Influence of Protein Crowder Size on Hydration Structure and Dynamics in Macromolecular Crowding

Po-hung Wang,^{†,} Isseki Yu,^{†,‡,*} Michael Feig,^{#,⊥} Yuji Sugita^{†,‡,⊥,v}*

[†]RIKEN Theoretical Molecular Science Laboratory, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan; [‡]RIKEN iTHES, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan; [#]Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, Michigan 48824, United States; [⊥]RIKEN Quantitative Biology Center (QBiC), Integrated Innovation Building 7F, 6-7-1 Minatojima-Minamimachi, Chuo-ku, Kobe, Hyogo, 650-0047, Japan; ^vRIKEN Advanced Institute for Computational Science (AICS), 7-1-26 Minatojima-Minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan.

Keywords: Macromolecular Crowding, Solvent Accessible Surface Area, Hydration, Diffusion coefficient, Protein Crowders

Corresponding Author: Yuji Sugita, RIKEN Theoretical Molecular Science Laboratory, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan; Phone: +81-48-462-1407; Fax: +81-48-467-4532; E-mail: sugita@riken.jp

* These two authors equally contributed the work.

ABSTRACT: We investigate the effects of protein crowder sizes on hydration structure and dynamics in macromolecular crowded systems by all-atom MD simulations. The crowded systems consisting of only small proteins showed larger total surface areas than those of large proteins at the same volume fractions. As a result, more water molecules were trapped within the hydration shells, slowing down water diffusion. The simulation results suggest that the protein crowder size is another factor to determine the effect of macromolecular crowding and to explain the experimental kinetic data of proteins and DNAs in the presence of crowding agents.

1. INTRODUCTION

Macromolecules occupy 20-40 % of the total volume of a cell corresponding to a concentration of up to 400 g/L[1]. In *in-vitro* experiments, the crowded environment has been mimicked by various crowding agents, such as polyethylene glycol (PEG) or Ficoll. Such environments often affect macromolecular structure and dynamics compared to those in dilute solution[2-5]. Many previous studies have suggested that compact globular conformations of macromolecules are favored mainly due to the excluded volume effect[2,3]. On the other hand, recent simulations as well as experimental studies have indicated that protein crowders may impact protein structure and dynamics in more complex ways due to non-specific protein-crowder interactions[6-9].

Even in cellular environments, 60-80 % of the total volume is occupied by water. The hydration structure and dynamics is a key factor for determining macromolecular structure, dynamics and functions. Therefore, if water properties are influenced by crowded cellular environments, there may be an indirect effect on macromolecular properties. Recently, highly concentrated protein solutions at different protein volume fractions were simulated by all-atom molecular dynamics (MD) simulations with explicit water[10,11]. In those studies, two small proteins, the villin headpiece subdomain (villin) and protein G, were used as protein crowders and the number of water molecules was adjusted in different simulation systems to effectively vary protein volume fractions. The main finding was that crowding reduces diffusion coefficients of water with increasing protein volume fractions[10]. These results were consistent with other studies that also suggested slow diffusion of surface water[12], and a lower dielectric constant upon confinement[13].

Feig et al. recently built a comprehensive atomistic model of the cytoplasm of *Mycoplasma genitalium*, which is based on genomic, proteomic, metabolomic, and structural information[14]. The same group simulated this model at different molecular levels, including hard-sphere Brownian dynamics simulations[15] and all-atom MD simulations[16]. In the resulting cytoplasmic system, macromolecular sizes are widely distributed ranging from small proteins to very large complexes, such as the molecular chaperon GroEL/ES or ribosomes. This raises the question of how important protein crowder sizes are. While crowder size has been considered before in the context of the volume exclusion effect[2,3], we are focusing here specifically on how protein crowder

size modulates hydration properties. More specifically we present results from atomistic simulations of multiple crowded protein systems consisting of smaller proteins (villin, protein G) or larger ones (bovine serum albumin (BSA)) at the same protein volume fractions to shed more light on the microscopic view of crowding effects in realistic biological environments.

This article is organized as follows: In the next section, the simulation and analysis protocols are described. In Section 3, water properties observed in different simulation systems are examined as a function of protein crowder sizes. In Section 4, we discuss the relationships between the simulation results and other relevant experimental data on macromolecular kinetics in *in-vitro* crowding conditions. Our work is concluded in Section 5.

