Please cite this article in press as: Kompella et al., What determines sub-diffusive behavior in crowded protein solutions?, Biophysical Journal (2024), https://doi.org/10.1016/j.bpj.2023.12.002

Biophysical Journal
Article

Biophysical Society

What determines sub-diffusive behavior in crowded protein solutions?

Vijay Phanindra Srikanth Kompella,[1,2] Maria Carmen Romano,[2,3] Ian Stansfield,[3] and Ricardo L. Mancera[1,*]
[1]Curtin Medical School, Curtin Health Innovation Research Institute, Curtin Institute for Data Science, Curtin University, Perth, Western Australia, Australia; [2]Department of Physics, Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen, United Kingdom; and [3]Institute of Medical Sciences, University of Aberdeen, Aberdeen, United Kingdom

ABSTRACT The aqueous environment inside cells is densely packed. A typical cell has a macromolecular concentration in the range 90–450 g/L, with 5%–40% of its volume being occupied by macromolecules, resulting in what is known as macromolecular crowding. The space available for the free diffusion of metabolites and other macromolecules is thus greatly reduced, leading to so-called excluded volume effects. The slow diffusion of macromolecules under crowded conditions has been explained using transient complex formation. However, sub-diffusion noted in earlier works is not well characterized, particularly the role played by transient complex formation and excluded volume effects. We have used Brownian dynamics simulations to characterize the diffusion of chymotrypsin inhibitor 2 in protein solutions of bovine serum albumin and lysozyme at concentrations ranging from 50 to 300 g/L. The predicted changes in diffusion coefficient as a function of crowder concentration are consistent with NMR experiments. The sub-diffusive behavior observed in the sub-microsecond timescale can be explained in terms of a so-called cage effect, arising from rattling motion in a local molecular cage as a consequence of excluded volume effects. By selectively manipulating the nature of interactions between protein molecules, we determined that excluded volume effects induce sub-diffusive dynamics at sub-microsecond timescales. These findings may help to explain the diffusion-mediated effects of protein crowding on cellular processes.

SIGNIFICANCE Protein crowding governs cell processes by altering the diffusion of biomolecules. Using Brownian dynamics simulations of cell-like densely crowded environments, we show that volume exclusion is solely responsible for sub-diffusion of proteins at the sub-microsecond scale and that molecules exhibit optimized search for interaction partners, wherein spatially confined movement at short timescales (sub-diffusion) maximizes encounter probability, and the normal diffusion observed at long timescales facilitates long-range search. These findings show the importance of accounting for sub-diffusion while estimating reaction rates, especially for the purposes of modeling biochemical reaction networks regulated by protein-protein interactions that are facilitated by diffusion. We also highlight that a well-defined macromolecular composition is necessary to accurately describe the crowding effect in both computational and experimental studies.

INTRODUCTION

Living cells contain a variety of macromolecules that maintain their functional and structural integrity. The concentration of macromolecules inside the cells varies from 90 to 450 g/L (1,2) depending on the type of cell (3), cell differentiation stage, and the organelle of interest (4), occupying nearly 5%–40% of cellular volume (2). The resulting intracellular macromolecular crowding is known to affect a variety of processes including but not limited to translation and growth (5), intracellular signaling (6), and transport mediated by molecular motors such as kinesin-1 (7). Moreover, extracellular crowding is known to play a significant role in accelerating the development of the extracellular matrix necessary in tissue-engineering approaches (8). These effects arise from altered molecular interactions, transport properties, and excluded volume effects due to crowding (9), (10–21). A more detailed understanding of the molecular mechanisms through which crowding mediates its effects is crucial to explain the above phenomena.

Diffusion is one of the major transport properties affected by crowding. Decreased diffusion rate as a result of crowding has been documented experimentally, through in vivo (22–25) and in vitro (26–28) approaches, as well as various

Submitted June 11, 2023, and accepted for publication December 4, 2023.
*Correspondence: r.mancera@curtin.edu.au
Editor: Rebecca Wade.
https://doi.org/10.1016/j.bpj.2023.12.002

Please cite this article in press as: Kompella et al., What determines sub-diffusive behavior in crowded protein solutions?, Biophysical Journal (2024), https://doi.org/10.1016/j.bpj.2023.12.002

computational approaches (20,29,30). Einstein's equation describes the relationship between the 3D diffusion coefficient (D) of a particle and its mean squared displacement (MSD) ($<x^2$ (t)$>$), which is the second moment of displacement distribution (Eq. 1) (31).

$$\langle x^2(t)\rangle = 6Dt^{\alpha} \quad (1)$$

Here, $<x^2(t)>$ and D are the ensemble average of the MSD (EAMSD) and diffusion coefficient, respectively. The diffusion coefficient is a measure of the rate (slow or fast) of diffusion and deviation from normal (linear) diffusion is determined by the α-exponent (Eq. 1). Normal diffusion is characterized by an α-exponent of 1.0, whereas $\alpha > 1.0$ and $0 < \alpha < 1.0$ correspond to super- and sub-diffusion, respectively.

Sub-diffusion has been observed in multiple experimental studies. Fluorescence recovery after photobleaching, extended to three dimensions, was used to establish sub-diffusion of green fluorescent protein (GFP) in HeLa cells (32). Using fluorescence correlation spectroscopy, sub-diffusion of microinjected dextran was established in vivo (33), and fluorescence-fluctuation analysis of raster scans was used to study GFP sub-diffusion (34). In all of these studies, the best time resolution was in the order of microseconds. Sub-diffusive behavior has been observed in coarse-grained simulations of bacterial cytoplasm (35), atomistic simulations of the cytoplasm of *Escherichia coli* (18); lipid bilayer systems (36,37), especially at low hydration (38); and Brownian dynamics and atomistic simulations of crowded protein solutions (20,39). Additionally, in the case of bacterial cytoplasm, Ando and Skolnick have shown, using both Brownian dynamics simulations without hydrodynamic interactions (HIs) and Stokesian dynamics simulations with HIs, that sub-diffusion is observed in timescales <1 μs (29). The sub-diffusive behavior in the cytoplasmic systems and protein solutions was observed in the sub-microsecond scale and, therefore, it is challenging to characterize this behavior using experimental approaches (39). Although sub-diffusion has been predicted in multiple simulation studies, there is limited understanding of the mechanism through which protein crowders induce this phenomenon.

Atomistic simulations have shown that chymotrypsin inhibitor 2 (CI2) exhibits significant sub-diffusive behavior in the presence of bovine serum albumin (BSA) as the crowder protein, whereas no appreciable sub-diffusive behavior was observed in the presence of lysozyme (20). We note that the spatio-temporal scale explored in these simulations was relatively small, whereby a single molecule of CI2 was simulated in the presence of eight protein crowder molecules for 117–244 ns at a crowder concentration of 100 g/L. Lysozyme exhibited a propensity to interact more strongly with CI2, whereas BSA crowders mostly interacted with each other. Nawrocki et al. later explained that the absence of substantial interactions between CI2 and BSA resulted in cage effects (i.e., where the motion of a molecule is akin to that of a particle trapped in a cage), which led to sub-diffusive dynamics (21). However, to the best of our knowledge, no attempt has yet been made to quantitatively associate cage effects and sub-diffusive behavior in crowded protein solutions.

There are multiple stochastic processes that display anomalous, sub-diffusive behavior. These include continuous time random walk (CTRW), fractional Brownian motion (fBm), heterogeneous diffusion process, and scaled Brownian motion, each exhibiting different features (40,41). For example, fBm is an ergodic process, whereas CTRW is nonergodic, and other stochastic processes such as scaled Brownian motion and the heterogeneous diffusion process exhibit sub-diffusion and weak ergodicity breaking (40). Anomalous diffusion in dextran solutions has been explained using fBm (42). A CTRW was used to explain the lateral diffusion of potassium channels in cells (43), and a hybrid CTRW and fBm approach was invoked to explain sub-diffusive transport of insulin granules inside cells (44). Therefore, before calculating the diffusion coefficient and α-exponent, it is important to verify whether the system is ergodic so that the EAMSD can be estimated using the time-averaged MSD (TAMSD) (40).

Slow diffusion has been explained in terms of excluded volume effects or cluster formation. Depending on the timescale, three types of cluster formation have been identified in proteins: transient, dynamic, and permanent clusters (in increasing order of lifetime) (45). Clusters that have a lifetime shorter than the time it takes for them to diffuse a distance equivalent to one protein diameter are defined as transient clusters, and the diffusion properties of the system mimic that of monomers. If the lifetime is longer, the short-time dynamic behavior is then determined by so-called dynamic clusters. On the other hand, permanent clusters have a lifetime longer than experimental timescales (45). The slow diffusion observed in crowded protein solutions of hen egg white lysozyme, ubiquitin, villin, and the third immunoglobulin G-binding domain of protein G headpiece has been explained by the formation of dynamic clusters (19). By contrast, formation of transient clusters was invoked to explain the slow diffusion in villin crowded systems at 135 g/L (21). Stokesian dynamics simulations of bacterial cytoplasm showed the importance of HI in regulating diffusion in crowded environments (29). On the other hand, the diffusion measured in the microsecond timescale in solutions crowded with myoglobin and hemoglobin was found to be largely dependent on excluded volume effects (46). However, the extent to which cluster formation and excluded volume effects influence sub-diffusion is not clearly understood; therefore, these phenomena need to be further characterized to explain sub-diffusion in crowded protein solutions.

We have investigated the causal relationships of sub-diffusion in crowded protein solutions. A CI2 tracer in a

Please cite this article in press as: Kompella et al., What determines sub-diffusive behavior in crowded protein solutions?, Biophysical Journal (2024), https://doi.org/10.1016/j.bpj.2023.12.002

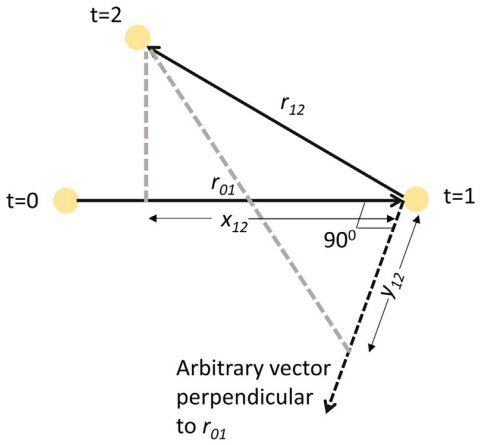

FIGURE 1 Quantification of cage effects. The particle is shown in yellow. The arbitrary vector perpendicular to r_{01} is shown as a dotted line vector. To see this figure in color, go online.

protein crowded environment of BSA and lysozyme were chosen as a model system since they have been studied experimentally (26). Simulations were performed at a range of concentrations (50–300 g/L) of both protein crowders using a grid-based Brownian dynamics approach (47,48). The role of excluded volume effects and protein clustering in inducing sub-diffusive behavior was discerned by selectively altering the forces between molecules. Our findings suggest that excluded volume effects, via cage effects, cause sub-diffusive dynamics at sub-microsecond timescales in crowded protein solutions.

MATERIALS AND METHODS

The experimentally determined 3D structures of BSA (PDB: 3V03), CI2 (PDB: 2CI2), and LYS (PDB: 1AKI) were obtained from the PDB. The Simulation of Diffusional Association (SDA, version 7.2.2) program was used to conduct Brownian dynamics simulations (47). Pre-processing of the proteins, as described below, was done with webSDA (48). The protonation states of amino acids in all proteins were assigned assuming a pH of 5.4 to emulate experimental conditions. Atomic charges and radii were taken from the AMBER force field 99 (49). Electrostatic grids of 1.0 Å resolution were calculated assuming an ionic strength of 200 mM (to also reproduce experimental conditions), with an ion radius of 1.5 Å, a protein dielectric constant of 4.0, a solvent dielectric constant of 78.0, and a temperature of 300 K, using the linearized Poisson-Boltzmann equation approach (50). The electrostatic grids of LYS and CI2 were $129 \times 129 \times 129$ Å^3 in size and the grid size of BSA was $193 \times 129 \times 161$ Å^3, reflecting the differences in size and shape of these proteins. Effective charges were calculated using webSDA. Electrostatic desolvation, hydrophobic desolvation, and Lennard-Jones energy grids were calculated at a resolution of 1.0 Å. The grid sizes of the electrostatic desolvation and Lennard-Jones (repulsive) energies of BSA, LYS, and CI2 were $133 \times 92 \times 109$ Å^3, $45 \times 55 \times 67$ Å^3, and $43 \times 44 \times 45$ Å^3, respectively. The size of the hydrophobic desolvation energy grids of BSA, LYS, and CI2 were $104 \times 76 \times 87$ Å^3, $45 \times 52 \times 60$ Å^3, and $44 \times 44 \times 45$ Å^3, respectively. The energy grid files obtained were then used to set up simulations with protein crowder concentrations of 50, 100, 200, and 300 g/L, with CI2 as the tracer. Initial configurations were generated using the *gen-box* tool in SDA by placing the proteins randomly in a cubic box of 350 Å length. To account for the potential influence of the initial configuration of the proteins in each system, three systems with different initial configurations were set up for every concentration.

The simulations were performed using SDAMM (a program used for simulations with multiple molecules) in SDA with a time step of 0.5 ps at the default SDA temperature of 300 K. Each of the simulations was run with the soft-core repulsive term only for 1 μs to remove any protein overlaps. The simulations were then run for 1 μs with the full energy term for equilibration purposes, followed by 5 μs of production runs. The self-diffusion coefficients of BSA and LYS were monitored to evaluate convergence, which was reached before 1 μs. Since both these crowders are larger than CI2, the convergence of the diffusion of BSA and LYS was expected to be slower and hence was used in this evaluation. Diffusion coefficients were calculated from the plots of TAMSD (obtained by averaging over all possible time origins) vs. time (lag time). The simulations with the soft-core repulsive term (decaying at a rate of $1/r^6$) only were performed using the same approach as above except that neither the equilibration nor the production run included attractive interactions (the scaling factor of electrostatic, electrostatic desolvation, and hydrophobic desolvation terms is set to zero) in the energy term. The trajectories were unfolded assuming that any given particle does not move more than half the simulation cell length between time frames considered (51).

Calculation of the α-exponent

The value of the α-exponent was calculated from the log(TAMSD/τ) vs. log(τ) curve (the linear relationship between the two variables can be derived from a time-averaged equivalent of Eq. 1, where EAMSD is replaced with TAMSD and t is replaced with lag time (τ)) using an approach similar to that of Balbo et al. (39). Since the α-exponent is a time-varying quantity in our simulations, the straight-line region of the plot is chosen by fitting the parts of the curve to a linear fit in such a way that the R^2 value is maintained above a cutoff of 0.95. The regions at long timescales usually showed high levels of noise, which affected the quality of the fit. This is due to the use of TAMSD in our calculations, such that the MSD calculation is affected at large lag time values due to poor statistics. Therefore, long-timescale regions with poor statistics were omitted from the calculations of the α-exponent. The average values of the α-exponents calculated using the data from the simulations with three different initial configurations are reported in Figs. 3, 5, S13, and S14, and the error bars in the plots correspond to the standard deviation (σ). All p values were calculated using two-tailed *t*-tests assuming unequal variance. It is important to note that, since log plots are used, the data at long timescales are crowded in a small region of the graph and, as a result, although one can reliably calculate diffusion coefficients up to the order of a microsecond (in the TAMSD vs. τ plots), it is not feasible to do a similar calculation of the α-exponent at long timescales with a stringent cutoff. However, since the α-exponent converges back to normal diffusion values within the range of timescales explored, this does not have any impact on our conclusions. Additionally, as elaborately discussed in the supporting material, power law fits were also used to confirm our findings on sub-diffusion that were inferred from the log-log plots.

