Dynamic cluster formation determines viscosity and diffusion in dense protein solutions

Sören von Bülow[a], Marc Siggel[a], Max Linke[a], and Gerhard Hummer[a,b,1]

[a]Department of Theoretical Biophysics, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany; and [b]Department of Physics, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany

We develop a detailed description of protein translational and rotational diffusion in concentrated solution on the basis of all-atom molecular dynamics simulations in explicit solvent. Our systems contain up to 540 fully flexible proteins with 3.6 million atoms. In concentrated protein solutions (100 mg/mL and higher), the proteins ubiquitin and lysozyme, as well as the protein domains third IgG-binding domain of protein G and villin headpiece, diffuse not as isolated particles, but as members of transient clusters between which they constantly exchange. A dynamic cluster model nearly quantitatively explains the increase in viscosity and the decrease in protein diffusivity with protein volume fraction, which both exceed the predictions from widely used colloid models. The Stokes–Einstein relations for translational and rotational diffusion remain valid, but the effective hydrodynamic radius grows linearly with protein volume fraction. This increase follows the observed increase in cluster size and explains the more dramatic slowdown of protein rotation compared with translation. Baxter's sticky-sphere model of colloidal suspensions captures the concentration dependence of cluster size, viscosity, and rotational and translational diffusion. The consistency between simulations and experiments for a diverse set of soluble globular proteins indicates that the cluster model applies broadly to concentrated protein solutions, with equilibrium dissociation constants for nonspecific protein–protein binding in the $K_d \approx 10$-mM regime.

MD simulation | protein crowding | diffusion | viscosity | dynamic clusters

The interior of cells is a densely crowded medium, in which macromolecular concentrations range from 90 mg/mL in red blood cells to 300 mg/mL in the mitochondrial matrix (1, 2). Macromolecular crowding influences the stability of proteins, reaction rates, the catalytic activity of enzymes, protein–protein association, and diffusion (3–13). Excluded volume through steric repulsion (14) and attractive protein–protein interactions as well as hydrodynamic interactions affect protein diffusion (6, 15–19). To address the influence of specific protein–protein interactions on protein diffusion (20), crowded solutions with proteins serving as both agents and readout have been studied (5, 14, 16, 21–30).

Experimental techniques to study the effects of macromolecular crowding on diffusion (15) include tracer boundary spreading (14), light scattering spectroscopy (31), fluorescence recovery after photobleaching (FRAP) (32–34), electron spin resonance (35), single-particle tracking (36), fluorescence correlation spectroscopy (FCS) (37–39), quasielastic neutron backscattering (27, 40), and NMR spectroscopy (24, 41, 42). Particle-based simulations complement these experiments (15), treating the proteins as spheres or ellipsoids (20, 43, 44), with residue-level coarse graining (45–47), or as rigid all-atom models (16, 48). Hydrodynamic interactions contribute significantly to the slowdown of protein diffusion in crowded environments (19). In implicit solvent, they are ignored or approximated via the diffusion tensor (16, 19, 44, 49).

Rapid advances in computing hardware and simulation algorithms have opened up the opportunity to study macromolecular crowding using atomistic molecular dynamics (MD) simulations. Explicit solvent accounts directly for excluded volume effects and hydrodynamic interactions and mediates short-range attractive and long-range electrostatic protein–protein interactions (5, 28–30, 50–52). Here, we use atomistic MD simulations of dense protein solutions to calculate the viscosity and protein diffusion coefficients as a function of protein concentration (Fig. 1). Ubiquitin (UBQ), the third IgG-binding domain of protein G (GB3), hen egg white lysozyme (LYZ), and villin headpiece (VIL) are used as model proteins.

Soluble proteins self-associate in concentrated solution to form transient and dynamic clusters (19, 24, 53–58). Clustering has also been reported for membrane proteins (59). The influence of cluster formation on the protein translational and rotational diffusivity has recently been addressed by atomistic simulation studies (29, 30). Here, we build on these findings and put cluster formation in the framework of the Stokes–Einstein relations connecting viscosity, cluster size, and diffusion. Central questions are (i) whether the Stokes–Einstein relations remain valid in concentrated protein solutions (60), (ii) how transient protein interactions affect the diffusivity and apparent hydrodynamic radii of proteins in concentrated solutions (42), (iii) how viscosity depends on protein concentration (39, 61, 62), and (iv) whether colloid models apply to concentrated protein solutions.

