Rotational and Translational Diffusion of Proteins as a Function of Concentration

Zahedeh Bashardanesh,† Johan Elf,† Haiyang Zhang,*‡ and David van der Spoel*†‡

1Uppsala Center for Computational Chemistry, Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, Box 596, SE-75124 Uppsala, Sweden
2Department of Biological Science and Engineering, School of Chemistry and Biological Engineering, University of Science and Technology Beijing, 100083 Beijing, China

ABSTRACT: Atomistic simulations of three different proteins at different concentrations are performed to obtain insight into protein mobility as a function of protein concentration. We report on simulations of proteins from dilute to the physiological water concentration (about 70% of the mass). First, the viscosity was computed and found to increase by a factor of 7−9 going from pure water to the highest protein concentration, in excellent agreement with in vivo nuclear magnetic resonance results. At a physiological concentration of proteins, the translational diffusion is found to be slowed down to about 30% of the in vitro values. The slow-down of diffusion found here using atomistic models is slightly more than that of a hard sphere model that neglects the electrostatic interactions. Interestingly, rotational diffusion of proteins is slowed down somewhat more (by about 80−95% compared to in vitro values) than translational diffusion, in line with experimental findings and consistent with the increased viscosity. The finding that rotation is retarded more than translation is attributed to solvent-separated clustering. No direct interactions between the proteins are found, and the clustering can likely be attributed to dispersion interactions that are stronger between proteins than between protein and water. Based on these simulations, we can also conclude that the internal dynamics of the proteins in our study are affected only marginally under crowding conditions, and the proteins become somewhat more stable at higher concentrations. Simulations were performed using a force field that was tuned for dealing with crowding conditions by strengthening the protein−protein interactions and the proteins become somewhat more stable at higher concentrations. Simulations were performed using a force field that was tuned for dealing with crowding conditions by strengthening the protein−protein interactions and the proteins become somewhat more stable at higher concentrations. Simulations were performed using a force field that was tuned for dealing with crowding conditions by strengthening the protein−protein interactions and the proteins become somewhat more stable at higher concentrations.

INTRODUCTION

The intracellular environment is densely packed with macromolecules like nucleic acids, proteins, and sugars. The macromolecule concentration in, for instance, Escherichia coli is ≈300−400 mg/mL,1 corresponding to ≈40% of the intracellular volume.2,3 It has been shown that a high concentration of macromolecules influences the thermodynamics and kinetics of cellular processes.3−7 Most in vitro biochemical investigations are performed using dilute solutions of up to 10 mg/mL. Despite this obvious difference, it is commonly assumed, due to the lack of available data, that biochemical quantities such as protein−ligand binding constants or enzymatic reaction rates in vivo are comparable to the in vitro values. Under in vivo conditions, steric effects and nonspecific interactions are significant, however, and these constraints likely affect cellular processes. Dynamics of biomolecules have been studied extensively by nuclear magnetic resonance (NMR) experiments,8,9 and this has led to models describing the internal protein dynamics as a combination of intramolecular vibrations and overall tumbling.10 The typical rotational tumbling time for a small protein of a few nanoseconds, which has emerged from NMR experiments,8,11 is not the same in a crowded environment.12 Therefore, an in-depth understanding of the diffusive properties of proteins under conditions mimicking the interior of a cell is of general interest for understanding cellular processes.

Diffusion of macromolecules in the cytoplasm can be measured through a variety of experimental techniques, including direct or indirect optical techniques such as fluorescence correlation spectroscopy (FCS),13 single-particle tracking (SPT),14,15 fluorescence recovery after photobleaching (FRAP),16,17 or spectroscopic techniques such as pulse field gradient (PFG) NMR via diffusion-ordered spectroscopy (DOSY).18 For example, the translational diffusion coefficient of green fluorescent protein (GFP) in E. coli was measured by FRAP to be reduced to 10% of the value in water.19,20

Theoretical models based on hard sphere (HS) models allow investigation of the steric effects on HS properties, due to...
amino acids in TIP4P/2005.42 It was found that the \( \Delta \) TIP4P/2005 combination is a good compromise that predicts amino acid side chains was evaluated alongside the diatoms.37,38 Best et al. proposed to replace the water model to strengthen the Lennard-Jones (LJ) interactions,36 to the interaction between proteins and water by selectively treated protein solutions to analyze dynamic properties.31 The and co-workers have presented large simulations of concentrations of glucose on protein dynamics and found significantly retarded protein dynamics and protein dehydration.26 Yu and co-workers reported on a model for a bacterial cytoplasm.27 From this work, they find that rotational as well as translational diffusion decrease in a similar manner due to crowding. Water dynamics in a crowded environment was addressed by the Feig group in multiple studies. In one study, the structure of water around proteins was investigated,28 while another focused on water dynamics by studying proteins of different sizes.29 One further simulation study of Villin, a 36-residue protein, under crowding conditions, showed that protein–protein interactions led to protein rotational motion being retarded more than translational motion.30 von Bülow and co-workers have presented large simulations of concentrated protein solutions to analyze dynamic properties.35 The findings from these papers will be discussed below.