Quantification of cage effects

Cage effects were quantified using Doliwa and Heuer's approach (52). Here, the displacement vector of a particle is given by $r_{mn}(\tau) = r(n\tau) - r(m\tau)$, where $r(n\tau)$ and $r(m\tau)$ are the position vectors at corresponding time points (Fig. 1). The first and second displacement vectors are therefore termed $r_{01}(\tau)$ and $r_{12}(\tau)$, respectively. The component of $r_{12}(\tau)$ along $r_{01}(\tau)$ is termed $x_{12}(\tau)$. The component of $r_{12}(\tau)$ along an arbitrary vector perpendicular to $r_{01}(\tau)$ is termed $y_{12}(\tau)$. According to this approach, it is expected that $x_{12}(\tau)$ be negative and decrease linearly with an increase in the magnitude of $r_{01}(\tau)$ in the presence of cage effects. This

Please cite this article in press as: Kompella et al., What determines sub-diffusive behavior in crowded protein solutions?, Biophysical Journal (2024), https://doi.org/10.1016/j.bpj.2023.12.002

A

	Diffusion coefficient in 10^{-7} cm^2s^{-1}			
	Simulation		Experiment	
Conc. Of the crowder in g/L	CI2 in BSA	CI2 in LYS	CI2 in BSA	CI2 in LYS
50	14.59	13.49	13.22	12.20
100	12.90	13.24	8.18	5.12
200	9.12	9.50	5.51	3.86
300	5.59	5.78	2.66	1.39

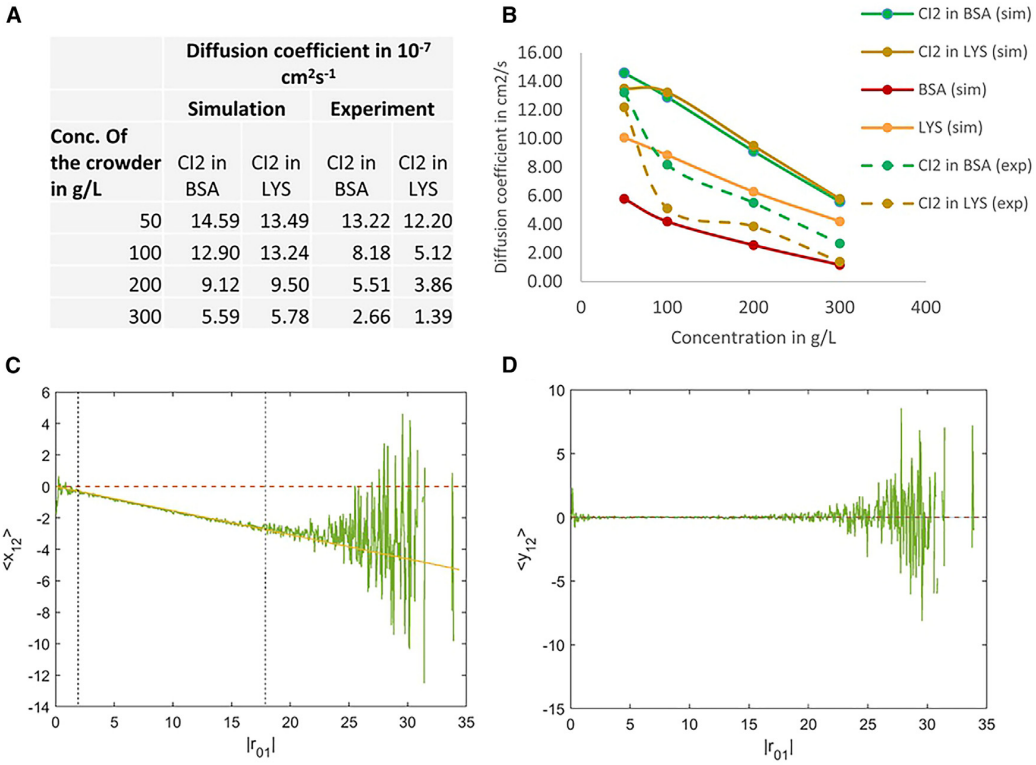

FIGURE 2 Predicted diffusion properties in crowded protein solutions. (*A*) Comparison of experimental and predicted CI2 diffusion coefficients. The predicted values are within the same order of magnitude of experiment, revealing good agreement. (*B*) The predicted and experimentally determined diffusion coefficients of the CI2 tracer in the presence of the protein crowders BSA and lysozyme and of the protein crowders themselves are plotted as a function of crowder concentration. As expected, the increase in crowder concentration results in a downward trend of the diffusion coefficient of CI2. (*C*) Average x_{12} as a function of $|r_{01}|$ (green), whereas the dashed red line corresponds to the reference $x = 0$ curve, and the dotted vertical line separates the regions of low and high noise. The yellow line corresponds to the linear fit for the less noisy region, whose slope is used in the calculation of α-exponent. The slope is negative, indicating the presence of cage effects. (*D*) Average y_{12} as a function of $|r_{01}|$ (green), whereas the dashed dotted line corresponds to the reference $y = 0$ line, and the blue dotted line (which is very close to the $y = 0$ line) corresponds to the linear fit of the less noisy region. $|r_{01}|$, x_{12}, and y_{12} have units of Å. Plots in (*C*) and (*D*) correspond to data at a BSA concentration of 300 g/L at $\tau = 5$ ns. To see this figure in color, go online.

anti-correlation is due to the rattling motion of the particles. The vector y_{12} (τ) acts as control since it is the component along an arbitrary vector, so it would be expected that, in the absence of cage effects, y_{12} (τ) and x_{12} (τ) exhibit similar behavior upon the increase in the magnitude of r_{01} (τ) (52,53) (Fig. 1). The $|r_{01}|$ vs. x_{12} (or y_{12}) plot is obtained by calculating the values of $|r_{01}|$ and x_{12} (or y_{12}) across all possible time origins along the length of the trajectory for a given protein and τ and combining the data for all the protein molecules (of a given species) in the simulation. The values of ($|r_{01}|$) are binned with a width of 0.05 Å and the corresponding x_{12} values are averaged. The plots are presented and discussed in Fig. 2. This process is repeated for different τ values. The approach described here is similar to that of Weiss (42), where the function used (42) to infer anti-correlation is given by

$$C_\tau(t) = \langle \frac{v_\tau(T)}{|v_\tau(T)|} \cdot \frac{v_\tau(T+t)}{|v_\tau(T+t)|} \rangle_T \quad (2)$$

where $v_\tau(t) = r(t+\tau) - r(t)$, r being the position vector. When $t = \tau$, the dot product (which is averaged across the time origins in the trajectory) is equivalent to the dot product of the unit vectors parallel to r_{01} and r_{12}, which should carry the same sign as x_{12}. In the presence of anti-correlation, $C_\tau(t) < 0$ when $t \approx \tau$, which implies that x_{12} is negative, which is consistent with the above approach.

Calculation of the α-exponent from cage effect

Weeks and Weitz have shown analytically that the slope of the $|r_{01}|$ vs. $<x_{12}>$ curve can be used to estimate the value of the α-exponent using the equation below (53):

$$\alpha_{cage}(\tau) = 1 + \frac{ln(1 + slope(\tau))}{ln(2)} \quad (3)$$

Using this approach, the α-exponent is calculated from cage effect data for a given lag time τ. The same data at different τ values is obtained by skipping the appropriate number of time frames in a simulation trajectory while calculating the displacement vectors.

RESULTS

Diffusion coefficients and sub-diffusive behavior

Time-averaged translational diffusion coefficients of CI2, LYS, and BSA were calculated from the curve of TAMSD vs. lag time(τ), averaging over all time origins and the

Please cite this article in press as: Kompella et al., What determines sub-diffusive behavior in crowded protein solutions?, Biophysical Journal (2024), https://doi.org/10.1016/j.bpj.2023.12.002

molecular species of interest. The experimental diffusion coefficient of CI2 at concentrations of 50, 100, 200, and 300 g/L of LYS and BSA had been determined previously (26). The long-time diffusion coefficients were calculated in the 0- to 1000-ns time range and the predicted diffusion coefficients of CI2 were compared with experimental values. Fig. 2 *A* shows that the predicted diffusion coefficients are of the same order of magnitude as experimental values. However, the difference in the predicted and experimental diffusion coefficients increased at higher concentrations. This could potentially be due to the lack of molecular flexibility in the simulated protein structures, which could contribute to a reduction in the tendency to form clusters. A more detailed description of the role played by such clusters is provided below. Fig. 2 *B* shows that the predicted diffusion coefficients of BSA and LYS decrease in magnitude with an increase in the concentration of the crowder, as expected.

The sub-diffusive behavior of the proteins was characterized. In solutions with a crowder concentration of 50 g/L, the α-exponent value of CI2 remained above 0.95 in the presence of both crowding proteins (Fig. 3 *A* and *C*). The same behavior was observed for the self-diffusion of the crowders, as shown in Fig. 3 *E* and *G*. The α-exponent did not exhibit pronounced variation with respect to lag time in each of the systems. The increase in the concentration of the crowder led to sub-diffusion. At a crowder concentration of 300 g/L, the value of the α-exponent decreased to 0.83 ($\sigma = 0.002$) in the range 10.4–38.8 ns for CI2 in BSA (Fig. 3 *B*), 0.87 ($\sigma = 0.002$) in the range 2.0–9.8 ns for CI2 in LYS (Fig. 3 *D*), 0.74 ($\sigma = 0.005$) in the range 8.0–39.8 ns for BSA (Fig. 3 *H*), and 0.80 ($\sigma = 0.001$) in the range 2.0–10.0 ns for LYS (Fig. 3 *F*), in all cases indicating the presence of sub-diffusive behavior. However, the observed sub-diffusion dynamics were transient and normal diffusion was gradually reached after a few hundreds of nanoseconds. In all of these cases, a clear trend can be discerned, whereby diffusion was close to normal at short timescales (less than 2 ns), sub-diffusive in the sub-microsecond timescale, and showed a reverting trend toward normal in the regime closer to 0.5 μs. This behavior is observed in all the three proteins, which are of different sizes and have a different total charge. An intermediate behavior was observed in crowder concentrations of 100 and 200 g/L (Fig. S13). Such transient sub-diffusive behavior has been predicted for γ-globulin and BSA self-crowded solutions (39).

Cage effects in the protein crowded solutions

The protein dynamics of the above-described crowded systems are consistent with sub-diffusive behavior arising due to macromolecular crowding. However, the underlying molecular mechanism by which protein crowding causes this phenomenon and its physical origins are not very well understood. The cage effect hypothesis is rigorously tested here. This hypothesis states that macromolecules in a crowded protein solution behave like colloidal particles and exhibit motion akin to rattling in a cage, termed cage effect (52), wherein they are trapped in a transient cage for a finite period of time before "hopping" to another cage. In contrast to regular Brownian motion, particles do not move freely while they are trapped in these cages. Therefore, these particles are expected to exhibit normal diffusion at very short timescales when they are not in close proximity to surrounding particles, but, at intermediate timescales, these particles would exhibit rattling dynamics before exhibiting normal Brownian motion at sufficiently long timescales. To quantitatively assess this, Doliwa and Heuer's approach (52) was used to investigate the presence of rattling-in-a-cage type of motion in our simulations. A plot of $<x_{12}>$ against $|r_{01}|$ is shown in Fig. 2 *C*, which was obtained from unfolded trajectories. It is evident from these plots that there is a clear anti-correlation between r_{01} and x_{12}. At higher values of $|r_{01}|$, the plots become noisy because there are very few particles that make very long jumps, reducing the number of data points available for analysis. There is also a higher probability for the particles that make long jumps to exit the transient cage, leading to cessation of the rattling motion (52). Fig. 2 *D* shows that, unlike x_{12}, y_{12} does not depend on the magnitude of r_{01}. These findings suggest the presence of a cage effect in crowded protein solutions. The slope of the linear section of the plot is an indicator of the strength of this cage effect. The slope calculated at different τ values in solutions with crowders at concentrations of 50 and 300 g/L is shown in Fig. 4. As expected, the slope of the tracer CI2 and protein crowders in the 50 g/L solutions was $\sim$0. In the 300-g/L solutions, the slope was initially $\sim$0 but, at intermediate timescales, the slope decreased substantially, indicating the existence of a strong cage effect, whereas, at longer timescales, the slope recovered back to $\sim$0. The x_{12} slopes calculated at intermediate timescales are significantly higher than y_{12} slopes calculated at the same timescales, indicating a pronounced cage effect as shown in Fig. 4. These observations indicate low cage effect at short timescales, followed by maximum cage effect at intermediate timescales, and restoration of low cage effect at long timescales (Fig. 4).

The predicted values of the α-exponent (calculated from the cage effect using Eq. 3) of the tracers and crowders in all the simulations (50–300 g/L of both crowders) were calculated and compared with the ones reported in the previous section, as shown in Figs. 3 and S13. The blue lines in Figs. 3 and S13 show the value of the α-exponent calculated from the log plots as described in the "calculation of the α-exponent" section, and the orange lines represent the α-exponents calculated using Eq. 3. The predicted values of the α-exponent are in good agreement with the calculated values for both crowders and tracer under all

Please cite this article in press as: Kompella et al., What determines sub-diffusive behavior in crowded protein solutions?, Biophysical Journal (2024), https://doi.org/10.1016/j.bpj.2023.12.002