To address these questions, we perform extensive MD simulations, develop a cluster model of concentration-dependent

Significance

For living cells to function, proteins must efficiently navigate the densely packed cytosol. Protein diffusion is slowed down by high viscosity and can come to a complete halt because of nonspecific binding and aggregation. Using molecular dynamics simulations, we develop a detailed description of protein diffusion in concentrated protein solution. We confirm that soluble proteins in concentrated solutions diffuse not as isolated particles, but as members of transient clusters between which they constantly exchange. Nonspecific protein binding and the formation of dynamic clusters nearly quantitatively account for the high viscosity and slow diffusivity in concentrated protein solutions, consistent with the Stokes–Einstein relations.

Fig. 1. Representative simulation snapshots of dense UBQ solution (200 mg/mL, $N = 405$) and GB3 solution (100 mg/mL, $N = 20$). Blue lines indicate the periodically replicated simulation boxes. Proteins are depicted as green surface (UBQ) or green ribbons (GB3); Na$^+$ and Cl$^-$ ions as blue and cyan balls, respectively; and TIP4P-D water as red sticks.

protein diffusion, and compare our results to the theoretical predictions for hard-sphere (HS) colloidal suspensions without and with attractive interactions. We put our findings in the context of a wide range of experimental and simulation studies and obtain a remarkably consistent picture of the diffusive dynamics in concentrated solutions of soluble globular proteins.

Results

Shear Viscosity Increases Strongly with Protein Volume Fraction. The viscosity of concentrated protein solutions and of TIP4P-D solvent at different ion concentrations was calculated from the pressure tensor fluctuations (*SI Appendix*, Fig. S5). The quadratic function Eq. **3** fits $\eta(\phi)$ well over the simulation range, $\phi \leq 0.15$ (Fig. 2, squares). We obtained similar fit parameters b for UBQ, GB3, and VIL solutions, whereas b of the LYZ solutions is significantly lower (Table 1). We calculated $\eta_0 = 0.937$ mPa·s from the average viscosity of solutions of TIP4P-D (65) water with 120 mM, 157 mM, and 200 mM NaCl. The values of the calculated LYZ viscosities are in remarkably good agreement with experimentally determined viscosities of the same protein (61). For all four proteins, the viscosities calculated from the pressure tensor fluctuations are bracketed by the experimental values for BSA (62), human serum albumin (HSA), and ribonuclease A (RNaseA) (39) solutions (Fig. 2).

The viscosity of dense protein solutions exceeds the Einstein prediction $\eta = \eta_0 + 2.5\phi$ for noninteracting HS colloids (66) even after correction for high concentration (63). The pronounced increase in the viscosity with protein concentration beyond the nonlinear HS viscosity model (63) indicates that short-range attractive interactions between the proteins cannot be ignored. For colloids, the second-order term $b\phi^2$ in Eq. **3** for the viscosity is related to the attraction strength, as measured by the osmotic virial coefficient (67–69). In the following, we use $\eta(\phi)$ in Eq. **3** with values of b listed in Table 1 to account for the dependence of the viscosity on the protein volume fraction.

Translational Diffusion Slows Down at High Protein Density. As shown in Movie S1 for GB3 at 200 mg/mL with $N = 540$ proteins, translational and rotational diffusion in concentrated solution is strongly impacted by protein interactions. For each protein in the simulation box, mean-squared displacement (MSD) curves were calculated and fitted to the Einstein relation in *SI Appendix*, Eq. S5 (*SI Appendix*, Fig. S6). The MSD curves of the dense protein solutions averaged over starting times and proteins are linear at times exceeding 10 ns. The translational diffusion coefficients D_t^{PBC} obtained by fitting the Einstein relation to the MSD from 10 ns to 30 ns are therefore long-time diffusion coefficients. The MSD curves of the dilute solutions (one protein in the simulation

box) are linear at small delays and were fitted to the Einstein relation from 0 ns to 5 ns. We corrected D_t^{PBC} for large finite-size effects using Eq. **7**, where we used $\eta(\phi)$ from the quadratic fit, Eq. **3**. The values before finite-size correction are listed in *SI Appendix*, Table S3.