An issue that has recently come to the light is that traditional additive force fields are not well suited to study proteins at high concentrations30,32,33 or disordered proteins.34,35 Three solutions to these problems have been proposed: to strengthen the interaction between proteins and water by selectively strengthening the Lennard-Jones (LJ) interactions,36 to strengthen the water–water dispersion interaction34 or to weaken the Lennard-Jones interaction between the protein atoms.37,38 Best et al. proposed to replace the water model to use in conjunction with the changed LJ interactions in the modified Amber force fields36 to TIP4P/200539 because the standard TIP3P model40 is known to have too fast kinetics.41 To validate this combination of the protein force field (ffam-ws) and water model, the free energy of hydration \( \Delta G_{\text{hyd}} \) of the amino acid side chains was evaluated alongside the diffusion of amino acids in TIP4P/2005.42 It was found that the ffam-ws + TIP4P/2005 combination is a good compromise that predicts \( \Delta G_{\text{hyd}} \) as well as diffusion constants reasonably well, but not perfectly.42 Nevertheless, we adopted this combination of models for the work here, after we evaluated it against reducing dispersion coefficient within proteins and organic liquids.38 A further route was taken by von Bülow et al., who used a recent Amber99 variant (Amber99SB*-ILDN-Q33–46) in combination with the dispersion-corrected TIP4P water model.34 As a side note, we would like to stress that evaluating force field predictive power using protein simulations is likely less efficient than using organic liquids 37,47–49

Using state-of-the-art computers and efficient simulation software,50,51 we report here on simulations in atomistic detail to evaluate biomolecular properties in conditions reminiscent of a physiological concentration. Systematic studies of three proteins are presented at increasing protein concentration and in three replicates. Mobility of biomolecules and water is analyzed as a function of protein volume fraction and related to the viscosity of the systems. The results are rationalized in terms of overall and internal motion.10

**METHODOLOGY**

**Molecular Dynamics (MD) Setup.** Three proteins were chosen (Figure 1) for which experimental rotational data are available from in vitro experiments, ubiquitin52 (76 amino acids, total charge 0), 2k5753 (55 amino acids, total charge \(-3 e\)), and 2kim54 (102 amino acids, total charge \(+7 e\)). The proteins were simulated (using three replicas) at four different concentrations using either 1, 2, 4, or 8 proteins in the same cubic box, which was then filled with water. For ubiquitin, a further three simulations were performed with 64 proteins at the same concentration as the eight-protein simulation. The highest protein concentration corresponds to a water mass fraction of less than 75%. An ionic strength of 0.15 M was obtained by adding Na\(^+\) and anion Cl\(^-\) to each box (Table 1).

![Structure of proteins](image-url)

**Figure 1.** Structure of proteins (A) 2k57, (B) ubiquitin, and (C) 2kim.

---

**Table 1.** Content of the Simulation Boxes as a Function of the Number of Proteins

<table>
<thead>
<tr>
<th></th>
<th>one protein</th>
<th>two proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOL Na(^+)/Cl(^-)</td>
<td>C</td>
<td>SOL Na(^+)/Cl(^-)</td>
</tr>
<tr>
<td>ubiquitin</td>
<td>16 422</td>
<td>46/46</td>
</tr>
<tr>
<td>2k57</td>
<td>10 871</td>
<td>34/31</td>
</tr>
<tr>
<td>2kim</td>
<td>23 549</td>
<td>66/73</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>four proteins</th>
<th>eight proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOL Na(^+)/Cl(^-)</td>
<td>C</td>
<td>SOL Na(^+)/Cl(^-)</td>
</tr>
<tr>
<td>ubiquitin</td>
<td>15 245</td>
<td>46/46</td>
</tr>
<tr>
<td>2k57</td>
<td>10 006</td>
<td>43/31</td>
</tr>
<tr>
<td>2kim</td>
<td>21 955</td>
<td>66/94</td>
</tr>
</tbody>
</table>

\(^a\)SOL refers to the number of waters and C the protein concentration (mM).
For all simulations, the Amber99SB-ws force field was used in combination with the TIP4P/2005 water model. Electrostatic interactions were treated using the particle mesh Ewald algorithm and Lennard-Jones interactions with a cut-off of 1 nm. All covalent bonds were constrained at their equilibrium length using the LINCS algorithm, allowing an integration time step of 2 fs. The temperature was controlled at 310 K using the v-rescale algorithm and a coupling time of 0.5 ps. The pressure was controlled at 1 bar using the Parrinello–Rahman algorithm with a time constant of 10 ps. Thirty-nine simulations of 1 μs each were performed on parallel computers using the GROMACS software.