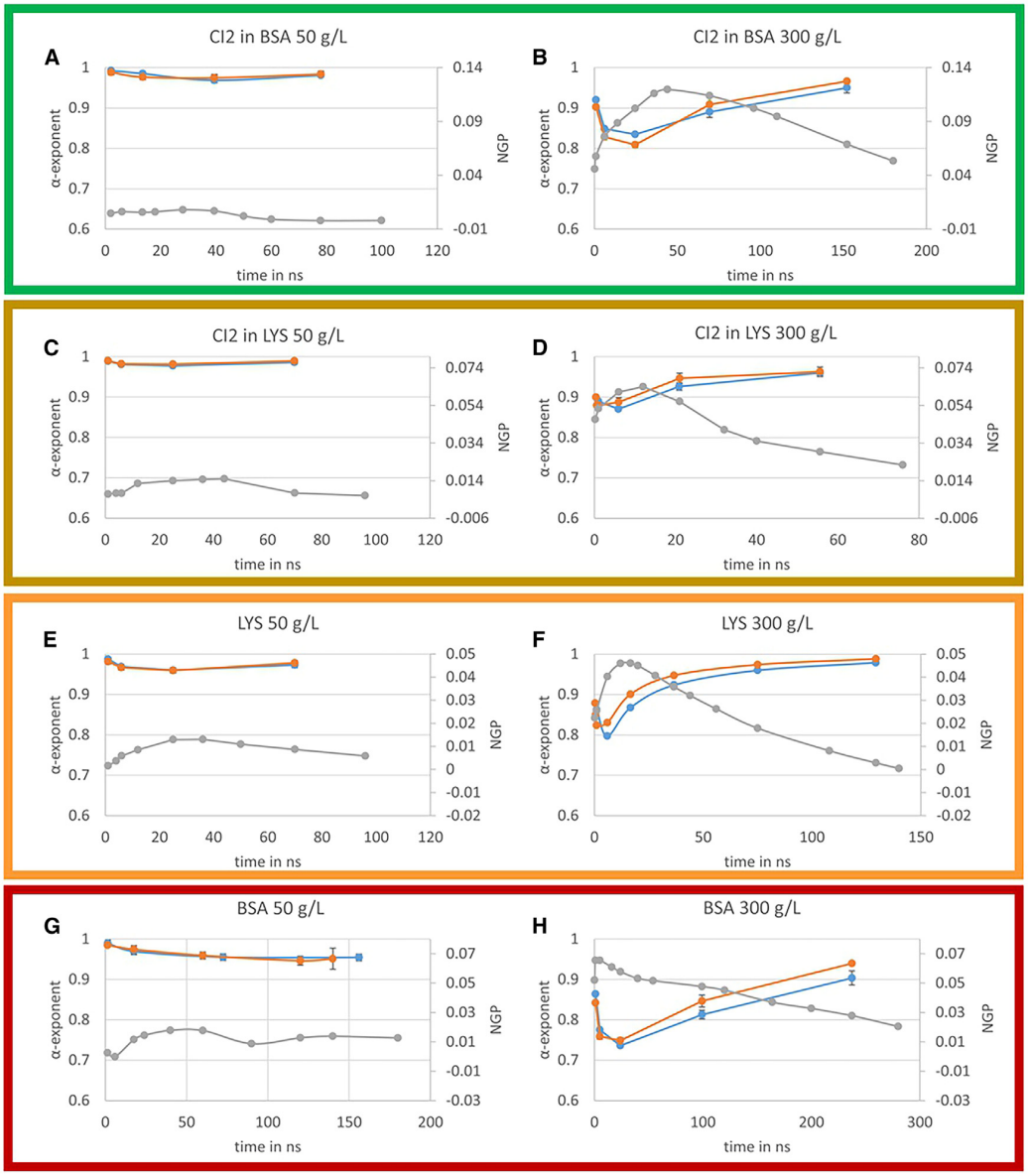

FIGURE 3 Sub-diffusive and non-Gaussianity properties of the crowders and tracer (50 and 300 g/L concentration of crowder). (*A–H*) The blue, orange, and gray lines in all the curves represent the α-exponent calculated from the log (TAMSD/τ) vs. log (τ) curves, α-exponent calculated from cage effects, and non-Gaussianity parameter (NGP) measured at different lag times, respectively. All the curves on the left side of the figure represent the data for low concentration of the crowder at 50 g/L and the ones on the right side represent data for high crowder concentration. The data for CI2 in BSA are in the first row highlighted in green, followed by data for CI2 in LYS in next row highlighted in yellow, followed by data for LYS and BSA highlighted in orange and red, respectively. Error bars represent the standard deviation of the value of the α-exponent between simulations started with different configurations. The time ranges in the individual graphs are different from each other due to the variation in the emergence of noise in the log (TAMSD/τ) vs. log (τ) curves. To see this figure in color, go online.

concentrations of the crowders at all lag times. The consistency in our predictions across different types of proteins with different sizes, net charges, and other properties is encouraging. Since sub-diffusive behavior is the manifestation of multiple mechanisms that do not necessarily constitute anti-correlated displacements (40), the fact that the computed value of the α-exponent obtained from anti-correlated displacements induced by cage effect is consistent confirms the validity of the hypothesis of cage effects causing sub-diffusive behavior in crowded protein solutions. The same approach described here was earlier used to establish cage effects using experimental data of

Please cite this article in press as: Kompella et al., What determines sub-diffusive behavior in crowded protein solutions?, Biophysical Journal (2024), https://doi.org/10.1016/j.bpj.2023.12.002

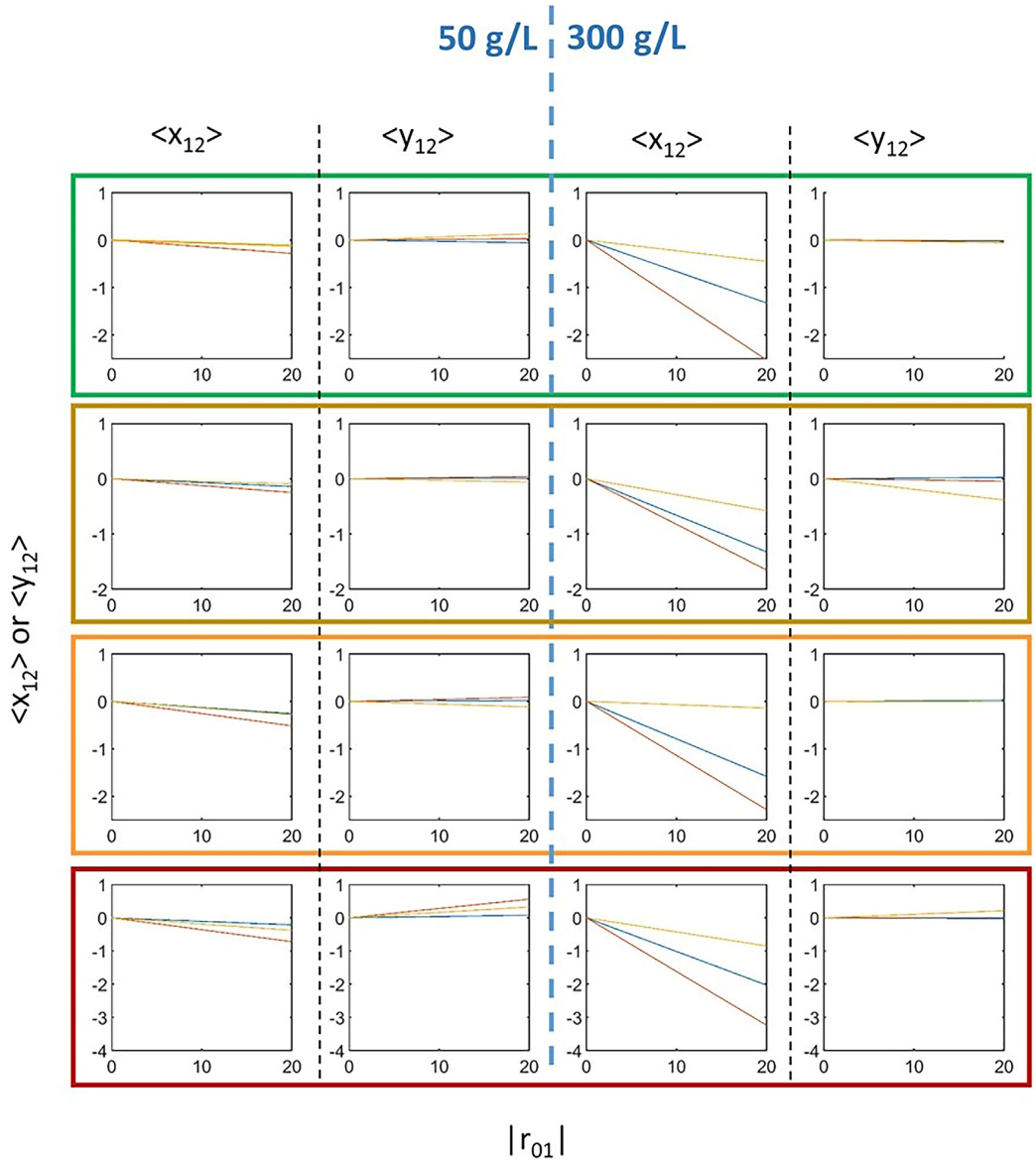

FIGURE 4 Variation of the intensity of cage effects with respect to time and crowder concentration. The straight lines plotted are representative of the slope calculated from the less noisy regions of plots of $<x_{12}>$ or $<y_{12}>$ vs. $|r_{01}|$. The blue, red, and yellow lines represent slopes at short, intermediate, and long timescales, respectively. The first and second rows highlighted in green and yellow represent the data for the diffusion of CI2 in BSA and lysozyme, respectively. The next two rows highlighted in orange and red represent the data for the self-diffusion of lysozyme and BSA, respectively. The first two columns of every row contain plots of $<x_{12}>$ vs. $|r_{01}|$ and $<y_{12}>$ vs. $|r_{01}|$ (in that order) at the low protein crowder concentration of 50 g/L. The next two columns contain the same plots at the high protein crowder concentration of 300 g/L. $|r_{01}|$, x_{12}, and y_{12} have units of Å. It should be noted that these plots are representative of slopes and the straight lines correspond to the fits made to the less noisy regions in the corresponding plots, as described in Fig. 2. The timescales corresponding to different colored lines are presented in Table S2. To see this figure in color, go online.

protein diffusion in the plasma membrane (54). However, the cage effect observed in those experiments was in the timescale of a few seconds. The observed anti-correlation of consecutive displacements is similar to the one noted in single-particle tracking experiments with dextran crowded solutions, which was explained by fBm (42).

Non-Gaussianity and ergodicity

To probe further the nature of the sub-diffusive behavior described in the previous section, we investigated the magnitude of deviations from a Gaussian distribution of displacements (Δr) by using a non-Gaussian parameter (NGP;

Please cite this article in press as: Kompella et al., What determines sub-diffusive behavior in crowded protein solutions?, Biophysical Journal (2024), https://doi.org/10.1016/j.bpj.2023.12.002

Eq. 4), in an approach similar to that of previous studies: (55).

$$\mathrm{NGP} = \frac{3\langle \Delta r^4(\tau)\rangle}{5\langle \Delta r^2(\tau)\rangle^2} - 1 \quad (4)$$

When the distribution is Gaussian, NGP $\approx$ 0, and when it deviates from Gaussian NGP is either above or below zero. The NGP of both crowders and tracers was calculated at all the concentrations and different lag times by choosing appropriate values of τ. It is clear from Figs. 3 and S13 that, at low concentrations, NGP is very low for both the crowders and tracer with no significant variation with respect to lag time. However, at the highest concentration of 300 g/L, there is a clear rise in NGP in all the cases at intermediate timescales. Non-Gaussianity is more prominent around the timescales where anomalous diffusion was identified, as described in the previous subsections. More importantly, there is a clear pattern with a slight deviation from Gaussian behavior at short timescales followed by a rise in non-Gaussianity that eventually reduces at longer timescales. Xue et al. observed increased non-Gaussianity in nanoparticles of comparable size to that of the mesh size of the polymer solution surrounding them (55). BSA molecules are larger than lysozyme molecules and, therefore, for a given concentration of the crowder, the BSA solution is expected to form larger voids compared to the lysozyme solution. Therefore, the tracer molecule CI2, which is a smaller protein than lysozyme, should exhibit a greater degree of non-Gaussianity in crowded BSA systems. In line with this argument, the maximum value of NGP for CI2 in a 300-g/L solution of BSA was predicted to be nearly twice as high as that predicted in LYS. The sudden increase in NGP observed in the case of BSA could be due to its larger size, which results in the molecule reaching the cage boundaries in a shorter time. This sudden increase in NGP also coincides with a steep decrease in the α-exponent and, in general, a negative correlation between NGP and α is noted in all the simulations. These observations point to a non-Gaussian origin of sub-diffusion, unlike fBm in the case of dextran solutions (42).

Stochastic processes such as fBm are predominantly ergodic in nature, whereas, in CTRW, deviation from ergodicity has been reported (40). While investigating the transport of insulin granules inside cells, Tabei et al. used the convergence of TAMSD, which was in turn averaged over the number of particles to infer ergodicity. The authors argued that, in an ergodic system, the average TAMSD calculated at a given lag time using simulation trajectories of different lengths should converge once sufficiently long trajectories are chosen (44). This approach mirrors the way we have assessed convergence in our simulations (Figs. S1–S4). We chose trajectories of different lengths and calculated diffusion coefficients in all these cases and, for all trajectories beyond a certain length, minimal variation in diffusion coefficients was observed. It can thus be inferred that TAMSD had converged for sufficiently long trajectories, implying ergodicity in our simulation systems.

The above findings on ergodicity, non-Gaussianity, and anti-correlation show that the behavior of our simulation systems is similar to that of fBm in finite time intervals, as reported with the numerical simulations of Guggenberger et al. (56). These authors showed that a space-confined particle, whose motion is calculated using a sub-diffusive fBm simulator, initially shows Gaussian behavior that becomes non-Gaussian at long timescales. This long-term non-Gaussianity is attributed to the presence of reflective boundaries. However, in our simulations, at longer timescales a trend pointing to recovery of Gaussianity is observed. This is due to the fact that, unlike in simulations with a strict reflective boundary, in the case of crowded solutions, a particle can cross this boundary at longer timescales and move to a different cage-like structure. Therefore, the movement of a particle at long timescales can be described as being more akin to slow Brownian motion, whereas, at intermediate timescales, non-Gaussianity manifests due to the reflective nature of cage-like structures.

Excluded volume effects

Protein molecules in crowded solutions are predicted to form dynamic/transient clusters and exhibit significantly low diffusion rate (19,39,46). The role played by attractive forces between protein molecules in regulating diffusion in timescales of the order of tens of nanoseconds has previously been reported (19). These studies indicate that the Stokes-Einstein equation is valid in crowded protein solutions, and the slow diffusion of proteins can be explained by the modified Stokes radius as a result of the formation of dynamic clusters (19). However, it is important to note that the pivotal role played by protein-protein interactions is dependent on the proteins under investigation. Furthermore, given that the timescales of dynamic cluster formation are predicted to be of the order of 1–50 ns (19), the effect of dynamic cluster formation on sub-microsecond-scale anomalous diffusion needs further investigation.

The slow diffusion, especially when the protein molecules form clusters with particularly slow diffusing partners, can potentially be modeled as trapping in CTRW (for a random amount of time), which then gives rise to anomalous diffusion, making cluster formation a possible cause of sub-diffusive behavior. On the other hand, cluster formation has been proposed as a potential hindrance to caging and, therefore, as reducing anomalous diffusive behavior (20,21). However, the role of protein shape and size in regulating sub-diffusive behavior has not been explored.