2. METHODS

2.1 Molecular models and simulations

The chicken villin headpiece subdomain (36 residues, villin; PDB id: 1VII[17]), segment B1 of streptococcal protein G (56 residues, protein G; PDB id: 1PGB[18]), and bovine serum albumin (583 residues, BSA; PDB id: 4F5S[19]), were used to build crowded protein systems for all-atom MD simulations. Villin and protein G were selected as small globular proteins, while BSA was chosen as a large protein. The native states of these proteins are known to be stable and they do not aggregate with each other. Therefore, the three proteins are suitable to examine the protein size effect in macromolecular crowding. In the three crowded systems whose protein volume fractions are set to around 30 %, 132 villin, 88 protein G, and 8 BSA were randomly positioned in each cubic periodic box and solvated with TIP3P water molecules[20]. Counterions were added to the systems to neutralize the total net charge. The charges of BSA, protein G, and villin are -12, -4 and +2, respectively. In order to eliminate the effect of counterions on water structure and dynamics, we also built another three crowded systems where villin, protein G, and BSA were neutralized via amino acid modification so that counterions were not necessary. The neutralization was accomplished by protonating Asp/Glu residues or deprotonating Lys at the protein surfaces. For villin, two lysines (Lys65 and Lys71) were deprotonated, and two aspartates (Asp36 and Asp40) and two glutamates (Glu19 and Glu42) in protein G were protonated. For BSA, two aspartates (Asp254 and Asp555) and ten glutamates (Glu17,

Glu45, Glu125, Glu182, Glu229, Glu243, Glu332, Glu464, Glu519, and Glu564) were protonated. The simulation systems are summarized in Table 1 and the snapshots after equilibration are shown in Figure 1a.

Table 1. Six simulation systems

Protein	f^a	L (Å) ^b	N_{atom}^c	N_{prot}^d	C_{prot} (g/L) ^e	N_{ion}^f	N_{water}^g	$t_s(\text{ns})^h$
BSA	30%	130.27	231571	8	399.7	96 Na ⁺	52561	100
Neutralized BSA	30%	130.20	230995	8	400.1	0	52369	100
protein G	30%	130.08	232504	88	411.2	352 Na ⁺	52304	100
Neutralized protein G	30%	130.76	234616	88	405.1	0	53008	100
villin	31%	131.04	236112	132	408.5	264 Cl ⁻	52392	100
Neutralized villin	31%	131.37	237168	132	405.2	0	52920	100

^aProtein volume fraction. ^bCubic box size. ^cTotal number of atoms. ^dTotal number of proteins. ^eConcentration of proteins. ^fTotal number of ions. ^gTotal number of water molecules. ^hSimulation time.

Figure 1. Snapshots of the systems and the theoretically accessible volume, $V^A(r)$. (a) Snapshots of the six simulation systems after equilibration. BSA is depicted with a blue surface representation while protein G and villin are shown as green and purple surface representations, respectively. Water molecules are shown as red dots. Sodium and chloride ions are represented as yellow and cyan spheres, respectively. (b) Outer surface of the theoretically accessible volume $V^A(r)$ (gray) at $r = 4.0$ Å from BSAs (whose structures are shown as blue lines) with the cross-sectional view of the regions inside the $V^A(r)$ is shown as a yellow mesh. (c) Schematic representation of the $V^A(r)$ with a certain thickness Δr . r is the distance from the center of the closest heavy atom (filled light blue circles). Numerically, $V^A(r)$ is calculated with a resolution of 0.2 Å^3 . The increment of the $V^A(r)$ was calculated with a thickness (Δr) of 0.5 Å . (d) A schematic picture of differences between villin and BSA crowders is shown.

The crowded systems were energy-minimized and then heated up to 300 K with three short simulations at increasing temperatures (10 ps at 100 K, 10 ps at 200 K, and 10 ps at 300 K). They were then equilibrated for 6 ns in the NPT ensemble at 300 K and 1 bar. The production runs continued in the NPT ensemble at the same temperature and pressure. All MD simulations were performed using the GENESIS software package[21]. The CHARMM 36 all-atom force field[22] was used. The bond lengths

involving hydrogen atoms were constrained using the SHAKE algorithm[23]. The Berendsen coupling scheme[24] was used for temperature and pressure control with a coupling time constant for temperature of 0.1 ps and a coupling time constant for pressure of 1.0 ps, respectively. The electrostatic interactions were treated using the Particle Mesh Ewald (PME) method[25]. The cutoff distance for the Lennard-Jones interactions was set at 12 Å and a switching function became effective at 10 Å. Each system was simulated for 100 ns using a time step of 2 fs, which was sufficiently long for examining water structure and dynamics in the crowded environments. Trajectory snapshots were saved every 10 ps.