**Hard Sphere Conformational Sampling.** Conformations of nonoverlapping “ubiquitin-like” hard spheres (HSs) were generated for N = 2, 4, or 8 particles. Two radii of gyration Rg = 1.2 and 1 nm were used as the sphere size, the average sizes of the simulation boxes were 7.96, 7.94, and 7.91 nm (box edge, respectively). Then, random coordinates for N HS particles were generated in cubic periodic boxes with the edge indicated. This was done 100 000 times to generate a reasonably sized sample. Distance analysis was performed as described below for the MD simulations.

**Translational Diffusion Coefficient.** A mean square displacement (MSD) analysis was performed to calculate the translational diffusion coefficient. The diffusion coefficients were extracted by a linear fit to the MSD by averaging blocks with a length of 100 ns. Errors in the diffusion coefficients were estimated by bootstrapping. The diffusion coefficient needs to be corrected for finite-size effects through eq 1

$$D_{\text{PBC}} = D_0 = \frac{2.837 k_BT}{6 \pi \eta L} \approx D_0 = 6.44 \times 10^{-22} / \eta L$$

where η is the viscosity, kBT = 1.38 × 10^{-20} m^2 kg/(s^2 K), T = 310 K, and L = 7, 8, 9 × 10^{-9} m, respectively, for 2k57, ubiquitin, and 2kim. Here, the shear viscosity is derived for each simulation through the Einstein relation,

$$\eta = \lim_{t \to \infty} \frac{1}{2 k_B T} \frac{d}{6} \sum_{i=1}^{N} \sum_{\beta=1, \alpha \neq \beta} \left( \int_{t_0}^{t_0+t} P_{i\beta}(t') dt' \right)^2 \chi_{i\beta}$$

where V is the volume of the simulation box and P_{i\beta} is the off-diagonal components of the pressure. Statistics is improved by taking the average over all six off-diagonal pressure tensor components.

**Rotational Correlation Time and Internal Dynamics.** To estimate the rotational correlation time (global tumbling), each trajectory of a protein was fitted rotationally and translationally to the initial structure. From these trajectories, the autocorrelation functions C(t) were calculated for all backbone amide (NH) vectors r_i using

$$C(t) = \int_0^\infty P_z(r(t) \cdot r(0)) dt$$

where P_z is the second-order Legendre polynomial, P_2(x) = (1/2)(3x^2 - 1). Subsequently, the order parameter (S_z^2) and the internal correlation time (τ_i) were derived by fitting to the function

$$C(t) = S_z^2 + (1 + S_z^2) \exp(-t/\tau_i)$$

for each NH vector r_i. These two values (S_z^2, τ_i) were averaged over all NH vectors to obtain ⟨S_z^2⟩ and ⟨τ_i⟩, and these were used in turn as the initial guess for calculating the overall tumbling time τ_M by fitting the averaged correlation function

$$\langle C(t) \rangle = \langle (S_z^2) + (1 + S_z^2) \exp(-t/\tau_i) \rangle \exp(-t/\tau_M)$$

(5)

to the autocorrelation function of the original trajectory averaged over the whole protein. This decomposition assumes that the overall tumbling is isotropic, which need not be the case; however, this assumption is applied commonly. The estimated overall tumbling time τ_M together with order parameters S_z^2 and internal correlation times τ_i obtained initially were then used to minimize χ^2 defined as

$$\chi^2 = \sum_{i=1}^{N} \left[ C_i(t) - \langle S_z^2 \rangle - (1 + \langle S_z^2 \rangle) \exp(-t/\tau_i) \exp(-t/\tau_M) \right]^2$$

(6)

with respect to the 2N + 1 parameters S_z^2, τ_i, and τ_M. The tumbling time that is proportional to the rotational diffusion coefficient through D_0^\text{rot} = 1/6τ_M needs to be corrected for finite-size effects through eq 7

$$D_{\text{PBC}}^\text{rot} = D_0^\text{rot} = \frac{k_B T}{6 \pi \eta V} \approx D_0^\text{rot} - 7.13 \times 10^{-22} / (\eta V)$$

(7)

**RESULTS**

**Protein Stability.** The mean square fluctuation of the proteins was analyzed at the residue level to see the effect of crowding on the stability of proteins (Figure 2). In the case of 2k57 and ubiquitin, the structures of proteins are intact and