After finite-size correction, the translational diffusion coefficient D_t of dilute UBQ is consistent with results of NMR spectroscopy (71, 72) in dilute solution. Our calculated D_t values of LYZ are bracketed by measurements in dilute and dense solutions (21, 24, 25, 71–76). The spread in the measured diffusion coefficients of LYZ is possibly due to differences in pH value, ionic strength, and temperature in the different experiments. All simulation values of D_t in the dilute solution are very close to Hydropro (70) predictions (Fig. 3A). The translational diffusion coefficients calculated for the large systems (with $N \geq 120$ proteins) at 200 mg/mL concentration are similar to the values for the small systems ($N \leq 20$) at the same concentration, decreasing slightly with increasing box sizes (number of N). For all proteins, our dynamic cluster model Eq. **8**, developed below, accounts nearly quantitatively for the slowdown of translational diffusion with increasing concentration.

Crowding Strongly Affects Rotational Diffusion. Rotational diffusion coefficients $\bar{D}_r$ and $\tilde{D}_r$ of the dense protein solutions were obtained by fitting quaternion correlations (80) and by integrating the orientational correlation function $\langle\langle P_1(\cos\theta(t))\rangle\rangle$, respectively, over the time range 0–100 ns (*SI Appendix*). Fits to elements of the quaternion covariance matrix are shown in *SI Appendix*, Fig. S7. The resulting rotational diffusion coefficients $D_1 - D_3$ are shown in *SI Appendix*, Fig. S8. Fits to the orientational correlation function are shown in *SI Appendix*, Fig. S9.

Orientationally averaged diffusion coefficients $\bar{D}_r$ and $\tilde{D}_r$ after finite-size correction (81) decrease strongly with increasing ϕ (Fig. 3B). At infinite dilution, the UBQ and GB3 results are bracketed by the rotational diffusion coefficients obtained from Hydropro (70) calculations and from NMR spectroscopy
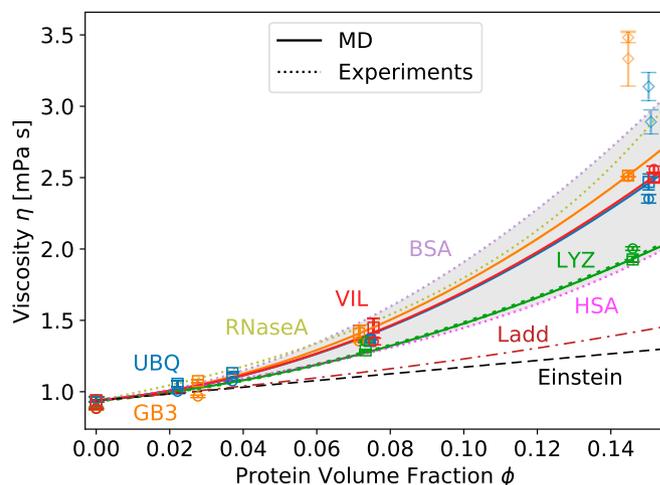
Fig. 2. Viscosity of concentrated protein solutions as a function of protein volume fraction from all-atom MD simulations (symbols), experiments (dotted lines and gray shading), and HS colloid theory (dashed lines). Blue, UBQ; orange, GB3; green, LYZ; red, VIL. The viscosity η was calculated from MD using fluctuations in the pressure tensor (*SI Appendix*, Eq. S3, squares; fit of Eq. **3**, solid lines) and the Stokes–Einstein (SE) relations (Eq. **6**; small systems $N \leq 20$, circles; large systems $N \geq 120$, diamonds). Green and purple dotted lines: fits of Eq. **3** to experimental data on concentrated LYZ (61) and BSA (62) solutions. Magenta and olive dotted lines: exponential fit [equation 4 in Zorrilla et al. (39)] to the experimental data on HSA and RNaseA solutions (39). Black dashed line: Einstein expression for HS colloidal suspensions (Eq. **4**). Brown dashed-dotted line: Ladd expression (63, 64) for HSs (Eq. **5**).