To delineate the effects of cluster formation from those arising from excluded volume, the same set of simulations as described above were conducted using only a soft-core repulsive term to remove attractive interactions between

Please cite this article in press as: Kompella et al., What determines sub-diffusive behavior in crowded protein solutions?, Biophysical Journal (2024), https://doi.org/10.1016/j.bpj.2023.12.002

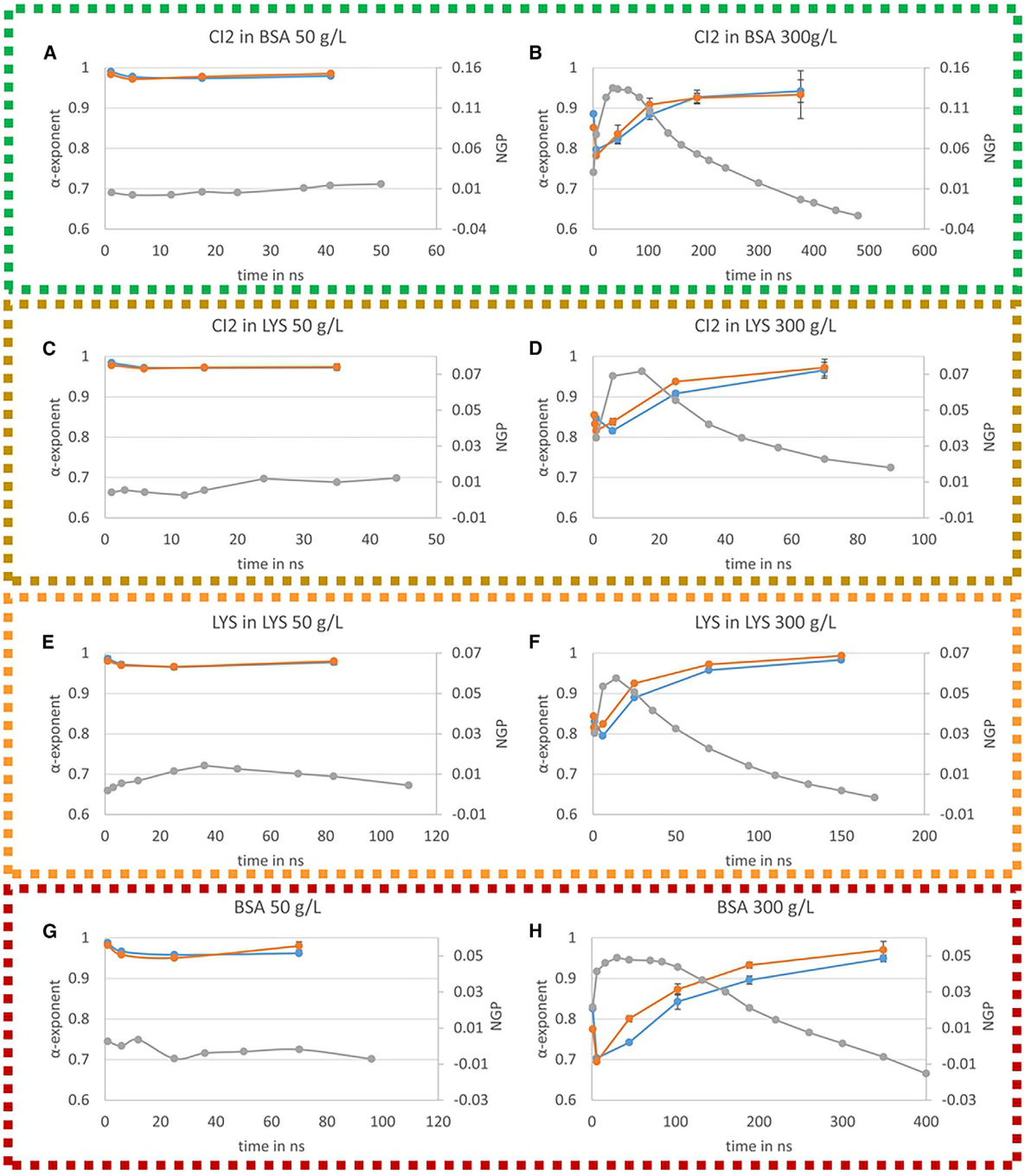

FIGURE 5 Properties of tracer and crowder in the absence of attractive interactions at concentrations of 50 and 300 g/L of the crowder. (*A–H*) The data are represented in the same way as in Fig. 3. The value of the α-exponent calculated using the log plot and cage effect, and NGP are computed for systems without attractive interactions. To see this figure in color, go online.

protein molecules. The α-exponent of crowders and tracers was calculated at all concentrations, as shown in Figs. 5 and S14. It can be seen that sub-diffusion persists despite the lack of attractive interactions. As expected, anti-correlation in the successive displacements due to cage effects is also observed in these simulations with a soft-core repulsive term. At a crowder concentration of 300 g/L, the anomalous diffusion coefficient of BSA reached a minimum value of

Please cite this article in press as: Kompella et al., What determines sub-diffusive behavior in crowded protein solutions?, Biophysical Journal (2024), https://doi.org/10.1016/j.bpj.2023.12.002

0.74 ($\sigma = 0.005$) in simulations with the full energy term, and a value of 0.70 ($\sigma = 0.0008$) in simulations with a soft-core repulsive term, whereas it had a value of 0.80 (σ is 0.001 for both the cases) in the case of LYS in both types of simulation. This is consistent with observations made by Feig and Sugita using all-atom simulations of a single CI2 molecule and eight molecules of BSA/LYS at a concentration of 100 g/L (20). In their simulations it was shown that BSA has stronger self-interactions than lysozyme does. Moreover, the dominance of monomers in LYS solutions at concentrations of less than 15% volume fraction (the maximum crowder concentration in our simulations is 13.5%) has also been shown experimentally (57). Therefore, the presence or absence of attractive forces did not significantly affect the α-exponents of LYS ($p = 0.1$). By contrast, due to the relatively stronger interactions between BSA molecules, the absence of attractive forces led to a significant drop in the value of the α-exponent ($p = 0.006$), indicating an increase in sub-diffusive behavior. In the presence of attractive forces, the value of the α-exponent of CI2 in the crowded environment of LYS was 0.87 ($\sigma = 0.002$), indicating minimal sub-diffusion behavior. However, when attractive forces were turned off, the value of the α-exponent reduced ($p = 0.002$) to 0.82 ($\sigma = 0.006$). With BSA as a crowder, the value of the α-exponent reduced ($p = 3 \times 10^{-5}$) from 0.83 ($\sigma = 0.002$) to 0.80 ($\sigma = 0.002$) when attractive forces were turned off. These observations are also consistent with the findings of Feig and Sugita (20), since CI2 interacts more strongly with LYS compared with BSA, and hence there is a larger effect on the value of the α-exponent when attractive forces are turned off. In addition, the value of the α-exponent of BSA was 0.70 and that of LYS was 0.80 in the absence of attractive forces (the statistical significance of this difference was measured using a t-test, $p = 1.5 \times 10^{-6}$). This suggests that cage effects vary between protein species even though neither of them forms clusters. The more pronounced sub-diffusion dynamics in BSA in the absence of attractive forces might be due to its larger size. Since large-sized crowders can create larger voids in the solution, the probability of protein localization is thus higher. This suggests that the extent of cage effects depends not only on the strength of protein-protein interactions but also on the size of the crowders (Table S3). Consequently, in systems with the full energy term, overall cage effects are likely to be a function of the basal cage effect (observed in the absence of attractive forces) and the strength of protein-protein interactions. Therefore, cage effects and sub-diffusion dynamics are specific to the crowders and tracers present. It is important to emphasize that the maximum cage effect in a given system is observed in the absence of attractive forces. Therefore, sub-diffusion beyond what is predicted from the maximum cage effect must arise from other phenomena. The more pronounced non-Gaussianity observed in the case of CI2 in BSA and LYS, compared with simulations with the full energy term, could be explained by an increase in excluded volume effects in the absence of attractive forces in the system, which can be inferred from the difference in the BSA-BSA and LYS-LYS radial distribution functions (at a concentration of 300 g/L) in the presence and absence of attractive interactions, as shown in Fig. 6.

The experimental diffusion coefficients of CI2 in the lower concentration (50 g/L) of BSA and LYS were 13.22 and 12.2 cm^2/s, respectively, which decreased to 2.66 and 1.39 cm^2/s, respectively, at the higher concentration of the crowders (300 g/L), resulting in a ~fivefold and ~ninefold decrease in diffusion rate in BSA and LYS, respectively. In our simulations without attractive forces, we noted a ~threefold and ~twofold decrease in the diffusion coefficient of CI2 in BSA and LYS environments, respectively (when going from a crowder concentration of 50 g/L to 300 g/L). These findings suggest that the contribution of excluded volume effects toward decreased microsecond-scale diffusion in crowded protein solutions is not insignificant.

DISCUSSION

Our findings suggest that sub-diffusive behavior is present in crowded protein solutions and the extent of it depends on the nature of the proteins under consideration. For a given protein solution with a certain crowder species at a given concentration, sub-diffusion dynamics mediated by cage effects have a maximum limit. This limit is a function of the proteins under consideration and, therefore, any sub-diffusion dynamics stronger than this limit would be the result of phenomena other than caging, such as nonspecific interactions mediating sub-diffusion and explained using CTRW models (58). However, it is evident from the use of a soft-core repulsive energy term only that such nonspecific interactions do not play a role in the sub-diffusion observed in our systems, reinforcing the role of cage effects. In our simulations, the α-exponent of BSA decreased in the absence of attractive interactions, whereas that of LYS remained the same. On the other hand, McGuffee and Elcock (18) showed that the α-exponent (of some cytoplasmic proteins) increased in the absence of attractive interactions. These observations indicate that sub-diffusive behavior is highly specific to the protein of interest and its crowded environment. As explained above in the case of a CTRW model, it is possible that the underlying sub-diffusive process in the cytoplasmic proteins (noted by McGuffee and Elcock) relied on attractive interactions.

Recently, it has been shown mathematically that extreme first-passage time, the minimum time taken by a searcher in a group of searchers to reach a target, is lower in the case of sub-diffusive searchers compared with normally diffusing counterparts (59). This suggests that sub-diffusive behavior has a vital role to play in biological systems, where molecular encounters drive cellular processes. The

Please cite this article in press as: Kompella et al., What determines sub-diffusive behavior in crowded protein solutions?, Biophysical Journal (2024), https://doi.org/10.1016/j.bpj.2023.12.002

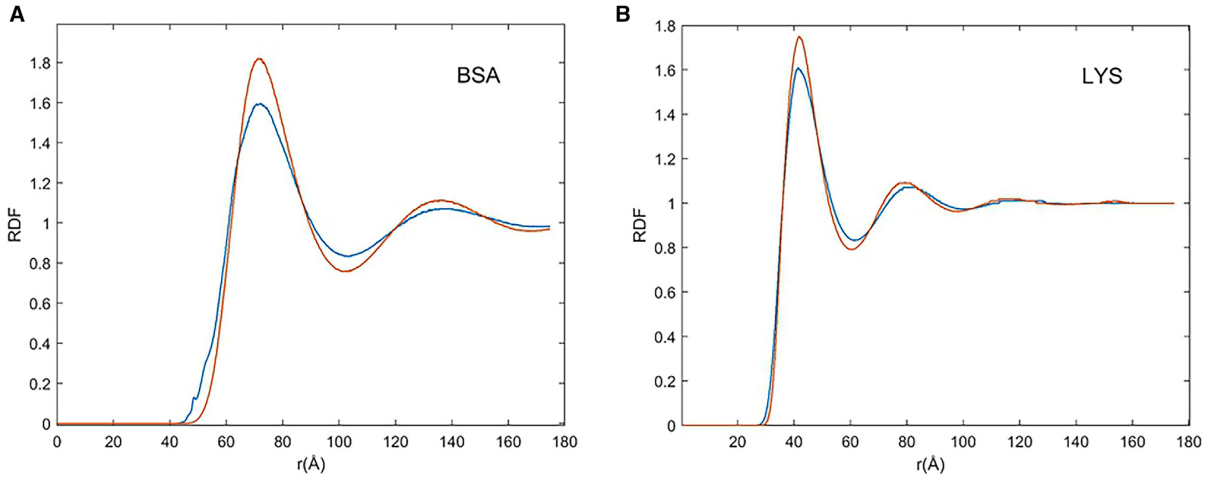

FIGURE 6 BSA-BSA (*A*) and LYS-LYS (*B*) radial distribution functions obtained from simulations at a concentration of 300 g/L using a soft-core repulsive term only (*red curves*) and the full energy term (*blue curves*). The effective radius was approximated as the maximum distance (r) at which radial distribution function ≈ 0, increasing by 2.8 Å and 1.6 Å in BSA and LYS, respectively, when only the soft-core repulsive term is used. Due to the larger size of BSA compared with LYS, the change in excluded volume due to the small change in the effective radius is more pronounced in the former. To see this figure in color, go online.

implications of cage effects and the subsequent sub-diffusive phenomenon are important in the context of diffusion-limited reactions. Normal diffusion is the underlying assumption made in the derivation of rate constants of diffusion-limited reactions. However, since deviations from normal diffusion are apparent and with varying intensity depending on the protein species and timescales investigated, it is important to account for such deviations using approaches such as Haugh's (60), especially in the framework of treating biological reaction networks as complex systems. Combining the fact that protein crowded systems emulate the cellular environment (26) and our findings indicating that the strength of sub-diffusion dynamics is a result of such crowding, in light of the above mathematical findings, it is possible to infer that cells should maintain crowding for optimal execution of the cellular processes. In fact, such a mechanism has already been proposed by Van Den Berg et al. and is termed homeocrowding (61).

The sub-diffusive behavior in our systems exhibited features of fBm. However, a more rigorous numerical approach is necessary to establish whether there is fBm in crowded protein systems. It would be interesting to use soft reflective walls that allow particles to escape the confined space to explain the restoration of normal diffusion with Gaussian behavior over long timescales.

Based on the simulations without attractive interactions, we have shown that excluded volume effects play a role in decreased diffusion. The diffusion coefficients predicted from the simulations are of the same order of magnitude as experimental values. However, due to the lack of precise agreement with experimental observations, further investigation is necessary to delineate the role of excluded volume effects and cluster formation.

CONCLUSIONS

The predictions of our Brownian dynamics simulation study clearly show that sub-diffusion in crowded protein solutions arises from cage effects. Moreover, deviations in the distributions of molecular displacements from that of Gaussian distribution is associated with the transient sub-diffusion. Based on our findings, it is clear that the sub-diffusive behavior in crowded protein solutions can be explained by volume exclusion. Since the sub-microsecond-scale anomalous diffusion observed is dependent on the properties of the proteins (i.e., surface properties such as charge, size, and shape), it is important to carefully account for the composition of the cytoplasmic protein and nucleic acid species when investigating the diffusive behavior of macromolecules in cell-like environments in these timescales.

DATA AND CODE AVAILABILITY

The data associated with this work, which include the input files, MSD data, analysis scripts, and other relevant files, are all publicly available on Zenodo (https://doi.org/10.5281/zenodo.8412942).

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2023.12.002.

AUTHOR CONTRIBUTIONS

V.P.S.K. contributed to the design of research and methodology, analysis, code development, visualization, and writing – original draft, and played a supporting role in resource acquisition. R.L.M., M.C.R., and I.S.

Please cite this article in press as: Kompella et al., What determines sub-diffusive behavior in crowded protein solutions?, Biophysical Journal (2024), https://doi.org/10.1016/j.bpj.2023.12.002

contributed to the design of research, funding and resource acquisition, critiquing the research, and manuscript preparation.

ACKNOWLEDGMENTS

This work used the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk), access to which was provided by the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim (https://www.hecbiosim.ac.uk/), supported by EPSRC (grant no. EP/R029407/1). Analysis and visualization of the simulation data were conducted at the Pawsey Supercomputing Centre, therefore this work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia, as well as resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government. V.K. gratefully acknowledges the receipt of a scholarship under the Aberdeen-Curtin Alliance collaborative PhD program.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPORTING CITATIONS

Reference (62) appears in the supporting material.