---

**Figure 2.** Mean square fluctuations of each protein at the residue level for different concentrations for (A) 2k57, (B) ubiquitin, (C) 2kim, and (D) 2kim (using Amber f99SB-ILDN). All plots except (D) are averages over three replica simulations of 1 μs. The shaded area shows the standard error.
also the internal dynamics of proteins are not affected by crowding (Figure 2A,2B). However, for 2kim, the region from residue 30 to 50, containing a loop and an α-helix, unfolds to some extent in the simulations at a low protein concentration (Figure 2C). Upon comparing the simulation of 2kim using the amber99sb-ws force field with a simulation using the standard Amber ff99SB-ILDN force field,43,45,64 we note that no unfolding occurs with the original force field (Figure 2C,D) as visualized in Figure 3. This may indicate that the water−protein interactions were strengthened too much in theflamm-ws force field.36 This notion is consistent with the finding that there is less unfolding as the water concentration decreases (Figure 2C).

**Internal Protein Dynamics.** The internal dynamics of the proteins was analyzed based on the motions of the backbone NH vectors using the Lipari and Szabo analysis10 (eq 4). Figure 4 shows the average order parameters $S^2$ for the three proteins as a function of concentration. In conjunction with this, Figure 5 shows the internal relaxation times $\tau_e$. The plots show the typical shape where residues in loop regions are more flexible with low $S^2$ (Figure 4) and high $\tau_e$ (Figure 5) than those in stable secondary structure elements. For ubiquitin, the region around residue 50 is more labile at low protein concentration than at high protein concentration (Figures 4B and 5B), this is, however, not apparent in the root mean square fluctuations (RMSFs) (Figure 2B) that are low at all ubiquitin concentrations. The same region, around residue 30−50, in 2kim that showed large RMSF (Figure 2C) also has low order parameters (Figure 4C) and slow internal tumbling (Figure 5C).

**Overall Mobility.** The shear viscosity $\eta$, the average translational diffusion constants $D_T$ for the proteins, and the overall tumbling times $\tau_M$ for each system are tabulated in Table 2. The viscosity in the single protein simulations, at low concentrations, is very similar as expected, at around $1.3 \times 10^{-3} \text{ kg/(m s)}$. For reference, the viscosity for TIP4P/2005 has been estimated to be $0.855 \times 10^{-3} \text{ kg/(m s)}$,35 and the addition of protein and salt increases the viscosity rather drastically to about a factor of 7−9 going from pure water to physiological protein concentration (Table 2). The viscosity is notoriously difficult to compute because the pressure fluctuations converge slowly.61 As a result, the computed $\eta$ for two ubiquitins and for two or four 2kim proteins are too low despite averaging over three replicas of 1 μs. This has an impact on the protein diffusion constants for these systems through eq 1 (Table 2), but the trends in Figure 6A are clear anyway.
To be able to compare the dynamics for different proteins, translational diffusion constants $D_p$ were normalized to be one at infinite dilution. $D_p/D_{p,0}$ is displayed as a function of protein volume fraction in Figure 6B. At a biomolecular mass fraction of more than 20%, the diffusion coefficient is reduced to 25–38% of dilute conditions (25% for 2k57, 38% for ubiquitin, and 35% for 2kim). For comparison, the gray line in Figure 6B shows the estimated decline in the diffusion coefficient for hard spheres (HSs) modeled according to the Enskog theory:

$$D_{HS} = D_{HS,0}(1 - \phi)^3/(1 - 0.5\phi^2)$$  \hspace{1cm} (8)

where $\phi$ is the protein volume fraction.

The average rotational tumbling times $\tau_M/\tau_{M,0}$ and corresponding uncertainties were obtained from all of the proteins in each simulation and are displayed as a function of biomolecular mass fraction in Figure 6C. At the highest concentration, we find that the global tumbling time $\tau_M$ increases by a factor of 21.9 ± 4.1 for 2k57, 11.5 ± 3.4 for 2kim, and 11.5 ± 3.4 for ubiquitin.