Table 1. Viscosity coefficient b, clustering propensity ζ, Baxter parameter τ, and effective dissociation constant K_d for nonspecific protein–protein binding in concentrated solutions of UBQ, GB3, LYZ, and VIL

Protein	b,mPa·s	ζ	τ^{cl}	K_d^{cl}, mM	τ^{RDF}	K_d^{RDF}, mM	τ^{D_t}	$K_d^{D_t}$, mM	$\tau^{k_{\mathrm{off}}}$	$K_d^{k_{\mathrm{off}}}$, mM	τ^{η}	K_d^{η}, mM
UBQ	55.3	8.2	0.12	19	0.13	21	0.13	21	0.49	$\sim$80	0.038	6.1
GB3	63.1	8.8	0.11	26	0.19	44	0.13	30	0.34	$\sim$80	0.033	7.6
LYZ	32.5	3.6	0.27	26	—	—	—	—	0.80	$\sim$80	0.071	6.9
VIL	56.2	11.6	0.09	28	—	—	0.13	43	0.24	$\sim$80	0.037	12.2

τ^{cl} and K_d^{cl} are from cluster sizes, τ^{RDF} and K_d^{RDF} are from radial distribution function $g(r)$, τ^{D_t} and $K_d^{D_t}$ are from $D_t'(\phi)$ at low protein concentration, $\tau^{k_{\mathrm{off}}}$ and $K_d^{k_{\mathrm{off}}}$ are from binding off rates, and τ^{η} and K_d^{η} are from viscosity term b.

(77, 78). The experimental rotational diffusion coefficient in dilute LYZ solution reported in ref. 73 is slightly lower than the calculated values, whereas the rotational diffusion coefficients reported in ref. 25 agree well with our data at all protein concentrations. The calculated rotational diffusion coefficients of dilute LYZ and VIL are in fair agreement with Hydropro (70) predictions. As for the translational diffusion, the rotational diffusion coefficients calculated for the large systems ($N \geq 120$) at 200 mg/mL concentration are close to the values of the small systems ($N \leq 20$) at the same concentration, being slightly lower for UBQ and slightly higher for GB3. The dynamic cluster model Eq. **9** predicts the rotational diffusion coefficients of UBQ, GB3, VIL, and LYZ accurately over the entire concentration range, except for the LYZ solution at 100 mg/mL concentration, where the effect of the weak clustering (Fig. 3C) is somewhat overestimated.

Diffusion in Dense Protein Solutions Follows the Stokes–Einstein Relation. Given translational and rotational diffusion coefficients, the viscosity can be estimated from the Stokes–Einstein relations (Eq. **6**). For all small systems ($N \leq 20$), we observe excellent agreement between the viscosity calculated from the autocorrelation of the pressure tensor fluctuations and from the diffusion coefficients (Fig. 2), indicating that the dense protein solutions show normal (Stokes–Einstein) diffusion for all protein volume fractions considered here. For the large systems ($N \geq 120$), slight deviations in D_t and D_r (Fig. 3) lead to strongly overestimated viscosity (diamonds in Fig. 2). Therefore, in practical calculations we advise against calculating the viscosity via the Stokes–Einstein relation, Eq. **6**, because the results are quite sensitive to the uncertainties in D_t and $\bar{D}_r$.

Hydrodynamic Radius, Cluster Size, and Diffusion Are Related. We obtained very similar hydrodynamic radii from the Stokes–Einstein relations for translation and rotation, Eqs. **10** and **11**, respectively (Fig. 3D). Therefore, after correcting for finite-size effects with actual shear viscosities $\eta(\phi)$, both translational diffusion and rotational diffusion follow Stokes–Einstein theory even at high protein volume fractions. Only values for the small systems ($N \leq 20$) are shown, because only for these systems was the computationally expensive calculation of the viscosity from the pressure fluctuation autocorrelation function performed.