REFERENCES

1. Kompella, V. P. S., I. Stansfield, …, R. L. Mancera. 2019. Definition of the Minimal Contents for the Molecular Simulation of the Yeast Cytoplasm. *Front. Mol. Biosci.* 6:97.
2. Ellis, R. J., and A. P. Minton. 2003. Join the crowd. *Nature.* 425:27–28.
3. Tabaka, M., T. Kalwarczyk, …, R. Holyst. 2014. The effect of macromolecular crowding on mobility of biomolecules, association kinetics, and gene expression in living cells. *Front. Physiol.* 2:1–14.
4. Nakano, S. I., D. Miyoshi, and N. Sugimoto. 2014. Effects of molecular crowding on the structures, interactions, and functions of nucleic acids. *Chem. Rev.* 114:2733–2758.
5. Klumpp, S., M. Scott, …, T. Hwa. 2013. Molecular crowding limits translation and cell growth. *Proc. Natl. Acad. Sci. USA.* 110:16754–16759.
6. Miermont, A., F. Waharte, …, P. Hersen. 2013. Severe osmotic compression triggers a slowdown of intracellular signaling, which can be explained by molecular crowding. *Proc. Natl. Acad. Sci. USA.* 110:5725–5730.
7. Nettesheim, G., I. Nabti, …, G. T. Shubeita. 2020. Macromolecular crowding acts as a physical regulator of intracellular transport. *Nat. Phys.* 16:1144–1151.
8. Satyam, A., P. Kumar, …, D. Zeugolis. 2014. Macromolecular crowding meets tissue engineering by self-assembly: A paradigm shift in regenerative medicine. *Adv. Mater.* 26:3024–3034.
9. Feig, M., I. Yu, …, Y. Sugita. 2017. Crowding in Cellular Environments at an Atomistic Level from Computer Simulations. *J. Phys. Chem. B.* 121:8009–8025.
10. Harada, R., Y. Sugita, and M. Feig. 2012. Protein crowding affects hydration structure and dynamics. *J. Am. Chem. Soc.* 134:4842–4849.
11. Asami, K., T. Hanai, and N. Koizumi. 1976. Dielectric properties of yeast cells. *J. Membr. Biol.* 28:169–180.
12. Tanizaki, S., J. Clifford, …, M. Feig. 2008. Conformational sampling of peptides in cellular environments. *Biophys. J.* 94:747–759.
13. Senske, M., L. Törk, …, S. Ebbinghaus. 2014. Protein stabilization by macromolecular crowding through enthalpy rather than entropy. *J. Am. Chem. Soc.* 136:9036–9041.
14. Wilcox, A. E., M. A. LoConte, and K. M. Slade. 2016. Effects of macromolecular crowding on alcohol dehydrogenase activity are substrate-dependent. *Biochemistry.* 55:3550–3558.
15. Nasreen, K., Z. A. Parray, …, A. Islam. 2020. Interactions under crowding milieu: Chemical-induced denaturation of myoglobin is determined by the extent of heme dissociation on interaction with crowders. *Biomolecules.* 10, 490.
16. Assunção, M., C. W. Wong, …, A. Blocki. 2020. Macromolecular dextran sulfate facilitates extracellular matrix deposition by electrostatic interaction independent from a macromolecular crowding effect. *Mater. Sci. Eng., C.* 106, 110280.
17. Qing, J., A. Chen, and N. Zhao. 2018. Quantifying the protein-protein association rate in polymer solutions: Crowding-induced diffusion and energy modifications. *Phys. Chem. Chem. Phys.* 20:27937–27948.
18. McGuffee, S. R., and A. H. Elcock. 2010. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput. Biol.* 6, e1000694.
19. von Bülow, S., M. Siggel, …, G. Hummer. 2019. Dynamic cluster formation determines viscosity and diffusion in dense protein solutions. *Proc. Natl. Acad. Sci. USA.* 116:9843–9852.
20. Feig, M., and Y. Sugita. 2012. Variable interactions between protein crowders and biomolecular solutes are important in understanding cellular crowding. *J. Phys. Chem. B.* 116:599–605.
21. Nawrocki, G., P. H. Wang, …, M. Feig. 2017. Slow-Down in Diffusion in Crowded Protein Solutions Correlates with Transient Cluster Formation. *J. Phys. Chem. B.* 121:11072–11084.
22. Swaminathan, R., C. P. Hoang, and A. S. Verkman. 1997. Photobleaching recovery and anisotropy decay of green fluorescent protein GFP-S65T in solution and cells: Cytoplasmic viscosity probed by green fluorescent protein translational and rotational diffusion. *Biophys. J.* 72:1900–1907.
23. Costantini, L., and E. Snapp. 2013. Probing endoplasmic reticulum dynamics using fluorescence imaging and photobleaching techniques. *Curr. Protoc. Cell Biol.* 60:21.7.1–21.7.29.
24. González-González, I. M., F. Jaskolski, …, J. M. Henley. 2012. Measuring membrane protein dynamics in neurons using fluorescence recovery after photobleach. *Methods Enzymol.* 504:127–146.
25. Plochowietz, A., I. Farrell, …, A. N. Kapanidis. 2017. In vivo single-RNA tracking shows that most tRNA diffuses freely in live bacteria. *Nucleic Acids Res.* 45:926–937.
26. Wang, Y., C. Li, and G. J. Pielak. 2010. Effects of proteins on protein diffusion. *J. Am. Chem. Soc.* 132:9392–9397.
27. Skóra, T., F. Vaghefikia, …, S. Kondrat. 2020. Macromolecular Crowding: How Shape and Interactions Affect Diffusion. *J. Phys. Chem. B.* 124:7537–7543.
28. Biswas, S., J. Kundu, …, P. K. Chowdhury. 2018. Mixed Macromolecular Crowding: A Protein and Solvent Perspective. *ACS Omega.* 3:4316–4330.
29. Ando, T., and J. Skolnick. 2010. Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. *Proc. Natl. Acad. Sci. USA.* 107:18457–18462.
30. Yu, I., T. Mori, …, M. Feig. 2016. Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *Elife.* 5, e19274.
31. Einstein, A. 1956. Investigations on the Theory of the Brownian Movement. Dover.
32. Daddysman, M. K., and C. J. Fecko. 2013. Revisiting point FRAP to quantitatively characterize anomalous diffusion in live cells. *J. Phys. Chem. B.* 117:1241–1251.
33. Weiss, M., M. Elsner, …, T. Nilsson. 2004. Anomalous Subdiffusion Is a Measure for Cytoplasmic Crowding in Living Cells. *Biophys. J.* 87:3518–3524.

Please cite this article in press as: Kompella et al., What determines sub-diffusive behavior in crowded protein solutions?, Biophysical Journal (2024), https://doi.org/10.1016/j.bpj.2023.12.002

34. Di Rienzo, C., V. Piazza, …, F. Cardarelli. 2014. Probing short-range protein Brownian motion in the cytoplasm of living cells. *Nat. Commun.* 5:5891.

35. Hasnain, S., C. L. McClendon, …, P. Bandyopadhyay. 2014. A new coarse-grained model for E. coli cytoplasm: Accurate calculation of the diffusion coefficient of proteins and observation of anomalous diffusion. *PLoS One.* 9, e106466.

36. Jeon, J. H., H. M. S. Monne, …, R. Metzler. 2012. Anomalous diffusion of phospholipids and cholesterols in a lipid bilayer and its origins. *Phys. Rev. Lett.* 109, 188103.

37. Jeon, J. H., M. Javanainen, …, I. Vattulainen. 2016. Protein crowding in lipid bilayers gives rise to non-Gaussian anomalous lateral diffusion of phospholipids and proteins. *Phys. Rev. X.* 6, 021006.

38. Stachura, S. S., C. J. Malajczuk, …, R. L. Mancera. 2019. Influence of Bilayer Size and Number in Multi-Bilayer DOPC Simulations at Full and Low Hydration. *Langmuir.* 35:2399–2411.

39. Balbo, J., P. Mereghetti, …, R. C. Wade. 2013. The shape of protein crowders is a major determinant of protein diffusion. *Biophys. J.* 104:1576–1584.

40. Metzler, R., J. H. Jeon, …, E. Barkai. 2014. Anomalous diffusion models and their properties: Non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Phys. Chem. Chem. Phys.* 16:24128–24164.

41. Bouchaud, J. P., and A. Georges. 1990. Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications. *Phys. Rep.* 195:127–293.

42. Weiss, M. 2013. Single-particle tracking data reveal anticorrelated fractional Brownian motion in crowded fluids. *Phys. Rev. E - Stat. Nonlinear Soft Matter Phys.* 88, 010101.

43. Weigel, A. V., B. Simon, …, D. Krapf. 2011. Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking. *Proc. Natl. Acad. Sci. USA.* 108:6438–6443.

44. Tabei, S. M. A., S. Burov, …, N. F. Scherer. 2013. Intracellular transport of insulin granules is a subordinated random walk. *Proc. Natl. Acad. Sci. USA.* 110:4911–4916.

45. Liu, Y., L. Porcar, …, P. Baglioni. 2011. Lysozyme Protein Solution with an Intermediate Range Order Structure. *J. Phys. Chem. B.* 115:7238–7247.

46. Mereghetti, P., and R. C. Wade. 2012. Atomic detail brownian dynamics simulations of concentrated protein solutions with a mean field treatment of hydrodynamic interactions. *J. Phys. Chem. B.* 116:8523–8533.

47. Martinez, M., N. J. Bruce, …, R. C. Wade. 2015. SDA 7: A modular and parallel implementation of the simulation of diffusional association software. *J. Comput. Chem.* 36:1631–1645.

48. Yu, X., M. Martinez, …, R. C. Wade. 2015. WebSDA: A web server to simulate macromolecular diffusional association. *Nucleic Acids Res.* 43:W220–W224.

49. Wang, J., P. Cieplak, and P. A. Kollman. 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* 21:1049–1074.

50. Jurrus, E., D. Engel, …, N. A. Baker. 2018. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* 27:112–128.

51. Allen, M. P., and D. J. Tildesley. 2017. Computer Simulation of Liquids, Second. Oxford University Press.

52. Doliwa, B., and A. Heuer. 1998. Cage effect, local anisotropies, and dynamic heterogeneities at the glass transition: A computer study of hard spheres. *Phys. Rev. Lett.* 80:4915–4918.

53. Weeks, E. R., and D. A. Weitz. 2002. Subdiffusion and the cage effect studied near the colloidal glass transition. *Chem. Phys.* 284:361–367.

54. Goiko, M., J. R. De Bruyn, and B. Heit. 2016. Short-Lived Cages Restrict Protein Diffusion in the Plasma Membrane. *Sci. Rep.* 6:34987.

55. Xue, C., X. Zheng, …, G. Hu. 2016. Probing Non-Gaussianity in Confined Diffusion of Nanoparticles. *J. Phys. Chem. Lett.* 7:514–519.

56. Guggenberger, T., G. Pagnini, …, R. Metzler. 2019. Fractional Brownian motion in a finite interval: Correlations effect depletion or accretion zones of particles near boundaries. *New J. Phys.* 21, 022002.

57. Porcar, L., P. Falus, …, Y. Liu. 2010. Formation of the dynamic clusters in concentrated Lysozyme protein solutions. *J. Phys. Chem. Lett.* 1:126–129.

58. Etoc, F., E. Balloul, …, M. Coppey. 2018. Non-specific interactions govern cytosolic diffusion of nanosized objects in mammalian cells. *Nat. Mater.* 17:740–746.

59. Lawley, S. D. 2020. Extreme statistics of anomalous subdiffusion following a fractional Fokker-Planck equation: Subdiffusion is faster than normal diffusion. *J. Phys. Math. Theor.* 53:385005.

60. Haugh, J. M. 2009. Analysis of reaction-diffusion systems with anomalous subdiffusion. *Biophys. J.* 97:435–442.

61. Van Den Berg, J., A. J. Boersma, and B. Poolman. 2017. Microorganisms maintain crowding homeostasis. *Nat. Rev. Microbiol.* 15:309–318.

62. Backlund, M. P., R. Joyner, and W. E. Moerner. 2015. Chromosomal locus tracking with proper accounting of static and dynamic errors. *Phys. Rev. E.* 91, 062716.

Biophysical Journal, Volume *123*

Supplemental information

What determines sub-diffusive behavior in crowded protein solutions?

Vijay Phanindra Srikanth Kompella, Maria Carmen Romano, Ian Stansfield, and Ricardo L. Mancera

Supplementary material

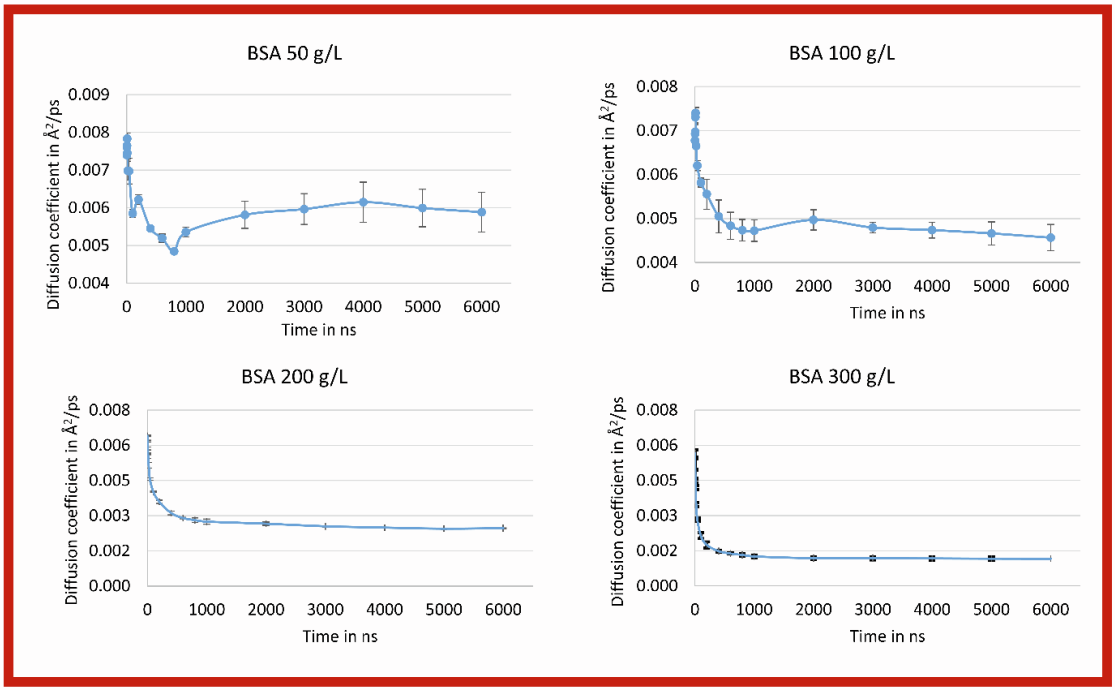

Figure S1. Convergence of the diffusion coefficient of BSA at different concentrations in simulations with the full energy term. The error bars represent the standard deviation (n=3).

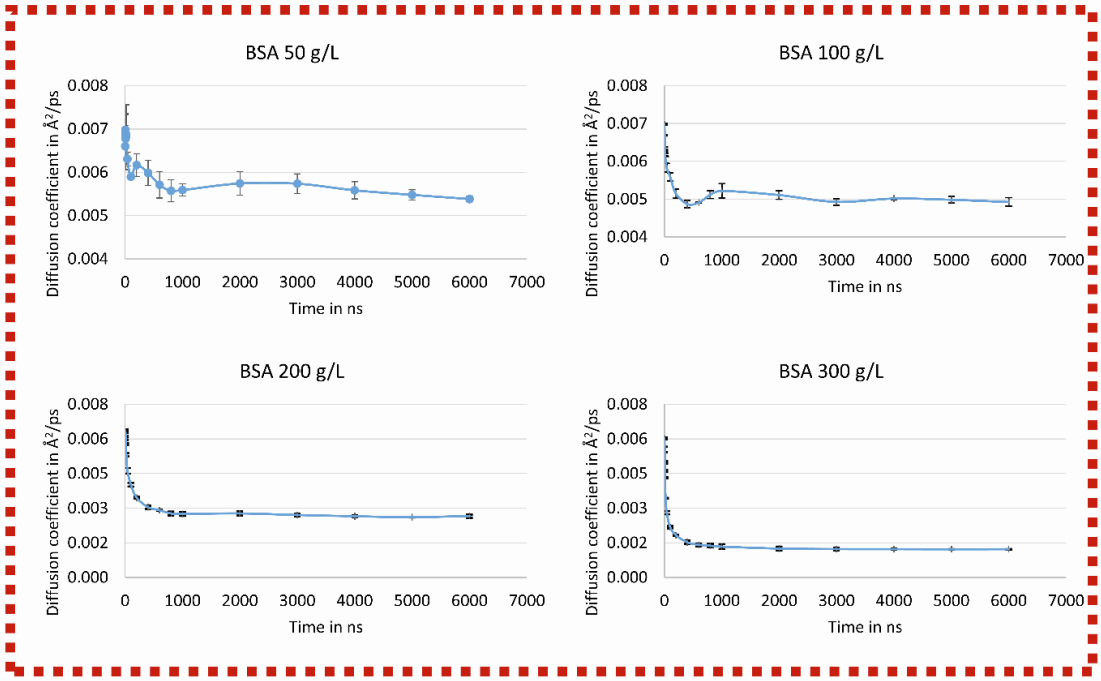

Figure S2. Convergence of the diffusion coefficient of BSA at different concentrations in simulations with the soft-core repulsive term only. The error bars represent the standard deviation (n=3).