2.2 Total solvent-accessible surface area in a crowded system

We calculated the total solvent-accessible surface areas (SASA) of proteins in the crowded systems. In the original method described by Lee and Richards[26], SASA is defined as the area described by the center of a probe sphere that rolls on the van der Waals surface of a protein. Typically, the probe radius is set to 1.4 Å to represent a water molecule. For a single protein, the Lee and Richards' method is easily applied during trajectory analysis of all-atom MD simulations under the periodic boundary conditions. In order to estimate the total SASA for multiple proteins, we calculated the atomic SASA for all atoms in the unit cell of periodic boundary condition (PBC). To consider the influence of PBC, atoms in the image cells (within 7.0 Å from the outer boundary of unit cell) are also included for the analysis. Then, only the atomic SASAs for atoms in the unit cell are summated to obtain the total SASA. The CHARMM van der Waals radii[27] for all protein heavy atoms were used for the SASA analysis.

2.3 Hydration structure of a crowded system

To analyze the hydration structure of a crowded system, we examine the theoretically accessible volume, $V^A(r)$, and the (proximal) radial distribution function, $g(r)$ as in our previous analysis of the villin/protein G systems[10]. To calculate $V^A(r)$, we define the accessible region with a given distance r away from the closest heavy atom of a protein[10] (the gray region in Figure 1b). Then, the number of grid points is counted in the region for obtaining $V^A(r)$. The radial distribution function of water molecules, $g(r)$, from proteins is calculated using the following equation:

$$g(r) = \frac{d(r)}{bV^A(r)}, \quad (1)$$

where b is the bulk water density ($0.034 \text{ \AA}^{-3}$) and $d(r)$ is the number of water molecules in the accessible regions with a given distance r away from the closest protein heavy atom.

2.4 The dynamics of water in a crowded system

The dynamics of water in a crowded system is examined by calculating the diffusion coefficient and the transition rate of water between protein surfaces and the distal regions. The diffusion coefficient of water is obtained using the Einstein relation:

$$D_w = \lim_{t \rightarrow \infty} \frac{1}{6t} MSD(t) + D_c, \quad (2)$$

$$MSD(t) = \left\langle |r(t' + t) - r(t')|^2 \right\rangle_{t'}, \quad (3)$$

$$D_c = \frac{k_B T \xi}{6\pi\eta L}, \quad (4)$$

where MSD stands for mean square displacement, D_c is a correction term taking into consideration of the finite periodic box size effects, k_B is the Boltzmann constant, T is the temperature, ξ is the correction factor, η is the shear viscosity of pure water, and L is the box dimension[28].

We calculated the average transition rate between surface-bound and freely diffusing water molecules:

$$\text{Surface} \xrightleftharpoons[r_{D \rightarrow S}]{r_{S \rightarrow D}} \text{Distal}, \quad (5)$$

where $r_{S \rightarrow D}$ and $r_{D \rightarrow S}$ are the transition rates for water moving from protein surfaces to distal (non-surface) regions and from distal (non-surface) regions to protein surfaces, respectively. A surface-bound water is defined if the minimum distance between the

water molecule and any protein heavy atoms is shorter than 4 Å. A distance between a distal water and any protein heavy atoms is 7 Å or longer. Note that this leaves a transition region, where water is neither near the surface nor in the distal regions. Only a trajectory crossing the transition region between the two well-defined regions is deemed to make a transition to the respective other state. This idea avoids apparent barrier recrossing events caused by a poorly defined boundary between states[29-31]. For each water, we obtained its time series of states according to the above definition and calculated average residence times for protein surfaces, $\tau_s = 1/r_{s \rightarrow D}$ and for distal regions $\tau_D = 1/r_{D \rightarrow s}$, respectively.

3. RESULTS

3.1 Protein structures in the crowded systems

In the all-atom MD simulations of the six crowded systems, villin, protein G, and BSA as well as their neutralized ones were generally stable. BSA shows slightly larger backbone root mean square deviations (RMSD) with respect to the initial structures in the MD simulation due to the flexible loop regions (see Table 2). The total SASA for the six crowded systems are also listed in Table 2. SASA values for single villin, protein G, and BSA copies in dilute condition, obtained from 20 ns MD simulation of each protein in water, are 3227.0 Å², 3770.6 Å², and 33170.1 Å², respectively. If the same number of proteins exists in dilute solution without any contacts, the total SASA (SASA_{dilute}) for villin, protein G, and BSA becomes 425964.0 Å², 331808.4 Å², and 265360.8 Å², respectively. These values are much larger than the total SASA in the actual crowding simulations (SASA_{crowding}), since proteins have numerous contacts between each other in the crowded environments. This difference, $\Delta = \text{SASA}_{\text{dilute}} - \text{SASA}_{\text{crowding}}$, increases as the number of copies of proteins increases, i.e. $\Delta_{\text{BSA}} < \Delta_{\text{protein G}} < \Delta_{\text{villin}}$. In the charge-neutralized systems, the total SASAs generally were smaller than those in the original systems. The protein G systems show the largest difference of the total SASA, probably due to the large difference in the surface polarity. However, the tendency of smaller proteins exhibiting larger total SASAs in the crowded environment at the same protein volume fraction is true in both the original and neutralized protein systems, suggesting that molecular geometry rather than electrostatic interactions is the determining factor.