<table>
<thead>
<tr>
<th>protein</th>
<th>exper.</th>
<th>1 copy</th>
<th>2 copies</th>
<th>4 copies</th>
<th>8 copies</th>
<th>64 copies</th>
</tr>
</thead>
<tbody>
<tr>
<td>2k57</td>
<td></td>
<td>1.4 ± 0.7</td>
<td>2.6 ± 0.3</td>
<td>4.7 ± 2.8</td>
<td>6.0 ± 2.2</td>
<td></td>
</tr>
<tr>
<td>ub</td>
<td></td>
<td>1.1 ± 0.6</td>
<td>1.2 ± 0.5</td>
<td>3.6 ± 1.8</td>
<td>6.3 ± 2.5</td>
<td>5.3 ± 2.3</td>
</tr>
<tr>
<td>2kim</td>
<td></td>
<td>1.3 ± 0.7</td>
<td>0.8 ± 0.7</td>
<td>1.0 ± 0.7</td>
<td>7.7 ± 4.3</td>
<td></td>
</tr>
<tr>
<td>2k57</td>
<td></td>
<td>0.09 ± 0.02</td>
<td>0.08 ± 0.01</td>
<td>0.06 ± 0.01</td>
<td>0.015 ± 0.001</td>
<td></td>
</tr>
<tr>
<td>ub</td>
<td></td>
<td>0.08 ± 0.04</td>
<td>0.08 ± 0.02</td>
<td>0.05 ± 0.01</td>
<td>0.026 ± 0.002</td>
<td>0.035 ± 0.002</td>
</tr>
<tr>
<td>2kim</td>
<td></td>
<td>0.06 ± 0.01</td>
<td>0.06 ± 0.01</td>
<td>0.04 ± 0.01</td>
<td>0.020 ± 0.003</td>
<td></td>
</tr>
<tr>
<td>2k57</td>
<td></td>
<td>5.1</td>
<td>2.9 ± 0.5</td>
<td>3.4 ± 0.0</td>
<td>6.8 ± 0.7</td>
<td>63.5 ± 4.3</td>
</tr>
<tr>
<td>ub</td>
<td></td>
<td>4.4</td>
<td>3.2 ± 0.4</td>
<td>3.9 ± 0.3</td>
<td>10.8 ± 2.4</td>
<td>36.7 ± 10.0</td>
</tr>
<tr>
<td>2kim</td>
<td></td>
<td>8.05</td>
<td>6.2 ± 0.9</td>
<td>7.2 ± 2.4</td>
<td>20.4 ± 2.3</td>
<td>41.1 ± 7.0</td>
</tr>
</tbody>
</table>

“*The diffusion coefficient is corrected for finite size effect according to eq 1. Experimental in vitro measurements of the global tumbling times are shown in the first column for comparison.*"
ubiquitin, and 6.6 ± 1.5 for 2kim. We note that the spread in the results, which is different from the protein translational diffusion, is compatible with NMR experiments.\(^6\) Mean square displacement plots of simulations of 2, 4, 8, and 64 ubiquitins are provided in the Supporting Information. Simulations of 8 and 64 proteins have the same protein concentration, and the fact that the results of both simulations are similar shows that the size of the systems is sufficient.

**Water Diffusion.** The water diffusion coefficient \(D_w\) is reduced by about 40% as the protein volume fraction increases to that corresponding to physiological conditions (Figure 6D). The HS model overestimates the reduction of the diffusion coefficient for water under crowding conditions slightly.

**Protein–Protein and Protein–Water Interactions.** Protein–protein hydrogen bonds were counted in all simulations with more than one protein, to test for clustering (Table 3). A geometric criterion was used where a hydrogen bond was counted when the donor–acceptor distance was less than or equal to 30°.\(^68,69\) In none of the cases is the number of hydrogen bonds larger than 1 on average, indicating that direct protein–protein interactions are infrequent and transient only.

Intramolecular hydrogen bonds were evaluated using the same criterion to investigate concentration dependence (Table 4). 2k57 and ubiquitin have a number of hydrogen bonds that do not depend on the concentration and has a low standard deviation, indicating a stable secondary structure. For 2kim, the numbers fluctuate significantly, and despite high standard deviation, there are three to four fewer hydrogen bonds in the single protein simulations than in the simulations at higher concentrations.

The number of protein–water hydrogen bonds was determined per protein as well (Table 5). For 2k57 and ubiquitin, the numbers are almost identical at all concentrations. For ubiquitin, there is a slight drop in the number of hydrogen bonds with concentration only, but for 2kim, the picture is more complex. Although there are no intermolecular protein–protein interactions (Table 3), the number of protein–water hydrogen bonds varies more for 2kim than for the other proteins due to again the partial unfolding that is also shown in Figures 2C and 3. The high number of protein–water hydrogen bonds in single-molecule 2kim simulations coincides with a reduced number of intramolecular hydrogen bonds (Table 4).

Radial distribution functions (RDFs) are provided in Figures S6–S8. The plots give center-of-mass distance RDFs for simulations of two and eight proteins. Except for one simulation of two proteins of 2kim (Figure S8), the probability of finding proteins close to each other is larger at the highest concentration (eight proteins) than at the lower concentration (two proteins), which is logical from the perspective of reduced volume available to the proteins at higher concentrations.