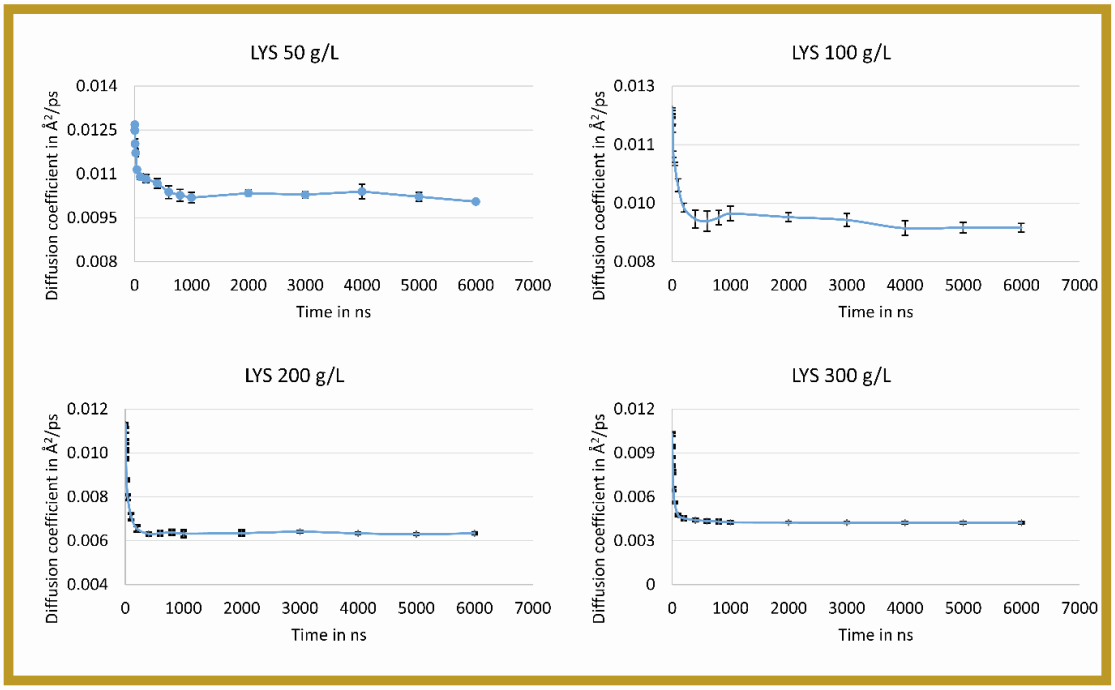

Figure S3. Convergence of the diffusion coefficient of lysozyme at different concentrations in simulations with the full energy term. The error bars represent the standard deviation (n=3).

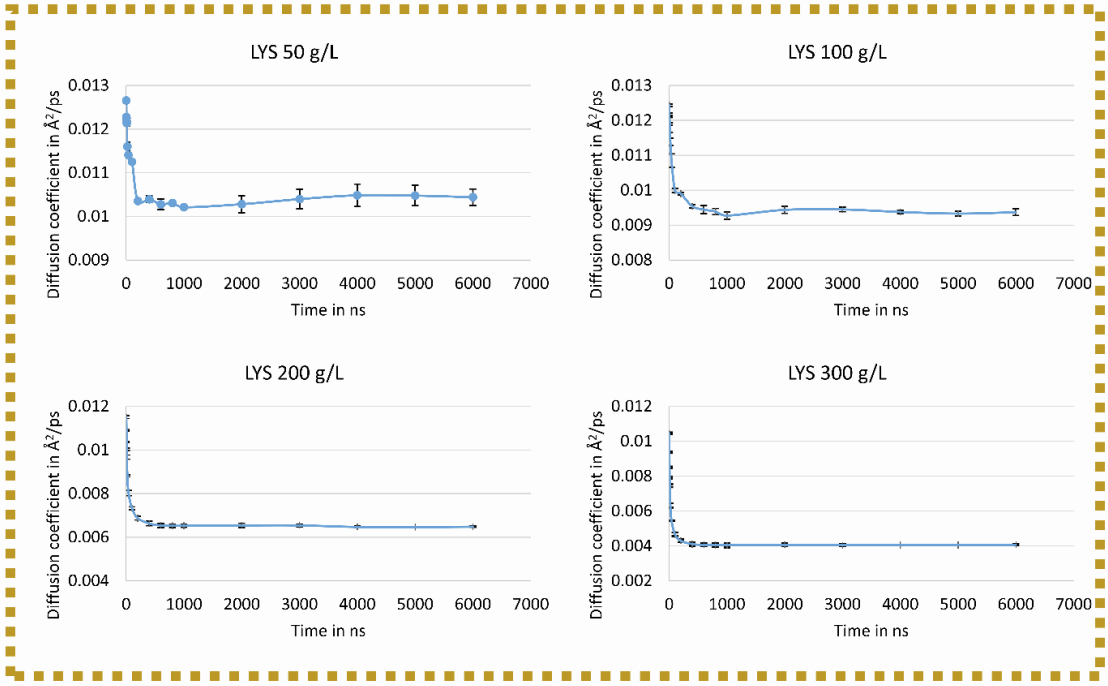

Figure S4. Convergence of the diffusion coefficient of lysozyme at different concentrations in simulations with the soft-core repulsive term only. The error bars represent the standard deviation (n=3).

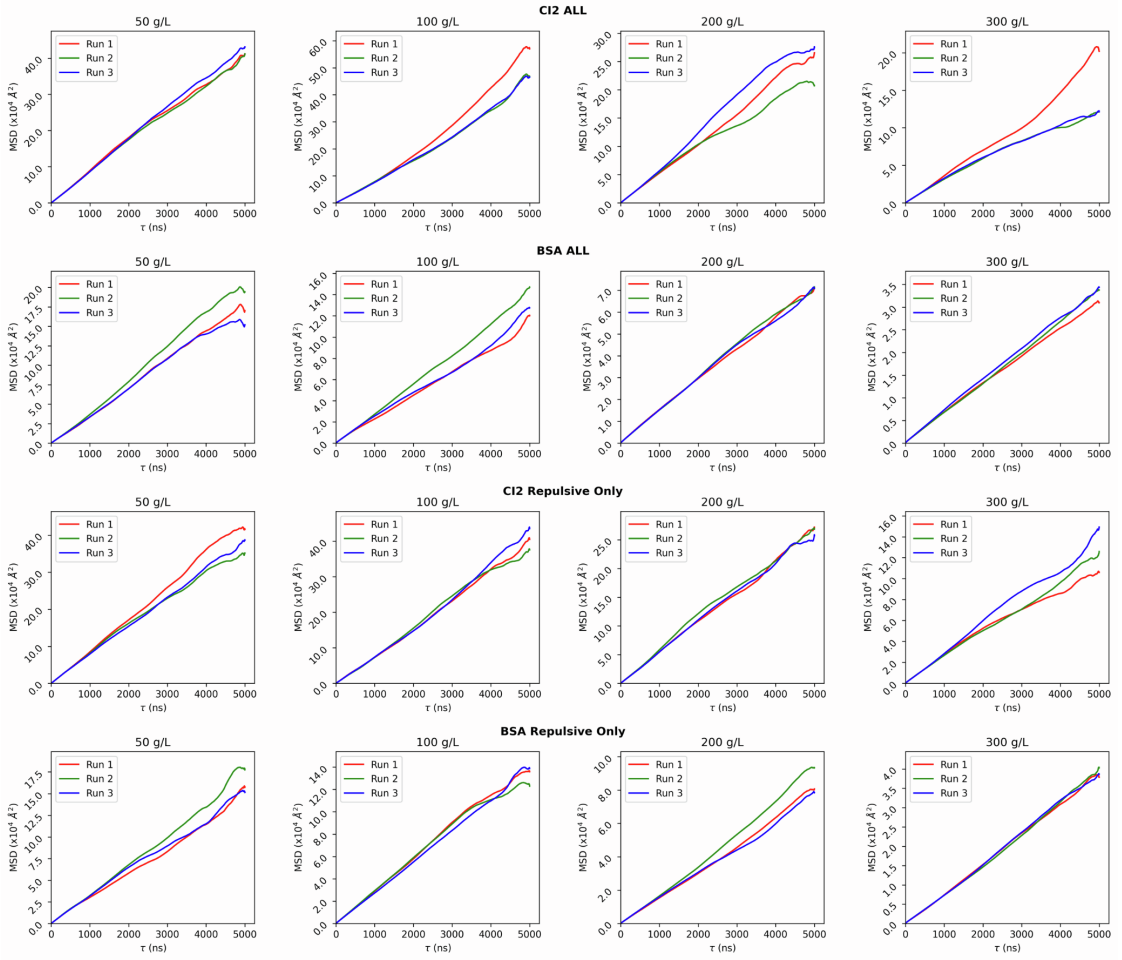

Figure S5. Time averaged mean-squared displacement (TAMSD) vs lag time plots of BSA (crowder) and CI2 in the presence of full energy term (ALL) and only soft-core repulsive (repulsive only) term.

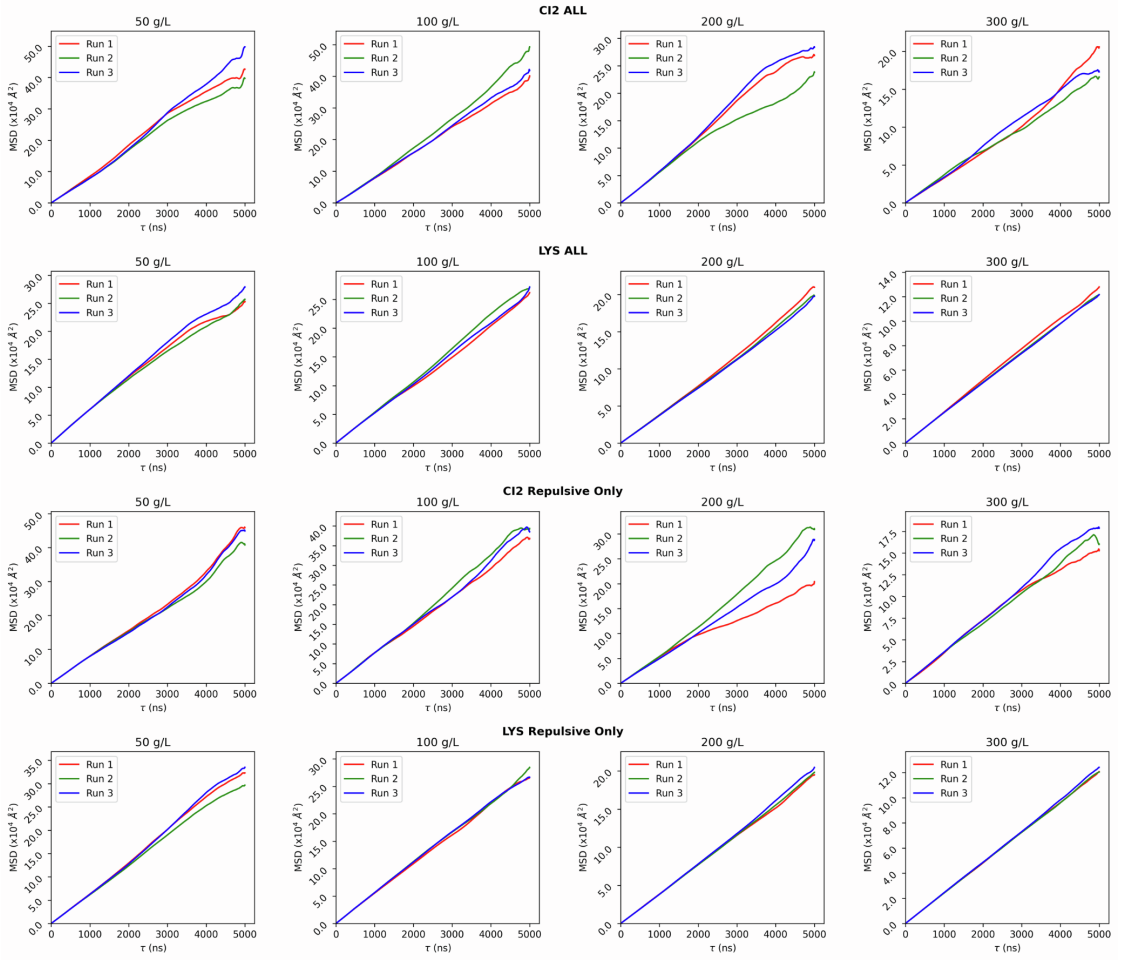

Figure S6. Time averaged mean-squared displacement (TAMSD) vs lag time plots of lysozyme (crowder) and CI2 in the presence of full energy term (ALL) and only soft-core repulsive (repulsive only) term.

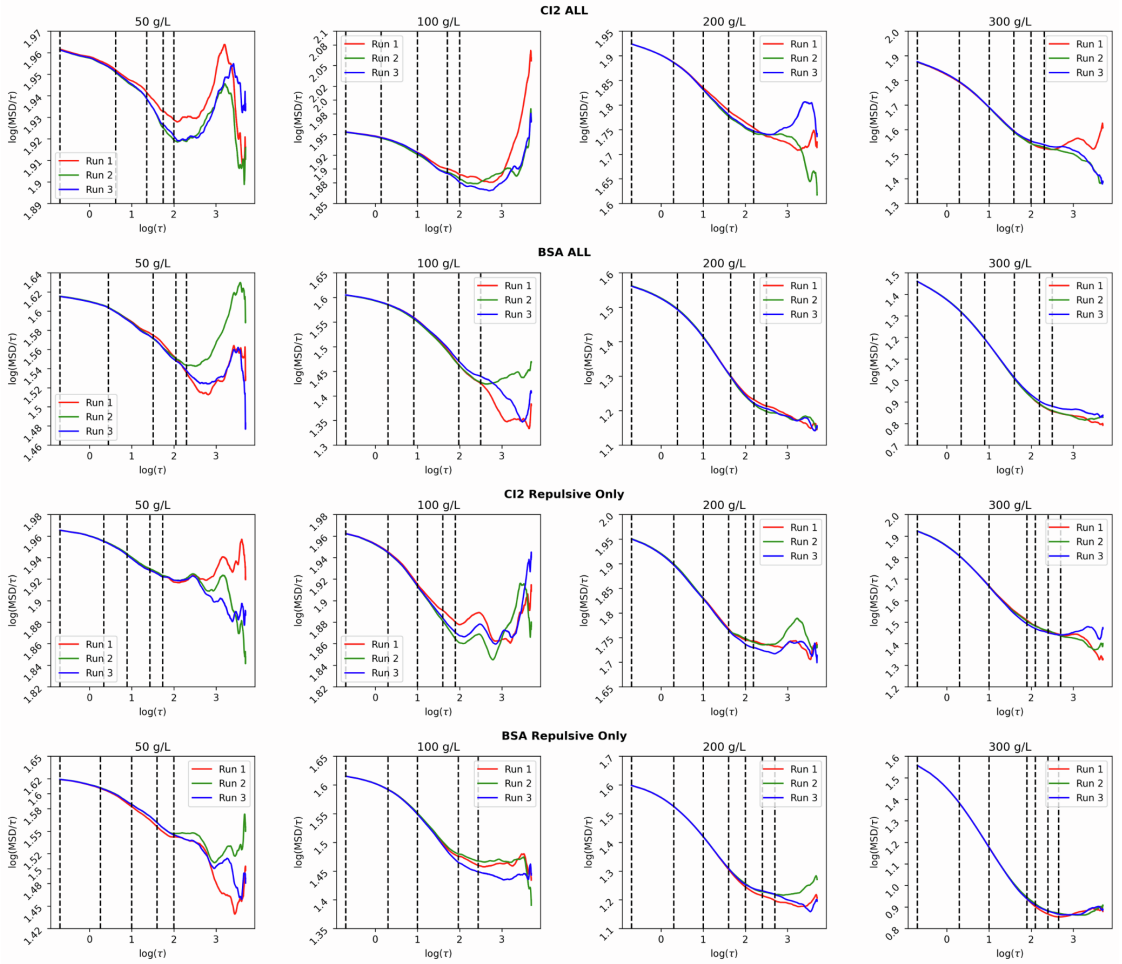

Figure S7. The log (TAMSD/τ) vs log (τ) plots of CI2 and BSA (crowder) under the conditions of full energy term (ALL) and only soft-core repulsive forces (repulsive only). The vertical dotted lines represent the time ranges used in the calculation of α-exponent.