If the increase in viscosity were to capture all factors that contribute to the concentration-dependent slowdown of protein diffusivity, then the hydrodynamic radius, calculated from the Stokes–Einstein relations, Eqs. **10** and **11**, should remain constant at all concentrations. Instead, we observe that the effective hydrodynamic radius cubed, $R_h{}^3$, increases with protein volume fraction ϕ. This dependence is consistent with protein cluster formation (Fig. 3C). Indeed, when calculating the cluster size distribution based on an α-carbon distance cutoff criterion, the cluster size distribution shifts to larger clusters at increasing protein volume fraction (*SI Appendix*, Fig. S10). For protein concentrations up to 100 mg/mL, the mean number of proteins in a cluster grows linearly as $\overline{m}(\phi) = 1 + \zeta\phi$, with clustering propen-

sity ζ listed in Table 1 (Fig. 3C). The highest concentration (200 mg/mL) was not included in the fit, because the close proximity of proteins causes a significant dependence of the calculated mean cluster size on the cutoff criterion (*SI Appendix*, Fig. S11). Given the linear increase of cluster size with protein volume fraction ϕ, the effective hydrodynamic radius cubed should likewise increase linearly with ϕ, $R_h{}^3(\phi) = R_{h,\phi=0}{}^3(1 + \zeta\phi)$, where we assumed that the hydrodynamic radius cubed is proportional to the cluster volume. Remarkably, the dynamic cluster model accounts nearly quantitatively for the relative increase of $R_h{}^3(\phi)$ (Fig. 3D).

As shown in *SI Appendix*, the clustering propensity ζ is related to an effective dissociation constant K_d via the protein volume v_p, $\zeta = 1/(v_p K_d)$. Given protein volumes $v_p(\mathrm{UBQ}) = 10.4$ nm^3, $v_p(\mathrm{GB3}) = 7.2$ nm^3, $v_p(\mathrm{LYZ}) = 17.2$ nm^3, and $v_p(\mathrm{VIL}) = 5.1$ nm^3 (*SI Appendix*, Fig. S12), we obtain dissociation constants of $K_d^{\mathrm{cl}}(\mathrm{UBQ}) = 1/(8.2 \times 10.4\ \mathrm{nm}^3 \times N_A) \approx 19$ mM, $K_d^{\mathrm{cl}}(\mathrm{GB3}) \approx 26$ mM, $K_d^{\mathrm{cl}}(\mathrm{LYZ}) \approx 26$ mM, and $K_d^{\mathrm{cl}}(\mathrm{VIL}) \approx 28$ mM with N_A Avogadro's constant (Table 1).

Effective Viscosity Accounts for Hydrodynamic Interactions. We investigated whether the effective viscosity $\eta(\phi)$ captures the indirect (hydrodynamic) effects of increased protein concentration and cluster formation, i.e., all of the effects that are not accounted for by an increased effective hydrodynamic radius. For each protein, we recorded the times at which it is free, i.e., not in any cluster (*SI Appendix*, Fig. S13). For these trajectory segments, we calculated MSD curves (*SI Appendix*, Fig. S14) and translational diffusion coefficients D_t^{free}. D_t^{free} could not be reliably determined for the proteins at the highest concentration (200 mg/mL) due to insufficient sampling of unbound proteins. For all four protein species, the product $\eta(\phi)D_t^{\mathrm{free}}(\phi)$ is approximately constant as a function of concentration (*SI Appendix*, Fig. S15). We conclude that the effective viscosity indeed accounts for the hydrodynamic contributions to the diffusivity slowdown.

Displacement Pair Correlation Shows Contribution from Direct and Hydrodynamic Interactions. We calculated the displacement pair correlation introduced by Ando and Skolnick (19) (*SI Appendix*). We analyzed the protein pair correlation for pairs at distances 0.6–3 nm. At distances corresponding to cluster formation, we observed highly correlated motion for all protein pairs at all concentrations and time delays (*SI Appendix*, Fig. S16). At increasing pair distance ($\sim$2–3 nm), the pair correlation decreased gradually.