**Distance Analysis.** We analyzed distances between protein heavy atom at different concentrations for different cut-off distances equal to 0.27, 0.57, and 0.87 nm in the same manner as Nawrocki et al.\(^70\) We considered distances at 0.27 nm to correspond to direct interactions and 0.57 and 0.87 nm for the first and second hydration layers, respectively. Increasing the cut-off leads to a significantly larger fraction of clustered proteins. A comparison with a ubiquitin-like hard sphere model (Figure 7) shows that the clustering cannot be explained by excluded volume as the largest clusters in the MD simulation are populated to much larger fractions than the HS model. We note that the radius used for the hard spheres was taken somewhat ad hoc as the radius of gyration. Using a smaller radius, corresponding to a lower concentration, leads to smaller and fewer protein clusters (Figure 7). The analysis at the shortest distance, 0.27 nm, corresponds to a strong hydrogen bond or salt-bridge. Apparently, this underestimated the clustering significantly, and therefore the hydrogen bonding analysis (Table 3) is not sufficient to draw conclusion about clustering behavior. Rather than direct interactions, long-range dispersion interactions between the proteins\(^58,71\) or simply the hydrophobic effect\(^25\) may influence the clustering. Figures S17–S19 show that the clustering occurs rapidly during the time of simulations and is reproducible in the three replicas of the protein simulations.

**DISCUSSION**

It has been stated that water determines the structure and dynamics of proteins.\(^25\) However, this picture is biased by the fact that most biochemical experiments are performed under dilute conditions.\(^58\) Protein properties depend on their environment, for instance, through electrostatic interactions with other constituents of that environment\(^75,76\) and the charge distribution on biomolecular surfaces has been suggested to affect macromolecular dynamics.\(^69\) Macromolecular crowding has also been reported to alter the translational and rotational diffusion of biomolecules in experimental studies.\(^12,75,76\) In this work, we employed molecular dynamics simulations to study proteins at high concentrations to mimic crowding effects. It is well established that MD simulations allow us to probe protein–water interactions in spatial and temporal details (see

### Table 3. Number of Intermolecular Protein–Protein Hydrogen Bonds Averaged over the Last 500 ns of the Simulations\(^a\)

<table>
<thead>
<tr>
<th>protein</th>
<th>two copies</th>
<th>four copies</th>
<th>eight copies</th>
</tr>
</thead>
<tbody>
<tr>
<td>2k57</td>
<td>0.4 ± 0.1</td>
<td>0.5 ± 0.4</td>
<td>0.5 ± 0.7</td>
</tr>
<tr>
<td>ubiquitin</td>
<td>0.7 ± 0.5</td>
<td>0.5 ± 0.5</td>
<td>0.6 ± 0.8</td>
</tr>
<tr>
<td>2kim</td>
<td>1.1 ± 1.0</td>
<td>0.6 ± 0.4</td>
<td>0.6 ± 0.6</td>
</tr>
</tbody>
</table>

\(^a\)Average and standard deviation over three replicas.

### Table 4. Number of Intramolecular Protein–Protein Hydrogen Bonds Averaged over the Last 500 ns of the Simulations\(^a\)

<table>
<thead>
<tr>
<th>protein</th>
<th>one copy</th>
<th>two copies</th>
<th>four copies</th>
<th>eight copies</th>
</tr>
</thead>
<tbody>
<tr>
<td>2k57</td>
<td>37.0 ± 1</td>
<td>36.4 ± 2</td>
<td>35.2 ± 2</td>
<td>36.7 ± 1</td>
</tr>
<tr>
<td>ubiquitin</td>
<td>55.5 ± 2</td>
<td>55.1 ± 2</td>
<td>55.2 ± 2</td>
<td>54.8 ± 2</td>
</tr>
<tr>
<td>2kim</td>
<td>53.1 ± 6</td>
<td>58.1 ± 4</td>
<td>55.7 ± 5</td>
<td>57.2 ± 4</td>
</tr>
</tbody>
</table>

\(^a\)Average and standard deviation over three replicas.

