c = 10 Å were treated as Coulombic [4,5]. All bonds involving hydrogen atoms were constrained with the LINCS algorithm to allow a time step of 2 fs [20]. The system was first energy minimized using steepest descent steps with a maximum force of 10.0 $kJ \cdot mol^{-1} \cdot nm^{-1}$ and a maximum of 5×10^6 steps, then equilibrated in the NVT ensemble at 300 K for 10 ns, and then in the NPT ensemble at p =1 bar for 10 ns. Then using the final structure of the protein molecule obtained in the

NPT equilibration as the starting structure, we performed 100 independent MD simulations with each being 100 ns long.

S6. SNAP-25

Symbiosomal nerve-associated protein 25 isoform A (SNAP-25), encoding by the SNAP-25 gene in humans, consists of N- and C-terminal α-helix connected by a small random coil linker [21]. The molecular weight of this intrinsically disordered protein (IDP) SNAP-25 is ~23 kD with a primary sequence length of 204 residues, and its radius of gyration (R_g) is ~3.3 nm. SNAP-25, together with syntaxin and synaptobrevin, composes an exceptionally stable four-helix bundle, SNARE complex, an intracellular membrane fusion protein, by pulling the two membranes tightly together to exert the force required for fusion [22,23]. Therefore, the dynamics of SNAP-25 protein is considered to be tremendously crucial for intracellular trafficking and vesicle disassembling. The three-dimensional structure of this disordered protein SNAP-25 (see Fig. 1d) was generated by homology modeling methods [24], based on the native amino acids sequence[25]. This building structure shows 98.53% sequence identity with template protein (a single chain in the PDB ID 6MDM). We performed one hundred 100 ns long trajectory for statistical analysis. All the simulations were carried by using the MD engine GROMACS (version 2016.3) [1] with the specific IDP force field [26] for the protein on a local computing cluster. The constant temperature was set to 300 K using a modified Berendsen thermostat [7] and the pressure to 1 bar using Parrinello-Rahman barostat [6]. The protein is centered in a rectangular water box (~ 6.5×6.5×14.2 nm^3) using the TIP3P explicit water model [3] and placed at least 1.0 nm from the box edge. The whole system contains slightly over 58000 atoms and 14 sodium ions to neutralize the total charge. Electrostatic interactions were treated with a cut-off of 1 nm, beyond which the PME method was used [4,5]. The VdW interactions were treated using a cut-off of 1 nm. For all the simulations, system coordinates were written into the trajectory file every 1 ps. We also performed ensemble simulations with the CHARMM36 force field in SNAP-25. Similar results (i.e., aging and non-ergodicity) are being observed.

Supplementary Notes

S7: Aging examination for different models

We also performed control numerical simulations on classical Brownian motion (BM), fractal Brownian motion (FBM, with a Hurst exponent of 0.3), continuous-time random walk (CTRW), and three different noise strengths CTRW to understand the aging dynamics. We simulated 100 independent trajectories to ensure efficient statistics, and each with 10,000 steps for these models, i.e., the same amount of data compared with simulations for each dataset of the molecule. As displayed in Fig. S1, the time-ensemble-averaged MSD (TEA-MSD) of ergodic Brownian motion or fractal Brownian motion scale linearly with observation time T, and converge during long observations, indicating the equivalence for short-time measurements and long-time observations. As expected, no aging is observed in BM and FBM (Figs. S1A and S1B); its time-averaged MSD is independent of observation time T, demonstrated by many theoretical works [27,28]. In sharp contrast, the random walk caused by trapping events (assuming the transition occurs with a scale-free trapping time distribution $P(\tau) \sim \tau^{-1.5}$) shows apparent aging behaviors (Fig. S1C). The TEA-MSD display power-law decay for different lag-times Δ, consistent with the analytical result of CTRW [27,28]. In other words, the effective mean waiting time was not constant but increased with the observation time, giving rise to aging dynamics [27-30].