Protein Binding Interfaces. The interactions between the proteins in clusters were loose but not entirely random in their orientation (*SI Appendix*, Fig. S17). For UBQ, the preferred binding interface coincides remarkably well with the noncovalent dimer interface reported from NMR measurements (82), more or less independent of protein concentration (Fig. 4). It includes residues 8–11, 20, 24, 25, 28, 31–42, 46–49, 54–60, and
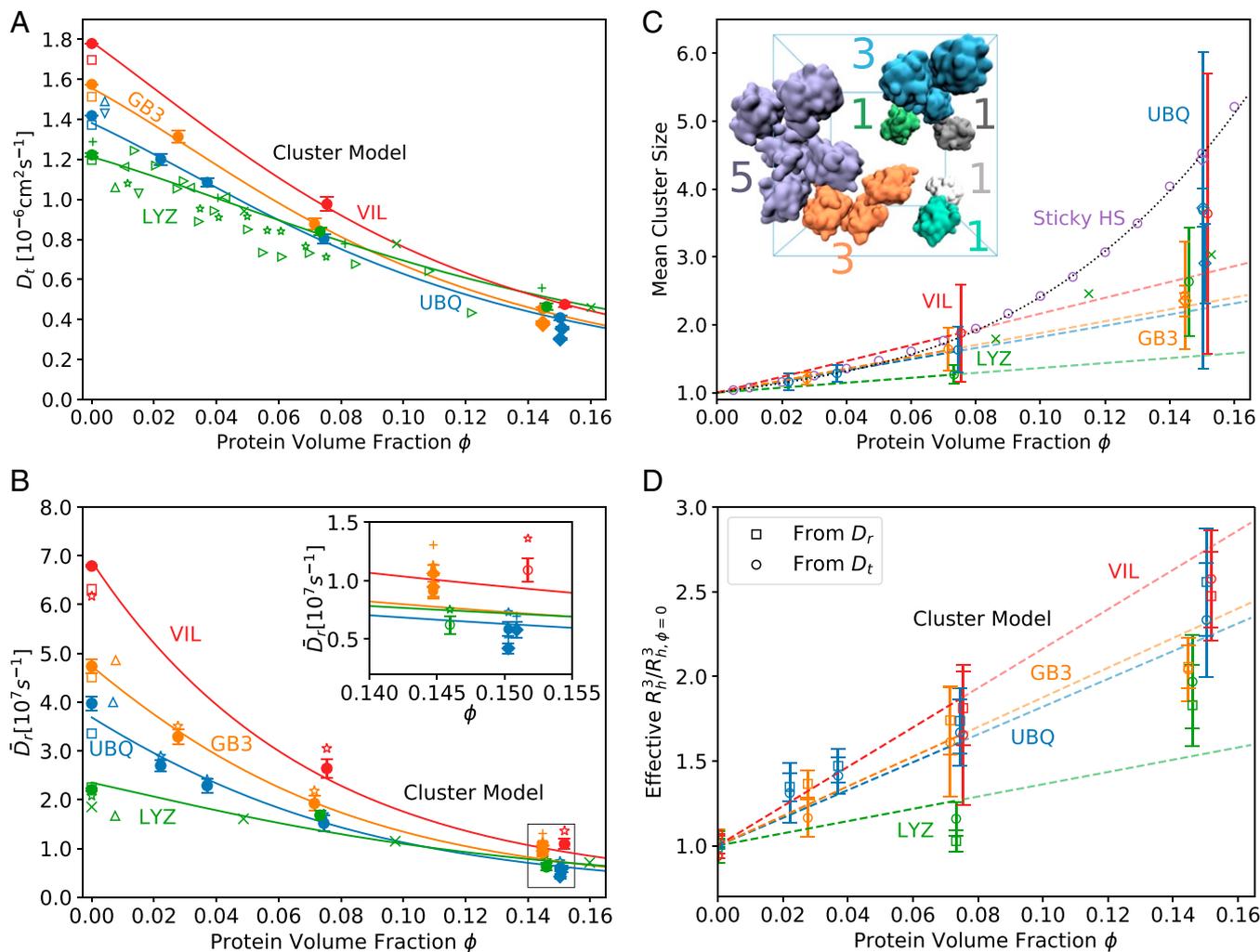
Fig. 3. Protein diffusion and clustering. (A and B) Concentration-dependent translational (A) and rotational (B) protein diffusion. (A) Dependence of the translational diffusion coefficient on protein volume fraction ϕ. Solid circles and diamonds show the finite-size corrected diffusion coefficients D_t of the small systems ($N \leq 20$) and large systems ($N \geq 120$), respectively. Open squares show Hydropro calculations (70), and other open symbols show experimental data. UBQ data: $\triangle$ (71), $\triangledown$ (72). LYZ data: $\triangle$ (73), $\triangledown$ (74), $\triangleleft$ (75), $\triangleright$ (21), $\star$ (76), $+$ (24), $\times$ (25). Lines show the predictions from the dynamic cluster model Eq. **8** with fitted $D_{t,\phi=0}$. (B) Dependence of the rotational diffusion coefficients on protein volume fraction ϕ. Solid circles and diamonds show the diffusion coefficients obtained from the anisotropic diffusion tensor and corrected for finite-size effects ($\bar{D}_r$) for the small systems ($N \leq 20$) and large systems ($N \geq 120$), respectively. Stars ($\star$) and plus signs ($+$) show the finite-size corrected diffusion coefficients $\bar{D}_r$ from integration of $\langle\langle P_1(\cos\theta)\rangle\rangle$ for the small and large systems, respectively. Triangles ($\triangle$) represent NMR data for dilute UBQ (blue) (77), dilute GB3 (orange) (78), and dilute LYZ (green) (73) solution. Green crosses ($\times$) show NMR data for LYZ (25). Open squares show results from Hydropro (70) calculations. Lines show the predictions from the dynamic cluster model Eq. **9** with fitted $D_{r,\phi=0}$. B, Inset shows a zoom-in at high protein volume fraction. (C and D) Protein-cluster model. (C) Dependence of mean cluster size $\bar{m}(\phi)$ on protein volume fraction ϕ [circles for small systems ($N \leq 20$), diamonds for large systems ($N \geq 120$)]. The dashed line shows a linear fit $1 + \zeta\phi$ to the data at 0–100 mg/mL with binding propensity ζ listed in Table 1. Purple circles and diamonds are results from a Monte Carlo simulation of Baxter's sticky HSs with $N = 16$ and $N = 120$ particles in the box, respectively, using $\tau = 0.15$ and finite-range attractive interactions up to 1.05 σ. The dotted curve shows the fit $\bar{m}(\phi) = 1 + \phi/\tau + 743.3\phi^3$ to the MC data up to $\phi = 0.15$. Green crosses show experimental data on LYZ cluster formation (79). (C, Inset) Representative simulation snapshot of dense UBQ solution (100 mg/mL). Blue lines indicate the periodically replicated simulation boxes. Colors and colored labels indicate transient UBQ protein clusters and cluster size m, respectively (solvent not shown). (D) Dependence of reduced hydrodynamic radii cubed, $R_h^3/R_{h,\phi=0}^3$, on protein volume fraction ϕ. The effective R_h is calculated from the Stokes–Einstein relations for translation, Eq. **10** (circles), and for rotation, Eq. **11** (squares). The dashed lines show the prediction $1 + \zeta\phi$ from C with ζ values from Table 1.