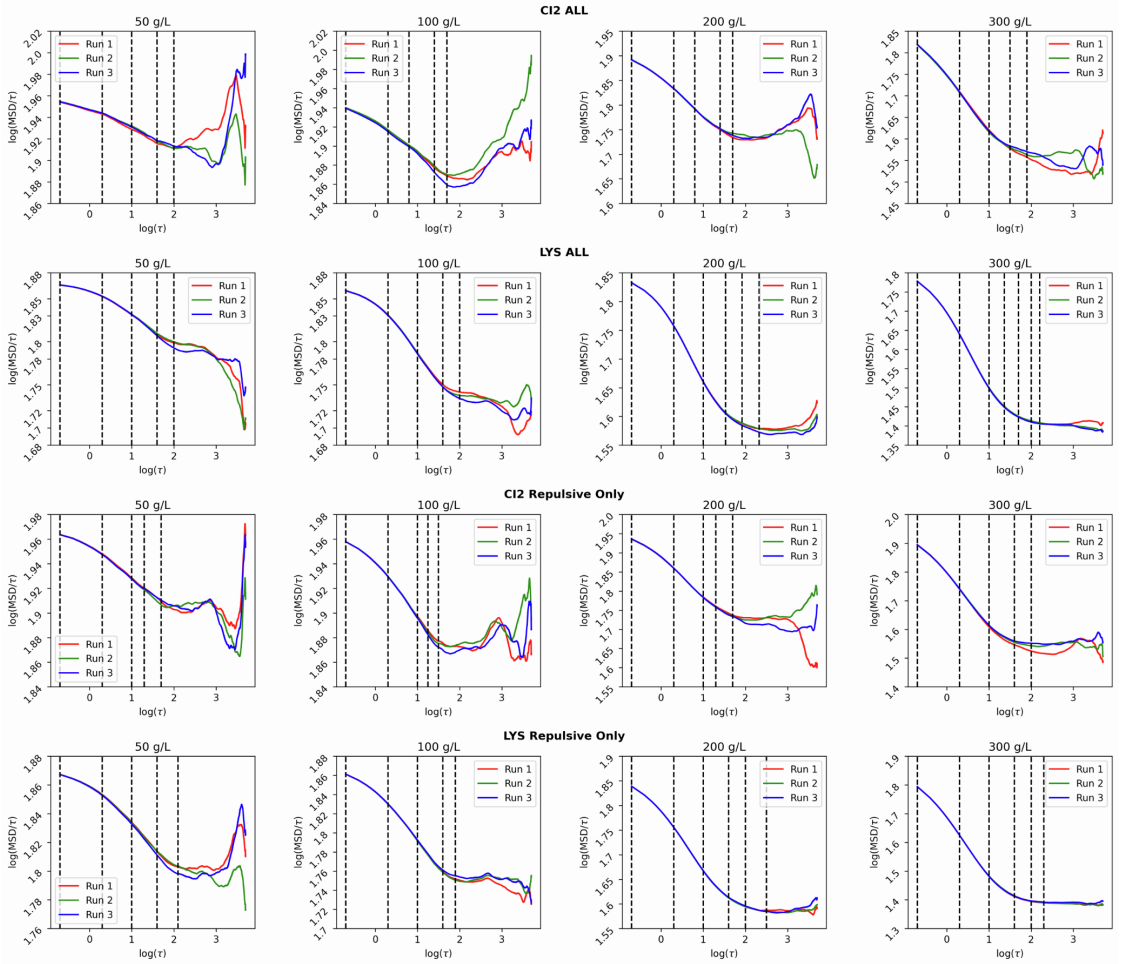

Figure S8. The log (TAMSD/τ) vs log (τ) plots of CI2 and lysozyme (crowder) under the conditions of full energy term (ALL) and only soft-core repulsive forces (repulsive only). The vertical dotted lines represent the time ranges used in the calculation of α-exponent. Fitting the power law function to the MSD vs lag time plots

Fitting of a power law function to the MSD v lag time plots

Sub-diffusion was estimated using the α-exponent calculated from the log-log plots as described in the Methods section in the main text. However, a static offset in the MSD can result in distorted log-log plots, which incorrectly implies the presence of sub-diffusive behaviour. A power law function was directly fit to the MSD vs lag time (τ) plots to calculate the α-exponent, without invoking log-log plots, to verify the validity of the inferences drawn from log-log plots. Using an approach similar to that of Backlund et al., (62) the MSD plots were fitted to equation 1 in the main text, with the two fit parameters being the α-exponent and the generalized diffusion coefficient. The fits were performed using the curve_fit function in the scipy library of Python, which uses the Levenberg-Marquardt algorithm.

Backlund et al. also account for static offset arising from zero mean Gaussian localization error (due to shot noise) by fitting the curves to the equation:

$$< x^2(t) > = 6Dt^{\alpha} + c \qquad \text{Equation S1,}$$

with 'c' being the static offset (62). Although the physical origins of the static offset in the simulations are different from that in the experiments, mathematically the offsets are equivalent. Therefore, we accounted for the static offset by following this approach using three parameter fits, with 'c' as the third parameter.

Both types of fits were performed by choosing data ranges in an incremental manner, i.e. the fit at any given lag time point, 'τ', was performed by using all the data points up to τ. In both types of fit, we noted sub-diffusive behaviour similar to that of our calculations using log-log plots, confirming the previously drawn conclusions (Figures S9-S12).

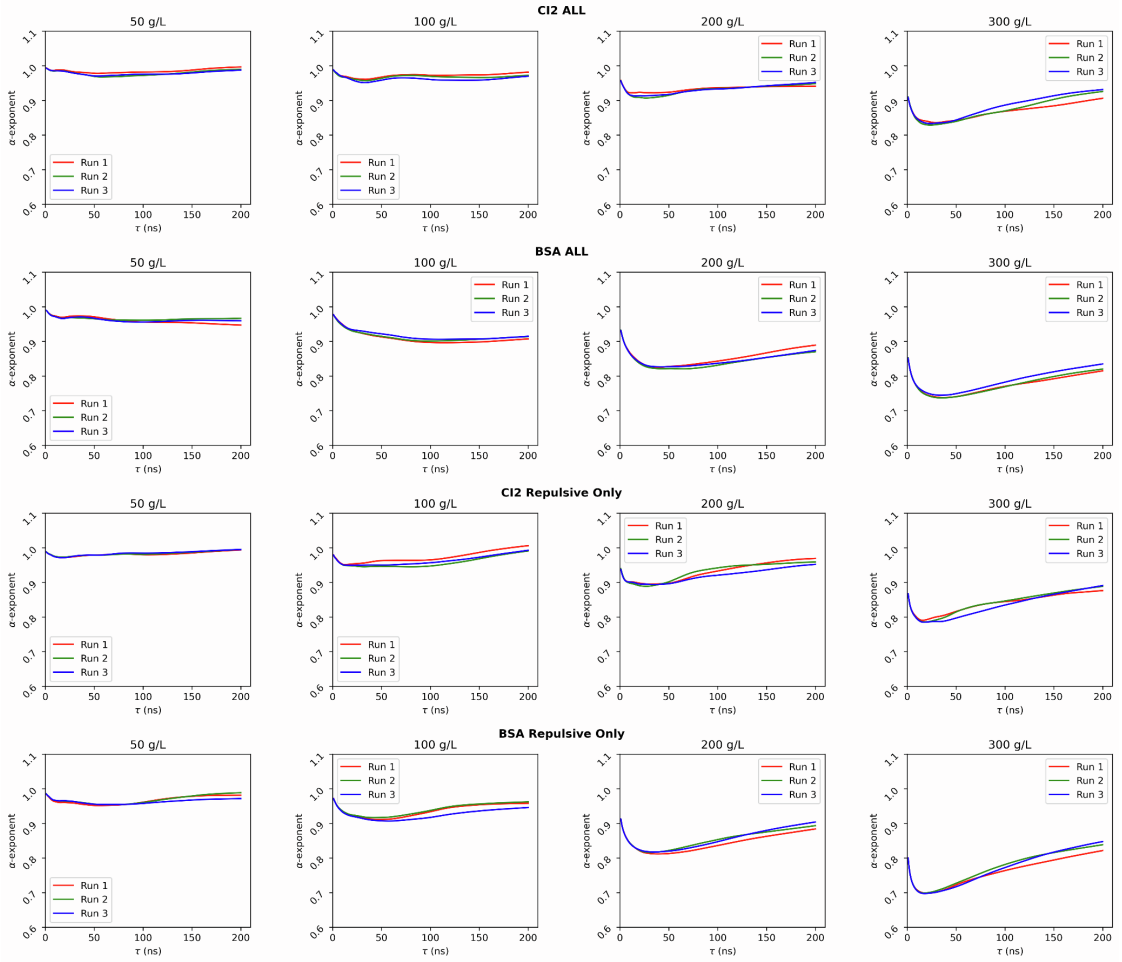

Figure S9. Calculation of α-exponent by fitting MSD vs lag-time plots to the power law equation (Equation 1 in the main text). The data shown here corresponds to BSA systems.

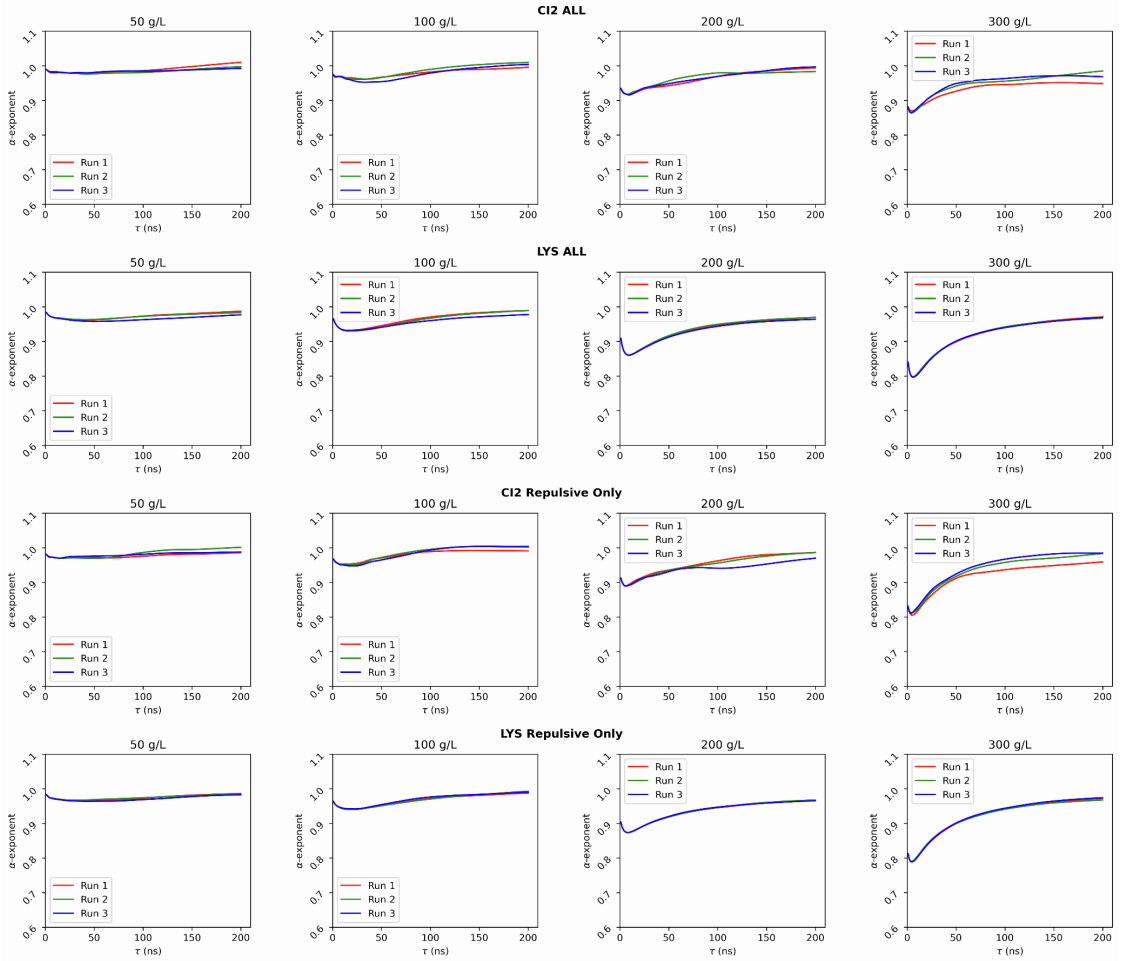

Figure S10. Calculation of α-exponent by fitting MSD vs lag-time plots to the power law equation (Equation 1 in the main text). The data shown here corresponds to lysozyme systems.