### Table 5. Number of Protein–water Hydrogen Bonds Averaged over the Last 500 ns of the Simulations\(^a\)

<table>
<thead>
<tr>
<th>protein</th>
<th>one copy</th>
<th>two copies</th>
<th>four copies</th>
<th>eight copies</th>
</tr>
</thead>
<tbody>
<tr>
<td>2k57</td>
<td>167 ± 2</td>
<td>167 ± 3</td>
<td>168 ± 4</td>
<td>163 ± 2</td>
</tr>
<tr>
<td>ubiquitin</td>
<td>186 ± 3</td>
<td>186 ± 3</td>
<td>185 ± 2</td>
<td>183 ± 4</td>
</tr>
<tr>
<td>2kim</td>
<td>257 ± 15</td>
<td>242 ± 11</td>
<td>249 ± 12</td>
<td>242 ± 9</td>
</tr>
</tbody>
</table>

\(^a\)Average and standard deviation over three replicas.
e.g., ref 77). Furthermore, MD simulations allow us to analyze the internal dynamics78 as well as the diffusive properties of proteins62,79 in comparison with, in particular, nuclear magnetic resonance (NMR) experiments. The choice of proteins was guided here by the availability of the experimental data for protein tumbling.

Simulations of highly concentrated systems with conventional force fields have been reported to lead to unphysical aggregation of biomolecules due to strong protein−protein interactions.32,33 Here, we used the force field by Best et al.36 that has been tuned for simulations of high concentrations of macromolecules, by strengthening the protein−water Lennard-Jones interactions. The force field is used in conjunction with the TIP4P/2005 water model 39 to obtain better kinetic properties. The combination was validated in two additional papers, one on hydration free energy of amino acids and diffusion of amino acids42 and one comparing the Best et al. force field36 to other approached for shifting the relative strength of interactions toward stronger protein−water interactions. 38 As noted in the introduction, alternative combinations of water models and protein force fields have been applied as well.28,31

The translational mobility of three small proteins, 2k57 (55 AA), 1ubq (76 AA), and 2kim (102 AA), are found here to be 6 times slower for 2k57, 3.1 times slower for ubiquitin, and 3.0 times slower for 2kim at a physiologically relevant protein volume fraction compared to the dilute conditions (Figure 6B). In other words, the slow-down of translational diffusion is more than the factor of two reported for the mini-protein Villin by Nawrocki et al.30 We do find a slow-down of rotational tumbling (Figure 6C) comparable to that reported by Nawrocki et al. for Villin.50 Protein concentration is an important factor and our simulations were performed at a somewhat higher concentration than those by Nawrocki30 or von Bülow,31 and therefore we find stronger retardation of mobility. Experimentally, a range of 4–17 times slow-down of translational diffusion has been reported.17,19,20,80,81 The fact that our simulation results are at the lower end of that range could be because there are only weak protein−protein interactions in our simulations (Figure 7). The clustering observed may explain that the rotational tumbling is retarded more than translational diffusion because direct interactions will hinder rotational motion. A cellular environment is crowded by molecules in a wide range of sizes and charge distributions, and transient interactions between cellular components could affect diffusional properties,76 in line with the result on the Villin protein.30 The slow-down in diffusion found in this work is correlated to the increase in viscosity (Figure 6A), although it is known that the Stokes−Einstein relation between diffusion and viscosity holds approximately only82,83 since the particle size that enters the equation is not an exact quantity. However, recent work by von Bülow suggests that the Stokes−Einstein relation may be applied to proteins in solution as well.71

The rotational diffusion slow-down for the different proteins (Figure 6C) has a large spread, a factor of 22 for 2k57, 11 for ubiquitin, and 7 for 2kim. These results are compatible with the NMR data that show that translational- and rotational diffusion may or may not be coupled, depending on the protein.77 For 2k57, the large increase in $\tau_M$ (Figure 6C and Table 2) can likely not be explained by the somewhat reduced hydration (Table S), as was noted for protein mobility in a crowded glucose solution,26 but rather by weak interactions as discussed above. It has been noted before that the hydration properties may differ between proteins of different sizes, leading to different diffusion properties.28,84 Harada et al.
suggest a relationship between protein—water interactions and viscosity, based on the Stokes—Einstein equation.28 Our results suggest that the increase of viscosity with protein concentration is very similar to the three proteins studied here (Figure 6A). Indeed, the increase of the viscosity by a factor of 7–9 is very similar to the factor of 8 reported from in vivo experiments.26 This suggests that the slow-down in rotation has another cause, e.g., the charge distribution on the surface or weak dispersion interactions. We find that the rotational diffusion of the proteins in our study seems to be size-dependent, unlike the translational diffusion and viscosity (Figure 6A–C). At the highest concentration of proteins, small proteins are slowed down more than big proteins for an unknown reason. The effect of protein–protein interactions on translational and rotational diffusion has previously been investigated by the addition of high concentrations of either synthetic polymers or protein crowders to dilute protein solutions as an attempt to mimic the cellular environment.12 These results showed that synthetic polymers led to translational diffusion being retarded more than rotational diffusion. However, using protein crowders or cell lysate, the rotational diffusion was retarded more than translational diffusion.12 These results seem to agree with the findings in this work (Figure 6B, C).