Table 2. Backbone RMSD and total SASA in the crowding simulations.

Name	Backbone RMSD (Å)	Total SASA (Å ²)
BSA	1.92±0.27	214267.3±329.9
Neutralized BSA	2.07±0.30	215604.3±297.6
protein G	0.84±0.03	253507.6±253.5
Neutralized protein G	0.86±0.03	237182.8±273.2
villin	1.61±0.01	264796.2±836.7
Neutralized villin	1.59±0.01	256342.4±595.3

3.2 Hydration structure of the crowded systems

Figure 2a shows the theoretically accessible volume $V^A(r)$ for each system. The volume quickly increases to the peak at $r = 1.4$ Å and then gradually decreases until it becomes zero at $r = 18$ Å. The only peak corresponds to the average van der Waals radii of the atoms at the protein surfaces and shows a small difference in height between the six systems. In Figure 2b, we show the difference of the calculated volumes with respect to that of the wild-type BSA system. At $r < 7$ Å, a larger volume is available for the protein G and villin systems, whereas these two have smaller accessible volumes at $r \geq 7$ Å. The accessible volumes of the BSA systems are the lowest within the second shell, but this trend is reversed in the distal regions. Schematic pictures in Figure 1d explain the accessible volume difference between the BSA and villin systems. BSA has fewer contacts with the neighboring proteins, while villin has more contacts with neighboring ones. In Figure 1d, the blue regions in the right panel is smaller than that in the left, since total SASA in the BSA system is much smaller than that of the villin system, as shown in Table 2. For $r \geq 7$ Å, The BSA system has a larger accessible volume because the packing is not ideal. On the other hand, villins occupy the box volume in a more efficient way, resulting in smaller distal spaces.

Figure 2. Hydration structures of the crowded systems. (a) Theoretically accessible volume, $V^A(r)$, (b) difference of $V^A(r)$ with respect to the wild-type BSA system, (c) percentage of water location, $X^w(r)$ as a function of the distance from the nearest protein atom, and (d) radial distribution function, $g(r)$. Black, red, and blue lines represent villin,

protein G, and BSA, and the full and dashed lines indicate the wild-type and neutralized systems, respectively.

Figure 2c shows the cumulative percentages of water location, $X^w(r)$ as a distance from the nearest protein heavy atom. Only a small percentage (2 %) of water molecules are located within 2.0 Å from the nearest protein heavy atoms and around 50 % of water is found within 5 Å. Table 3 summarizes the accessible volumes and water locations in three well-defined regions: the first shell ($r \leq 4$ Å), the second shell ($4 < r \leq 7$ Å), and the distal region ($r > 7$ Å). The values were in good agreement with our previous studies of the villin/protein G crowding simulations at similar protein volume fractions[10]. The BSA systems have less water molecules within 7 Å of the proteins than the protein G and villin systems.

Table 3. Theoretically accessible volume and water locations in the three well-defined regions.

Protein	1 st shell ($r \leq 4$ Å)		2 nd shell ($4 < r \leq 7$ Å)		distal ($r > 7$ Å)	
	$V^A(r)$ (%) ^a	$X^w(r)$ (%) ^b	$V^A(r)$ (%)	$X^w(r)$ (%)	$V^A(r)$ (%)	$X^w(r)$ (%)
BSA	54.4	35.6	18.9	26.5	26.7	37.9
Neutralized BSA	54.7	35.9	19.1	26.9	26.2	37.2
protein G	59.1	42.4	22.1	30.9	18.8	26.7
Neutralized protein G	56.8	39.1	20.3	28.3	22.9	32.6
villin	60.4	43.5	20.6	29.1	19.0	27.4
Neutralized villin	59.0	41.6	19.6	27.7	21.4	30.7

^aThe percentage of theoretical accessible volume, $V^A(r)$, for the first and second shells and distal regions. ^bThe percentage of water locations, $X^w(r)$, for the first and second shells and distal regions.

Interestingly, Figure 2d shows that the radial distribution function, $g(r)$, is almost unchanged between the six crowded systems. In Eq. (1), the changes in numerator, $d(r)$, are cancelled with the denominator, $V^A(r)$. The difference of crowder size distribution could change the accessible volumes and the number of water molecules at distal regions, but the water densities as a function of the distance from protein surfaces remain the same.