71–76. The C-terminal tail (residues 71–76) and an adjacent relatively hydrophobic surface patch show strong involvement, in line with experimental evidence (82). In dense LYZ solutions, we observed that residues Asp48 and Arg73 contribute most to LYZ–LYZ interaction (*SI Appendix*, Fig. S17). In a Brownian dynamics study (83), these residues were found to play crucial roles in the formation of a LYZ–LYZ encounter complex.

Colloidal Suspension Model. Baxter's attractive (sticky) HSs (68, 69, 84–86) are widely used as a model for suspensions of interacting colloidal particles. Their association constant is related

to the dimensionless Baxter parameter τ as $\tau = v_{\mathrm{HS}}K_d$ (derived in *SI Appendix*), where v_{HS} is the HS volume. Low τ indicates high protein stickiness. We conducted Monte Carlo (MC) simulations of Baxter sticky HSs with $\tau = 0.15$ and finite-range attractive interactions up to 1.05 σ of $N = 16$ and $N = 120$ particles in a simulation box. The mean cluster size grows linearly at low protein concentrations, deviating for $\phi > 0.075$ (Fig. 3C). Considering the rough approximations of this model, the cluster sizes of MC simulations of sticky HSs are in surprisingly good agreement with those of the atomistic MD simulations. We can also relate τ and K_d to the second virial coefficient
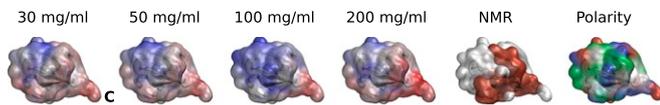
Fig. 4. UBQ protein–protein interaction interfaces. Red, white, and blue indicate strong, intermediate, and weak involvement in UBQ–UBQ contacts. The NMR experimental noncovalent dimer interface is derived from Liu et al. (82). The rightmost view indicates hydrophobic (white), polar (green), acidic (red), and basic (blue) surface residues. C indicates the C terminus.