A careful comparison between the MD-derived results (Fig. 2f) and non-ergodic CTRW subdiffusion from numerical simulation (Fig. S1C) shows some qualitative differences. Firstly, the decay of TEA-MSD over T at different values of Δ in CTRW is parallel to each other in the double logarithmic scale (Fig. S1C). In contrast, the MD-derived TEA-MSD among different lag-times Δ is gradually reduced upon the increase of T (Fig. 2f, main text). The difference results from the fact that each state in the CTRW is static, but in protein, it involves a group of similar conformations with slight differences due to the thermal fluctuation, i.e., one conformational state coupled with a small amplitude of the noise. Here, we modified bare CTRW by assuming Ornstein-Uhlenbeck noise occurring when trapping in one state [31], where the amplitude of the noise is controlled by a noise strength parameter (η). After such treatment, the numerical results (Figs. S1E) are in much better agreement with these derived directly from MD observed (Fig. 1f). Thus, the noisy CTRW model indeed predicts that the difference of TEA-MSD at different lag times Δ is reduced upon increasing T. OU noise subordinated on broadly distributed trapping events dominated CTRW show closest results (Fig. S1E). Furthermore, adding more significant noise (Fig. S4F, η=0.5) to non-ergodic CTRW, no aging is observed and ergodic [31].

Supplementary Figures

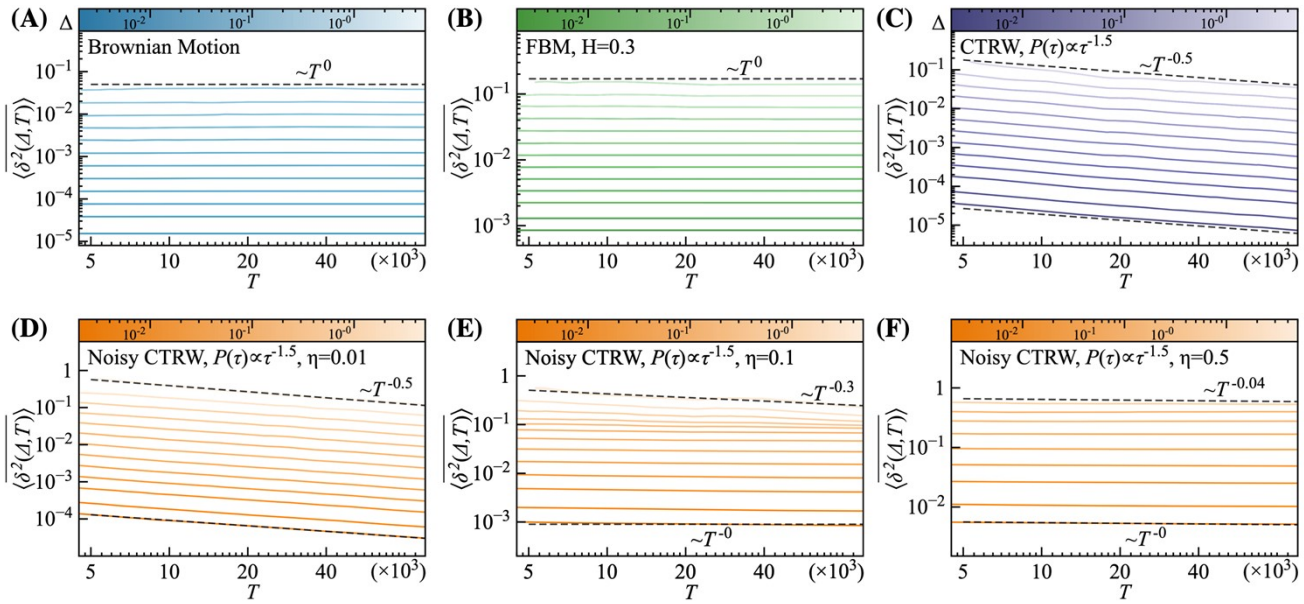

Figure S1 | The aging plots obtained from the numerical simulation of different models. The aging plot of time-ensemble-averaged MSD (TEA-MSD) derived from numerical simulation of Brownian motion (**A**), and fractal Brownian motion (**B**). (**C**) TEA-MSD as a function of T obtained from the continuous-time random walk (CTRW). (**D-F**) Three different simulated noisy CTRW by addition Ornstein Uhlenbeck noise to each state, the respective noise strengths (η) are indicated. All dashed lines represent power-law fits and guide the aging exponent of TEA-MSD vs. T.