The differences between our simulations and the HS model predictions for diffusion are small for both proteins (Figure 6B) and water (Figure 6D). The reasonable correspondence between full atomistic MD and HS models suggests that the excluded volume explains the drop in diffusion to some extent,24 however, interactions are important as well (Figure 7). Indeed, Brownian dynamics simulations, where only steric effects are taken into account, form a simple tool to test Enskog predictions for the effects of crowding. It has been shown that the diffusion of simple hard sphere molecules slows down as a function of molecular volume occupation in agreement with Enskog theory.85 Hard sphere Brownian dynamics simulation coupled with hydrodynamic interactions can in principle be tuned to give a good estimate on slow-down of translational diffusion as a function of crowding; however, these techniques cannot give insight into the role of explicit waters or into the rotational mobility of biomolecules.

McGuffee and Elcock studied the effects of crowding in a cytoplasmic model system of E. coli containing the 50 most abundant proteins, but without explicit water.53 Steric forces with and without electrostatic interactions cause different dynamics as shown by the difference in the diffusion coefficient of GFP as one of the constituents of their model, however, the GFP diffusion coefficient predicted by that model23 was still not correct when compared to measurements.7,19,20 The addition of an intramolecular interaction by Lennard-Jones potential was used to obtain a correct diffusion coefficient of GFP.23,24,25 However, for this, the Lennard-Jones parameter had to be exaggerated since the role of water was neglected.27,28 Skolnick highlights in a review the importance of hydrodynamic interactions for understanding biomolecular mobility under crowding conditions. However, these models predict, for instance, that the decay of translational diffusion is size-dependent, something that seems to be at odds with our findings (Figure 6B) although the range of protein sizes studied here is limited and crowders of different sizes and chemical composition can have different effects. It should be noted that, in general, hydrodynamic interactions need to be considered only when no explicit solvent is present.86

It has been shown that proteins become more stable under crowding conditions effectuated through inert synthetic polymers, where only the steric effects are important.99–92 However, when doing the measurement in a cellular-like environment, it was found that electrostatic interactions destabilize protein structures when opposite charges on the surface attract each other.93 In some cases, electrostatic interactions may stabilize the structures due to repulsive forces.3,94,95 These findings emphasize that the effect of direct intermolecular interactions on stability is not negligible. Few intermolecular protein–protein hydrogen bonds were found in this study (Table 3), but solvent-separated interactions are important (Figure 7), and there is an effect of protein concentration on protein stability (Figure 2). It cannot be excluded that due to the force field used for this study, the solute–solvent interactions are strengthened so much that the proteins have the tendency to bind to water more than to the other proteins and, as a result, it seems that the region containing a loop and α-helix (residue 30–50) of the 2kim protein is destabilized (Figure 2C). This behavior was not observed when the original Amber ff99SB-ILDN was used (Figures 2D and 3). Interestingly, the stability increases when the protein concentration goes up for both 2kim (Figure 2C) and ubiquitin (Figure 4B), in line with the experimental studies mentioned.

Without a doubt, the relationship between protein stability and the environment or protein dynamics and the environment is very complex.25,54,70 Although MD simulations are ideally suited to study large and complex systems of biomolecules, be it virus particles90–99 or crowding effects77,28 in atomistic detail, it will take considerable effort to tune force fields to become completely transferable between different chemical environments.100–105 It may be necessary to consider long-range dispersion interactions explicitly37,38,55,102 or to include higher-order terms of the London dispersion forces to obtain more accurate simulations.105 However, to study the effect of crowding without altering the van der Waals parameters, using a model with higher-order dispersion coefficients101 or a polarizable force field104,105 might give results closer to experimental values, even though conventional force field improvement is still on-going as well.106

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.9b02835.

Mean square displacement of proteins as a function of time (Figures S1–S13); protein–protein radial distribution functions (Figures S14–S16); and maximum protein–protein cluster size as a function of time (Figures S17–S19) (PDF)

AUTHOR INFORMATION

Corresponding Author
E-mail: david.vanderspoel@icm.uu.se. Phone: +46 18 4714205.

ORCID

Haiyang Zhang: 0000-0002-2410-7078
David van der Spoel: 0000-0002-7659-8526
### Funding
The Swedish research council is acknowledged for financial support to DvdS (grant 2013-5947) and for a grant of computer time (SNIC2017-12-41) through the parallel computer center, Stockholm, Sweden. H.Z. acknowledges computer time (SNIC2017-12-41) through the parallel support to DvdS (grant 2013-5947) and for a grant of financial assistance.

### Notes
The authors declare no competing financial interest.

### REFERENCES