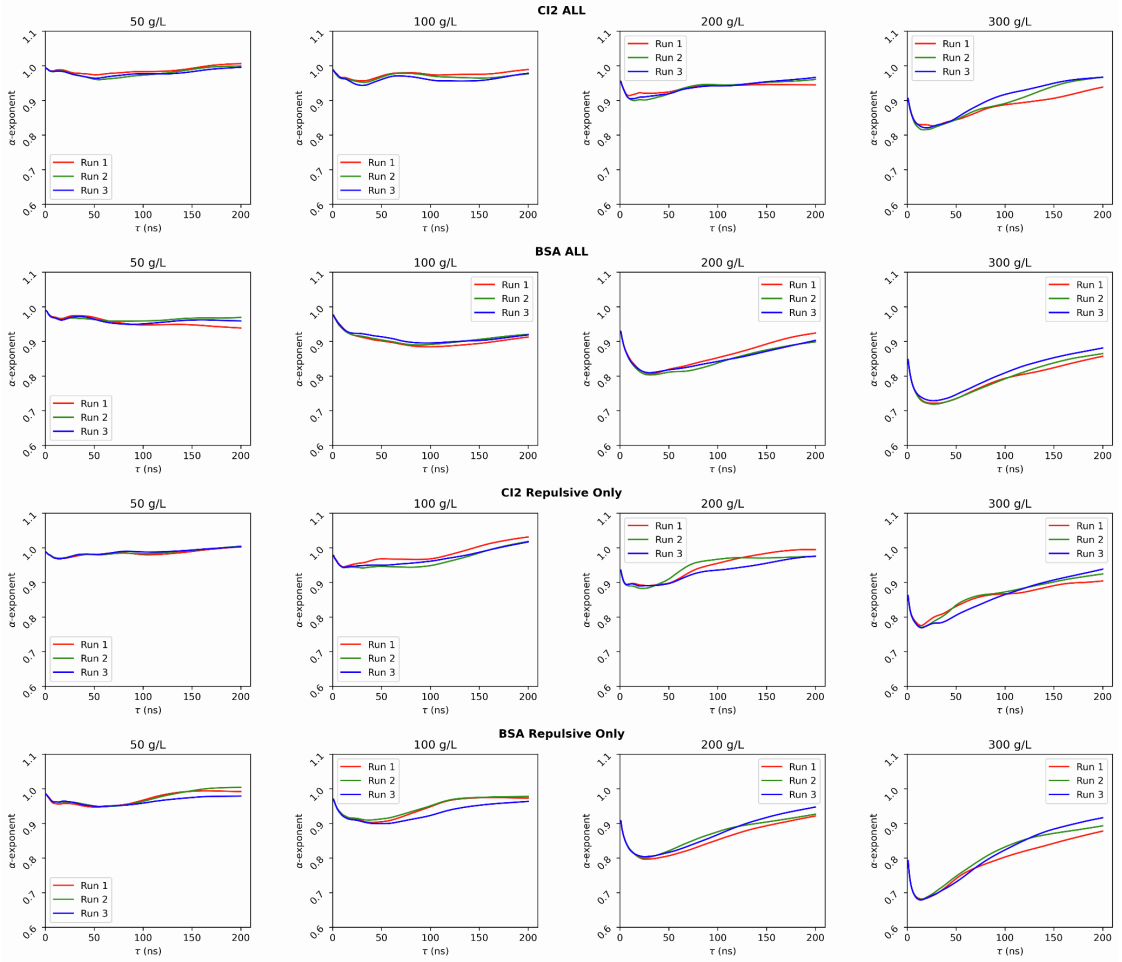

Figure S11. Calculation of α-exponent by fitting MSD vs lag-time plots to the power law equation with a static offset (Equation S1). The data shown correspond to BSA systems.

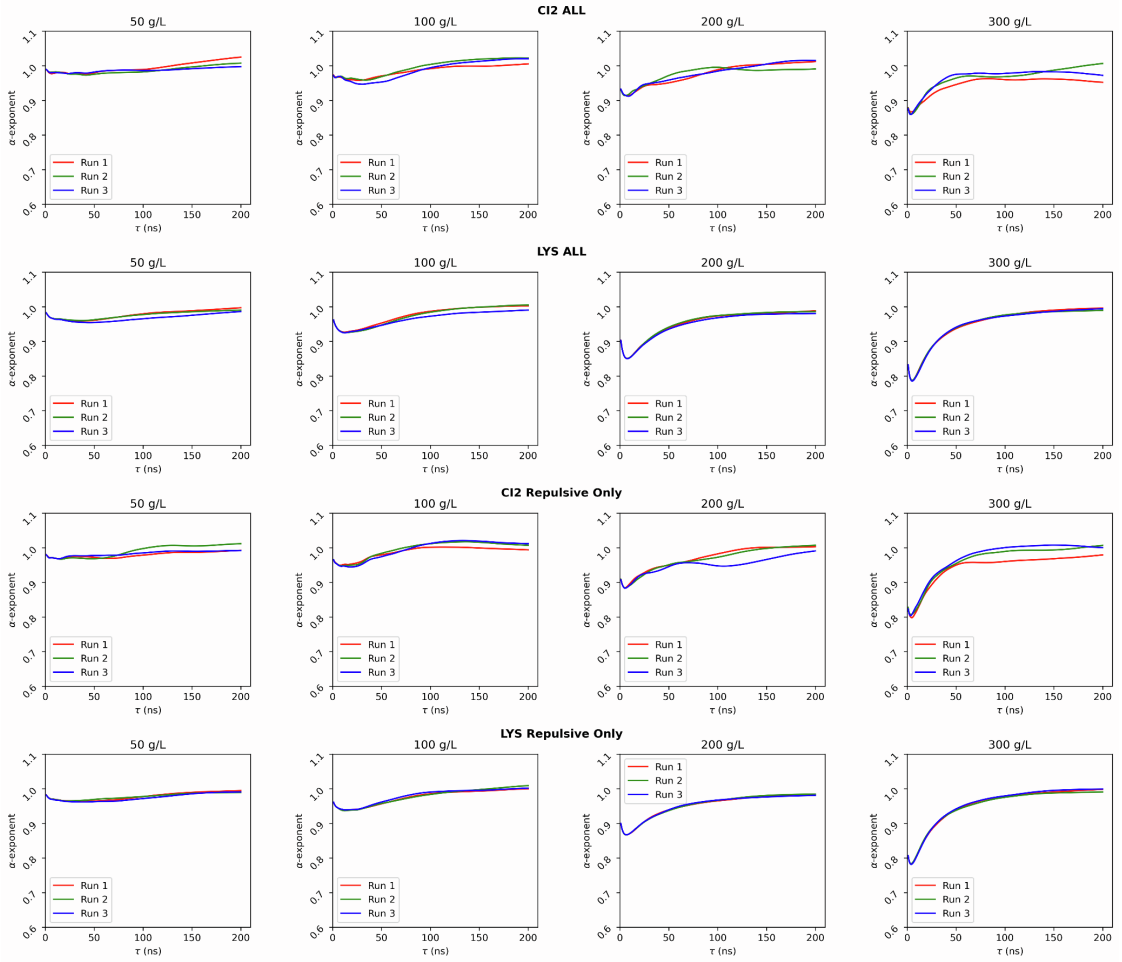

Figure S12. Calculation of α-exponent by fitting MSD vs lag-time plots to the power law equation with a static offset (Equation S1). The data shown here correspond to lysozyme systems.

Table S1. Number of crowder protein molecules, tracer molecules, and total molecules in the simulations.

Concentration	**Crowder**	**No. of crowder molecules**	**No. of tracer (CI2) molecules**	**Total no. of protein molecules**
50 g/L	BSA	21	26	47
100 g/L	BSA	43	26	69
200 g/L	BSA	86	26	112
300 g/L	BSA	128	26	154
50 g/L	LYS	96	26	122
100 g/L	LYS	193	26	219
200 g/L	LYS	385	26	411
300 g/L	LYS	579	26	605

Table S2. Timescales corresponding to the colors used in Figure 4.

Crowder/tracer	**50 g/L**			**300 g/L**		
	blue	red	yellow	blue	red	yellow
CI2 in BSA	2 ns	13.4 ns	39.4 ns	1 ns	24.6 ns	152.2 ns
CI2 in LYS	1 ns	6 ns	70 ns	0.4 ns	6 ns	55.6 ns
LYS	1 ns	6 ns	70 ns	0.4 ns	1 ns	129.4 ns
BSA	1.4 ns	17.6 ns	60 ns	1 ns	24 ns	237.4 ns

Table S3. Properties of the proteins simulated. Molecular surface area and molecular weight were calculated using Pymol. The charge data was taken from the PQR files generated by PDB2PQR.

	Molecular surface area ($Å^2$)	Molecular weight (kDa)	Net charge (pH=5.4)
CI2	7,554	7.28	-1
BSA	66,730	66.22	-2
LYS	14,197	14.31	+10

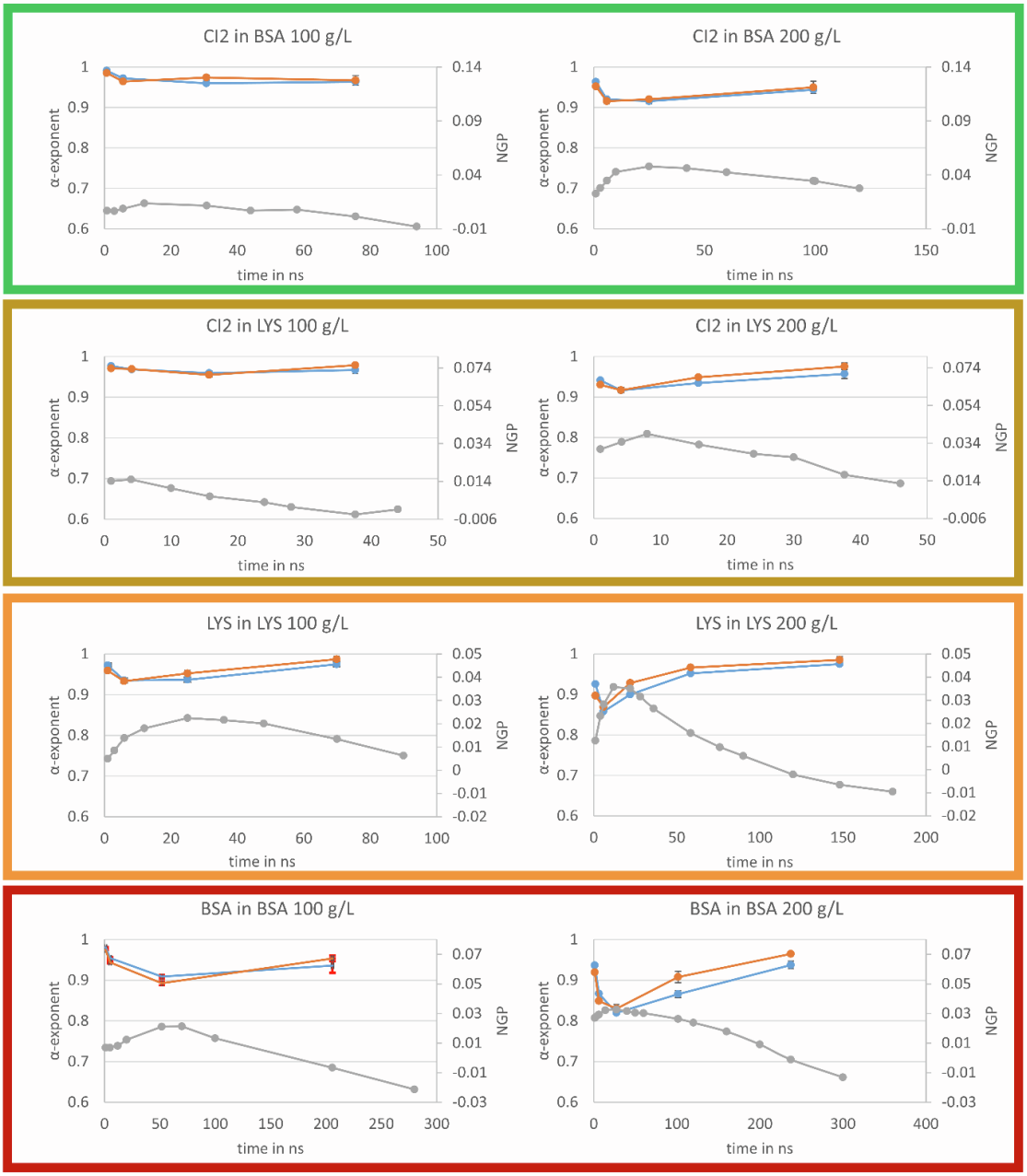

Figure S13. Sub-diffusive and non-Gaussianity properties of the crowders and tracer (at concentrations of 100 and 200 g/L of the crowder). The data is represented in the same way as in Figure 3 in the main text.

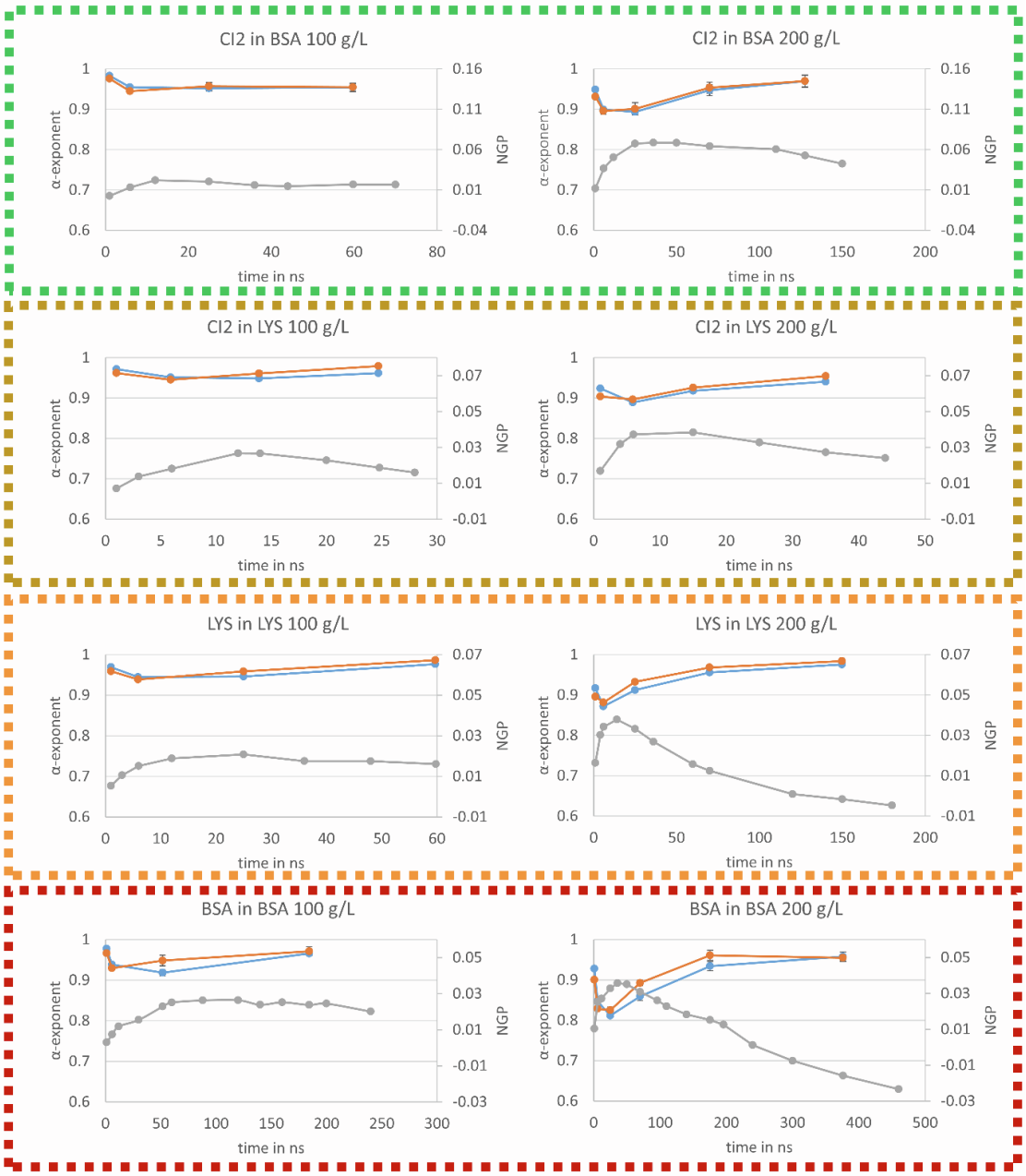

Figure S14. Properties of tracer and crowder in the absence of attractive interactions (100 and 200 g/L of the crowder). Error bars represent standard deviation (n=3). The data is represented in the same way as in Figure 5 in the main text.

References

62. Backlund, M.P., R. Joyner, and W.E. Moerner. 2015. Chromosomal locus tracking with proper accounting of static and dynamic errors. *Phys Rev E*. 91:62716.