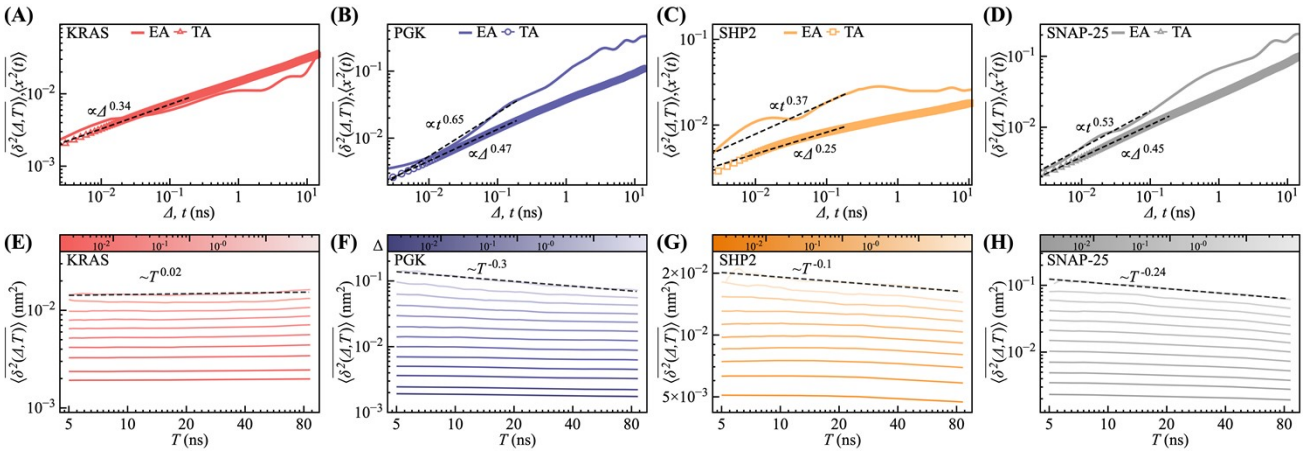

Fig. S2 | Using end-to-end distance to characterize ergodicity and aging dynamics in all proteins. The ensemble-averaged MSD (EA-MSD, solid lines) vs. time-ensemble averaged MSD (TA-MSD, marker) for the KRAS (**A**), PGK (**B**), SHP2 (**C**), and SNAP-25 (**D**). The overlap between EA-MSD and TA-MSD indicates ergodic motion in KRAS, revealing no appreciable non-ergodicity. While the gap between TA-MSD and EA-MSD confirms the ergodic breaking on the MD time window (**B-D**). (**E-H**) The aging TA-MSD are plotted against observation time, T (i.e., the time used in the moving average), with different fixed lag times Δ. The dashed lines in (**E-H**) guide the trend of power-law decay, and suggest non-aging (**E**) and aging (**F-H**) behaviors at the associated biomolecules. Although the sub-diffusive and aging power-law exponents are slightly different from Fig. 1 (main text), the qualitative results, i.e., the presence of non-ergodicity and aging in larger proteins, are similar to the manuscript.

Supplementary Tables

Table S1 | The distance *x*(*t*) used to characterize the internal dynamics of six biological molecules. The reasons why we chose these different distances $x(t)$ for respective proteins is described below. For the distance $x(t)$ of four protein cases (KRAS, PGK, SHP2, SNAP-25), we use the same examination of Ref. [8,32,33]. More specifically, to examine the overall intra-domain structural dynamics of KRAS like the inter-domain motions in the globular proteins, we divided the protein into two nearly equal-weighted segments, i.e., residues 1-76 and 77-167, calculated the center-of-mass distance from the MD trajectories, and compared the results to the inter-domain dynamics of PGK and SHP2. The PGK distance $x(t)$ is selected as the center of mass distance between two fist-like N- and C-domains (i.e., residues 1-185 and 200-389), which is functionally important [8]. Since the substrate covalently binds to the N- and C-domain to perform its catalytic function. The SHP2 distance $x(t)$ is between two selected residues (Q87 and K266) located in the N-SH2 and PTP domains. As previously reported [32], the relative motion between the N-SH2 and PTP domains is essential for its function and shows crucial open and close conformational change. The SNAP-25 distance $x(t)$ is between two selected residues (Q20 and I139), which have demonstrated conformational switching behavior and structural heterogeneity as recently reported [33].

Molecules	Distance, $x(t)$
ssDNA	End-to-end distance
dsDNA	End-to-end distance
KRAS	Center of mass distance between residues 1-76 and 77-167
PGK	Center of mass distance between residues 1-185 and 200-389
SHP2	Center of mass distance between residues 87 and 266
SNAP-25	Center of mass distance between residues 20 and 139

Table S2 | The subdiffusive exponents *α* for different molecules. The α_e and α_t represent the ensemble and time values of the subdiffusive exponent, respectively. We estimated α using simple power-law scaling to fit the two type MSD curves (Figs. 1c-d) at times ranging from 1 - 100 ps.

	ssDNA	dsDNA	KRAS	PGK	SHP2	SNAP-25
α_e	0.69	0.65	0.22	0.95	0.38	0.65
α_t	0.68	0.62	0.22	0.59	0.27	0.41

SI REFERENCESw

1 M. Abraham, D. van der Spoel, E. Lindahl and B. Hess, *Google Scholar There is no corresponding record for this reference*.