3.3 The dynamics of water in the crowded systems

The average MSD of water molecules is plotted as a function of time in Figure 3a, indicating a linear trend for all the systems. Linear fits give the slope of each curve and the diffusion coefficients were calculated using Eq. (2)-(4). Figure 3a also shows a

slowdown in the diffusion depending on the size of the protein crowders. Plots of $\log[\text{MSD}(t)/t]$ vs. $\log t$, shown in Figure 3b, give an illustration of how anomalous diffusion occurs at short times ($t < 10$ ns). On this time scale, $\log[\text{MSD}(t)/t]$ is decreasing as a function of $\log t$. For longer times ($t > 10$ ns), $\log[\text{MSD}(t)/t]$ becomes gradually converged.

In Table 4, the diffusion coefficients of water, D_w , the residence time of water near protein surfaces, τ_s , and the distal regions, τ_D , are summarized. We calculated the following quotient Q ,

$$Q = \frac{r_{s \rightarrow D} X_s^w}{r_{D \rightarrow s} X_D^w} \approx \frac{\tau_D (X_{1st}^w + X_{2nd}^w)}{\tau_s X_D^w}, \quad (6)$$

where X_s^w , the percentage of water locations at the protein surface, is approximated by summation of X_{1st}^w and X_{2nd}^w (the percentage of water locations at 1st and 2nd hydration shells). X_D^w is the percentage of water locations in the distal region. The resulting Q values are close to 1 for all 6 systems, suggesting the detailed balance relation described with the transition rates and the probabilities of water locations. Larger values of D_w were observed in the BSA systems compared to the protein G and villin systems. The residence times, τ_s , were 160-170 ps for villin and BSA but somewhat shorter, around 110-120 ps, for protein G, presumably due to a reduced surface polarity. These results suggest that the residence time of water near protein surfaces is mainly determined by protein-water interactions. In contrast, τ_D for the BSA systems are around 100 ps, while those for the villin and protein G systems are shorter than 60 ps. Large accessible volume in the distal regions in the BSA system likely allows water molecules to stay there longer.

We calculate the rotational relaxation of water using Eq. 7 in ref. [32]

$$P_2(t) = \langle 1.5 \cos^2 \theta(t) - 0.5 \rangle, \quad (7)$$

and the O-H vectors of water were used in the calculation. We then fit the correlation curves using an exponential function to obtain the relaxation times. The rotational relaxation times for water in the distal region are around 0.5 ps for all systems, comparable to the relaxation time of bulk water reported in ref. [32]. The rotational relaxation times for water at the surface are 92.8 ps, 113.4 ps, and 122.2 ps for BSA, protein G, and villin systems, respectively. These values are 30 times larger than the

relaxation time of water at the surface of lysozyme in dilute condition. Similar for the different proteins studied in the simulations, this slow-down in relaxation time corresponds to the long residence time scales τ_s and further indicates that the overall motion of water molecules at the protein surface is retarded significantly over dilute conditions. This large decrease in rotational diffusion of water may also help explain the 22-to-64-fold slowdown of rotational diffusion of ubiquitin at a glucose concentration of 325 g/L found in a combined NMR and MD simulation study[33]. Another experimental study also has shown a tenfold reduction in rotational diffusion of macromolecules at physiologically relevant concentrations of crowders[34].

Figure 3. Mean square displacement (MSD) and $\log[\text{MSD}(t)/t]$ as a function of time.

(a) The mean square displacement of water for the systems, (b) the anomalous diffusion indicator, $\log[\text{MSD}(t)/t]$ against $\log t$ in (b). The color code is the same as in Figure 2.

Table 4. Diffusion kinetics of water molecules, number of contacts for each protein, and the exchange of 3 contacting partners of each protein (N_{ex}^3).

Protein	D_w ($\text{\AA}^2/\text{ps}$)	τ_s (ps) ^a	τ_D (ps) ^a	Contacts ^b	N_{ex} ³
BSA	0.306±0.015	161.0	96.6	3.850±0.207	1
Neutralized BSA	0.312±0.016	162.0	97.8	5.430±0.307	1
protein G	0.265±0.013	119.6	52.0	4.185±0.114	50
Neutralized protein G	0.289±0.011	109.6	60.3	10.805±0.231	50
villin	0.261±0.011	178.6	52.5	4.998±0.103	71
Neutralized villin	0.278±0.014	172.9	57.6	12.045±0.208	71

^aObtained by averaging over times when water molecules stay at protein surface (τ_s) and at distal region (τ_D), assuming Eq. (5). Please refer to Section 2.4 for the calculation of the times. ^bThe average number of contacts as determined by the shortest C α distances between proteins using a cutoff distance of 9 $\text{\AA}$. More details for the calculation can be found in Section 4.1.

4. DISCUSSION

4.1 Force-field limitations

The current study is limited by the accuracy of molecular force fields describing protein-protein and protein-water interactions. As for the protein-protein interaction, it has been reported that the currently used force fields have a tendency to overestimate aggregation and compaction[35-37]. Furthermore, the TIP3P model, which we used to describe water interactions in the simulations, is known to provide faster kinetics and diffusion of water compared to the experiments and simulations with other water models. However, our focus in this study is the relative difference of hydration dynamics in the crowded protein systems at the same protein volume fraction but with different sizes of crowders. For investigating such differences, we consider that the current models and force-field parameters are sufficiently accurate and can provide meaningful results. To further support our argument, we calculate the shortest distance between C α atoms of proteins and count the number of contacts between proteins using a cutoff of 9 $\text{\AA}$. The averaged contacts are summarized in Table 4. We also search for the exchange of 3 contacting partners of each protein (N_{ex} ³). The 3 closest proteins (contact partners) of each protein were first defined using the shortest distances as criteria. If all the three proteins of a target protein are different between at 50 ns and 100 ns, we count one exchange for the protein, as shown in Table 4. We show that more than 50 % of villin and protein G experienced at least one exchange in the simulation. The number of exchanges in BSA in Table 4 is small since the total number of contacts in BSA crowded systems is much smaller than other two systems. These results suggest

that, in our simulations, no strong aggregation was observed. We additionally note that further optimization of the protein-water interactions is underway by us as well as others to better describe protein folding and stability in water and cellular environments.

4.2 Mechanisms for slow down of water diffusion in the presence of small protein crowders

How are the diffusion coefficients of water determined in the current crowding simulations? To answer this question, we examine the correlations between D_w and other properties. First, we plot the correlation between D_w and the total SASA in Figure 4a, showing a linear dependency for the results from the six crowding simulations. The error bars of D_w , calculated using block analysis using five blocks (each block had 10 ns simulation time), are not shown in the figure because they are within 5 % of the calculated diffusion constants. Second, we correlate D_w with the percentage of water in the distal region, $X^w(r)$ for $r > 7$ Å (see Figure 4b). The six data points again fit well to a straight line with $R^2 = 0.96$. An extrapolation of the fitted line gives a theoretical limit of $D_w = 0.1456$ Å²/ps. when no water can reside in the distal region. Third, we correlate D_w with the average residence time of water molecules at the distal region, τ_D , shown in Figure 4c. Again we apply a linear fit to the six data points. All the three plots show good linear fits, suggesting that the difference of water diffusion upon protein crowder sizes is determined by the volume of distal regions from proteins and the number of water residing at the space. Good linear correlation also supports the assumption that the slowdown effect on water molecules at the distal region from different interactions between water and protein side-chains in all systems can be negligible. At the protein surface, water dynamics is determined by protein-water interactions that may differ between proteins, as shown by elongated τ_s and rotational times of water. At the distal region, water dynamics remains mostly the same, as shown by the unaffected rotational times of water. SASA affects the distribution of the water molecules at different regions, and the distribution determines the slowdown effect of water diffusion.

Figure 4. Correlation between the diffusion coefficient of water and total SASA, percentage of water in the distal region, and the times in the distal region. The correlation between the diffusion coefficient of water in crowded protein systems, D_w , and the total SASA of proteins is shown in (a). A linear fit of six data points gives a

good correlation with $R^2 = 0.93$, shown as a dashed line in (a). The best fitting formula is $D_w = -9.4 \times 10^{-7} \text{ SASA} + 0.5111$. The correlation between D_w and the percentage of water in the distal region, X , is shown in (b). The six data points, corresponding to the six systems, fit well to a straight line with $R^2 = 0.96$. The best fitting formula, shown as a dashed line in (b), is $D_w = 4.4 \times 10^{-3} X_D + 0.1456$. The correlation between D_w and τ_D , is shown in (c). Details of the calculation of τ_D , the average time when water is in the distal region, can be found in Section 2.4. The best fitting formula with $R^2 = 0.87$ is $D_w = 9.0 \times 10^{-4} \tau_D + 0.222$. The error bars of D_w , calculated using block analysis (five blocks), were not included in the figure because they are less than 5 % of the calculated D_w .