2 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 712-725.

3 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926-935.

4 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089-10092.

5 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577-8593.

6 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182-7190.

7 H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola and J. Haak, *J. Chem. Phys.*, 1984, **81**, 3684-3690.

8 X. Hu, L. Hong, M. D. Smith, T. Neusius, X. Cheng and J. C. Smith, *Nat. Phys.*, 2016, **12**, 171.

9 G. Luo, I. Andricioaei, X. S. Xie and M. Karplus, *J. Phys. Chem. B*, 2006, **110**, 9363-9367.

10 R. Satija, A. Das and D. E. Makarov, *J. Chem. Phys.*, 2017, **147**, 152707.

11 T. Neusius, I. Daidone, I. M. Sokolov and J. C. Smith, *Phys. Rev. E*, 2011, **83**, 021902.

12 T. Neusius, I. Daidone, I. M. Sokolov and J. C. Smith, *Phys. Rev. Lett.*, 2008, **100**, 188103.

13 T. Neusius, I. M. Sokolov and J. C. Smith, *Phys. Rev. E*, 2009, **80**, 011109.

14 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins: Struct., Funct., Bioinf.*, 2010, **78**, 1950-1958.

15 H. Watson, N. Walker, P. Shaw, T. Bryant, P. Wendell, L. Fothergill, R. Perkins, S. Conroy, M. Dobson and M. Tuite, *The EMBO journal*, 1982, **1**, 1635-1640.

16 A. D. MacKerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo and S. Ha, *J. Phys. Chem. B*, 1998, **102**, 3586-3616.

17 A. D. Mackerell Jr, M. Feig and C. L. Brooks III, *J. Comput. Chem.*, 2004, **25**, 1400-1415.

18 J. Xie, X. Si, S. Gu, M. Wang, J. Shen, H. Li, J. Shen, D. Li, Y. Fang and C. Liu, *Journal of medicinal chemistry*, 2017, **60**, 10205-10219.

19 S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511-519.

20 B. Hess, *J. Chem. Theory Comput.*, 2008, **4**, 116-122.

21 D. Fasshauer, R. B. Sutton, A. T. Brunger and R. Jahn, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 15781-15786.

22 P. Washbourne, P. M. Thompson, M. Carta, E. T. Costa, J. R. Mathews, G. Lopez-Benditó, Z. Molnár, M. W. Becher, C. F. Valenzuela and L. D. Partridge, *Nat. Neurosci.*, 2002, **5**, 19.

23 K. Weninger, M. E. Bowen, U. B. Choi, S. Chu and A. T. Brunger, *Structure*, 2008, **16**, 308-320.

24 P. Benkert, M. Biasini and T. Schwede, *Bioinformatics*, 2010, **27**, 343-350.

25 D. Fasshauer, W. K. Eliason, A. T. Brünger and R. Jahn, *Biochemistry*, 1998, **37**, 10354-10362.

26 D. Song, R. Luo and H.-F. Chen, *J. Chem. Inf. Model.*, 2017, **57**, 1166-1178.

27 Y. Meroz and I. M. Sokolov, *Phys. Rep.*, 2015, **573**, 1-29.

28 R. Metzler, J.-H. Jeon, A. G. Cherstvy and E. Barkai, *Phys. Chem. Chem. Phys.*, 2014, **16**, 24128-24164.

29 A. V. Weigel, B. Simon, M. M. Tamkun and D. Krapf, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 6438-6443.

30 S. A. Tabei, S. Burov, H. Y. Kim, A. Kuznetsov, T. Huynh, J. Jureller, L. H. Philipson, A. R. Dinner and N. F. Scherer, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 4911-4916.

31 J. H. Jeon, E. Barkai and R. Metzler, *J. Chem. Phys.*, 2013, **139**, 121916.

32 Y. Tao, J. Xie, Q. Zhong, Y. Wang, S. Zhang, F. Luo, F. Wen, J. Xie, J. Zhao and X. Sun, *J. Biol. Chem.*, 2021, **296**, 100538.

33 N. Saikia, I. S. Yanez-Orozco, R. Qiu, P. Hao, S. Milikisiyants, E. Ou, G. L. Hamilton, K. R. Weninger, T. I. Smirnova and H. Sanabria, *Cell Reports Physical Science*, 2021, **2**, 100616.