4.3 The size effect on kinetics in the crowded systems

The current MD simulations suggest that systems with smaller-sized proteins (protein G and villin) have a stronger effect on water dynamics in the crowded systems. This is an additional insight to our previous observations on the reduced diffusion as a

function of protein volume fractions[10]. There are several experimental reports in which smaller rather than larger crowders induce higher solvent viscosity[38-41], or where large crowders reach an upper limit in terms of exerting crowding effects[42]. An important question in the current study is how the reduction of water diffusion is related to various properties of macromolecules in the crowded environments. As far as the mobility of macromolecules is concerned, proteins tend to diffuse slower in the presence of crowders. This tendency is more enhanced when the crowder size is smaller[34,43-45], which is in agreement with our hypothesis. Measured micro-viscosity for calmodulin as assessed by altered translational mobility also shows an increasing trend as crowder size becomes smaller[38]. The hydration dynamics of ubiquitin is also affected more by small crowders[46]. Crowding also reduces binding rates, and small crowders result in more reduction than large crowders[41,43,47], probably due to slow diffusion of binding partners as a result of slow diffusion of water observed in the all-atom MD simulations.

Our work may have some implications for correlated diffusion of water, ligands, and proteins in a living cell. Water diffusion correlates with those of ions and ligands, so varying water diffusion near proteins of different sizes would correlate with varying ligand diffusion near proteins of different sizes, thus affecting the diffusion-limited enzymatic reaction. Moreover, our work implies the importance of heterogeneity of proteins or other macromolecules in a cell, as suggested in a recent study[15]. If only small globular proteins of the same size exist in a cell, large number of protein-protein contacts would enhance their aggregations and may hinder most of the important biological functions.

Another implication from our work is to help design *in vitro* experiments, especially when crowders are added into the test tube to establish a crowded environment. It can work as a general guideline and help the experimentalist decide the best crowder for a specific target protein, with crowder size taken into consideration.

5. CONCLUSIONS

In the present study, we have examined the effect of protein crowder size on the hydration structure and dynamics by carrying out all-atom MD simulations of six

crowded systems with different protein crowder sizes at the same protein volume fraction. The systems with smaller protein crowders show larger surface areas in total, limiting the number of freely diffusing water molecules. As a result, the diffusion of water molecules in the system slows down compared those with larger protein crowders. This microscopic view of the difference in hydration structure and dynamics explains the influence of crowder sizes on various kinetic properties of macromolecules in the crowded environment.

ASSOCIATED CONTENTS

AUTHOR INFORMATION

Corresponding Author

*E-mail: sugita@riken.jp (Yuji Sugita).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The simulations and analysis were carried out using the RIKEN Integrated Cluster of Clusters (RICC) and RIKEN HOKUSAI supercomputer systems through the HPCI strategic research project (hp140229, hp150233) and HPCI general trial use project (hp150145, hp160120) and **FLAGSHIP 2020 project focused area 1 “Innovative drug discovery infrastructure through functional control of biomolecular systems (hp160207)”**. This work was supported in part by RIKEN QBiC, iTHES **and pioneering project “Dynamic structural biology”**(to YS), a Grant-in-Aid for Scientific Research on Innovative Area “Novel measurement techniques for visualizing ‘live’ protein molecules at work” (No. 26119006) (to YS), a grant from JST CREST on “Structural Life Science and Advanced Core Technologies for Innovative Life Science Research” (to YS), a Grant-in-Aid for Scientific Research (C) from MEXT (No. 25410025) and **Incentive Research Projects from RIKEN** (to IY), and support from the U.S. National

Institutes of Health (NIH, GM092949, GM084943) and the U.S. National Science Foundation (NSF, MCB 1330560) (to MF).

References:

- [1] R.J. Ellis, A.P. Minton, *Nature* 425 (2003) 27.
- [2] A.P. Minton, J. Wilf, *Biochemistry* 20 (1981) 4821.
- [3] H.X. Zhou, G. Rivas, A.P. Minton, *Annu. Rev. Biophys.* 37 (2008) 375.
- [4] F. Despa, D.P. Orgill, R.C. Lee, *Ann. N. Y. Acad. Sci.* 1066 (2005) 54.
- [5] J. Spitzer, *Microbiol. Mol. Biol. Rev.* 75 (2011) 491.
- [6] M. Feig, Y. Sugita, *J. Phys. Chem. B* 116 (2012) 599.
- [7] A.C. Miklos, M. Sarkar, Y. Wang, G.J. Pielak, *J. Am. Chem. Soc.* 133 (2011) 7116.
- [8] M. Sarkar, A.E. Smith, G.J. Pielak, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 19342.
- [9] M. Sarkar, J. Lu, G.J. Pielak, *Biochemistry* 53 (2014) 1601.
- [10] R. Harada, Y. Sugita, M. Feig, *J. Am. Chem. Soc.* 134 (2012) 4842.
- [11] R. Harada, N. Tochio, T. Kigawa, Y. Sugita, M. Feig, *J. Am. Chem. Soc.* 135 (2013) 3696.
- [12] V.A. Makarov, M. Feig, B.K. Andrews, B.M. Pettitt, *Biophys. J.* 75 (1998) 150.
- [13] F. Despa, A. Fernandez, R.S. Berry, *Phys. Rev. Lett.* 93 (2004) 228104.
- [14] M. Feig, Y. Sugita, *J. Mol. Graph. Model.* 45 (2013) 144.
- [15] T. Ando, I. Yu, M. Feig, Y. Sugita, *J. Phys. Chem. B* 120 (2016) 11856.
- [16] I. Yu, T. Mori, T. Ando, R. Harada, J. Jung, Y. Sugita, M. Feig, *eLife* 5 (2016) e19274.
- [17] C.J. McKnight, P.T. Matsudaira, P.S. Kim, *Nat. Struct. Biol.* 4 (1997) 180.
- [18] T. Gallagher, P. Alexander, P. Bryan, G.L. Gilliland, *Biochemistry* 33 (1994) 4721.
- [19] A. Bujacz, *Acta Crystallogr. Sect. D-Biol. Crystallogr.* 68 (2012) 1278.
- [20] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, *J. Chem. Phys.* 79 (1983) 926.
- [21] J. Jung, T. Mori, C. Kobayashi, Y. Matsunaga, T. Yoda, M. Feig, Y. Sugita, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 5 (2015) 310.
- [22] A.D. Mackerell, Jr., M. Feig, C.L. Brooks, 3rd, *J. Comput. Chem.* 25 (2004) 1400.
- [23] J.-P. Ryckaert, G. Ciccotti, H.J. Berendsen, *J. Comput. Phys.* 23 (1977) 327.
- [24] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, J.

- Chem. Phys. 81 (1984) 3684.
- [25] T. Darden, D. York, L. Pedersen, J. Chem. Phys. 98 (1993) 10089.
 - [26] B. Lee, F.M. Richards, J. Mol. Biol. 55 (1971) 379.
 - [27] M. Nina, D. Beglov, B. Roux, J. Phys. Chem. B 101 (1997) 5239.
 - [28] I.-C. Yeh, G. Hummer, J. Phys. Chem. B 108 (2004) 15873.
 - [29] N.-V. Buchete, G. Hummer, J. Phys. Chem. B 112 (2008) 6057.
 - [30] M. Sarich, F. Noé, C. Schütte, Multiscale Model. Simul. 8 (2010) 1154.
 - [31] P.-h. Wang, R.B. Best, J. Blumberger, J. Am. Chem. Soc. 133 (2011) 3548.
 - [32] M. Marchi, F. Sterpone, M. Ceccarelli, J. Am. Chem. Soc. 124 (2002) 6787.
 - [33] E. Spiga, L.A. Abriata, F. Piazza, M. Dal Peraro, J. Phys. Chem. B 118 (2014) 5310.
 - [34] Y. Wang, C. Li, G.J. Pielak, J. Am. Chem. Soc. 132 (2010) 9392.
 - [35] D. Petrov, B. Zagrovic, PLoS Comput. Biol. 10 (2014) e1003638.
 - [36] L.A. Abriata, M. Dal Peraro, Sci. Rep. 5 (2015) 10549.
 - [37] R.B. Best, W. Zheng, J. Mittal, J. Chem. Theory Comput. 10 (2014) 5113.
 - [38] A.B. Goins, H. Sanabria, M.N. Waxham, Biophys. J. 95 (2008) 5362.
 - [39] J. Shin, A.G. Cherstvy, R. Metzler, Soft Matter 11 (2015) 472.
 - [40] E. Dauty, A.S. Verkman, J. Mol. Recogn. 17 (2004) 441.
 - [41] N. Kozar, Y.Y. Kuttner, G. Haran, G. Schreiber, Biophys. J. 92 (2007) 2139.
 - [42] A. Miklos, C. Li, C. Sorrell, L.A. Lyon, G. Pielak, BMC Biophys. 4 (2011) 13.
 - [43] Y.Y. Kuttner, N. Kozar, E. Segal, G. Schreiber, G. Haran, J. Am. Chem. Soc. 127 (2005) 15138.
 - [44] D.S. Banks, C. Fradin, Biophys. J. 89 (2005) 2960.
 - [45] D. Lavalette, M. Hink, M. Tourbez, C. Tétreau, A. Visser, Eur. Biophys. J. 35 (2006) 517.
 - [46] Luciano A. Abriata, E. Spiga, Matteo D. Peraro, Biophys. J. 111 (2016) 743.
 - [47] N. Kozar, G. Schreiber, J. Mol. Biol. 336 (2004) 763.