(69) $B_2 = 2\pi \int_0^\infty r^2\, dr\, [1 - g(r)] = 4v_{HS} - v_{HS}/\tau = 4v_{HS} - 1/K_d$. Integration of the radial distribution function of the UBQ centers at 30 mg/mL and GB3 centers at 39 mg/mL over the first peak gives us values of $B_2(UBQ) = -38.9$ nm^3 and $B_2(GB3) = -8.9$ nm^3, respectively (*SI Appendix*, Fig. S18) (no data for LYZ and VIL, as we did not conduct simulations at low concentrations for these proteins). With $v_p(UBQ) = 10.4$ nm^3 and $v_p(GB3) = 7.2$ nm^3 and under the assumption of the sticky HS model, we thus obtain $K_d^{RDF}(UBQ) = 21$ mM and $K_d^{RDF}(GB3) = 44$ mM, in good agreement with the values from the cluster model (Table 1).

For the sticky HS model, Cichocki and Felderhof (68) derived a low-density expansion of the long-time single-particle diffusion coefficient in terms of the volume fraction,

$$\frac{D_t(\phi)}{D_{t,\phi=0}} \approx 1 - \left(2.0972 + \frac{0.562}{\tau}\right)\phi. \qquad [1]$$

We determined the slope $D'_{t,\phi=0}/D_{t,\phi=0} \approx 6.3$ (Fig. 5A, dashed-dotted line) from a fit to our protein simulation data for UBQ and GB3 at low ϕ, which gives us a value of $\tau^{D_t} \approx 0.1$. From τ^{D_t}, $v_p(UBQ) = 10.4$ nm^3, and $v_p(GB3) = 7.2$ nm^3, we obtain apparent dissociation constants of $K_d^{D_t}(UBQ) = \tau/(N_A v_p) = 21$ mM and $K_d^{D_t}(GB3) \approx 30$ mM that are in good agreement with the K_d values obtained directly from the cluster model and by integration of $g(r)$ (Table 1). Since the relative concentration-dependent slowdown of VIL diffusivity is similar to that of UBQ and GB3 (Fig. 5A), we also calculated $K_d^{D_t}(VIL) \approx 43$ mM using the same τ^{D_t}. This does not apply for LYZ solutions.

Cichocki and Felderhof (68) also evaluated the quadratic term in the viscosity expansion,

$$\frac{\eta(\phi)}{\eta_0} \approx 1 + 2.5\phi + \left(5.931 + \frac{1.899}{\tau}\right)\phi^2. \qquad [2]$$

Equating $b = 5.931 + 1.899/\tau$ with the coefficient b from the ϕ^2 term in the viscosity expansion fitted to MD data (Eq. **3** and Table 1), we obtain values of τ^η and K_d^η (Table 1), which are again close to those obtained from the cluster model, $g(r)$, and $D_t(\phi)$ (Table 1).

Dissociation Constant from Off Rate of Nonspecific Complexes. The cumulative distribution functions of the lifetimes of protein pairs (*SI Appendix*, Fig. S19) show that most pairs stay together for 1–50 ns, indicating dynamic clustering according to Liu et al.'s (24) terminology. The lifetimes of protein pairs are independent of the protein concentration, supporting the presence of dynamic protein clusters rather than protein aggregation. From the cumulative distribution function, we obtained the same median protein pair lifetime of $\tilde{t}_{off} \approx 5$ ns for all proteins and concentrations and defined an off rate for nonspecific complexes as $k_{off} = 1/t_{off} = \ln(2)/\tilde{t}_{off} \approx 1.4 \times 10^8$ s^{-1}, assuming exponential kinetics. Assuming in addition a Smoluchowski on rate, $k_{on} = 4\pi D_t R_h$, and substituting the Stokes–Einstein relation for translational diffusion, we expect an on rate for nonspecific complexes of $k_{on} = 2k_B T N_A/3\eta_0 \approx 1.8 \times 10^9$ M$^{-1}\cdot$s^{-1}. The resulting "kinetic" dissociation constant $K_d^{k_{off}} = k_{off}/k_{on} \approx 80$ mM agrees well with those obtained from the other methods (Table 1).
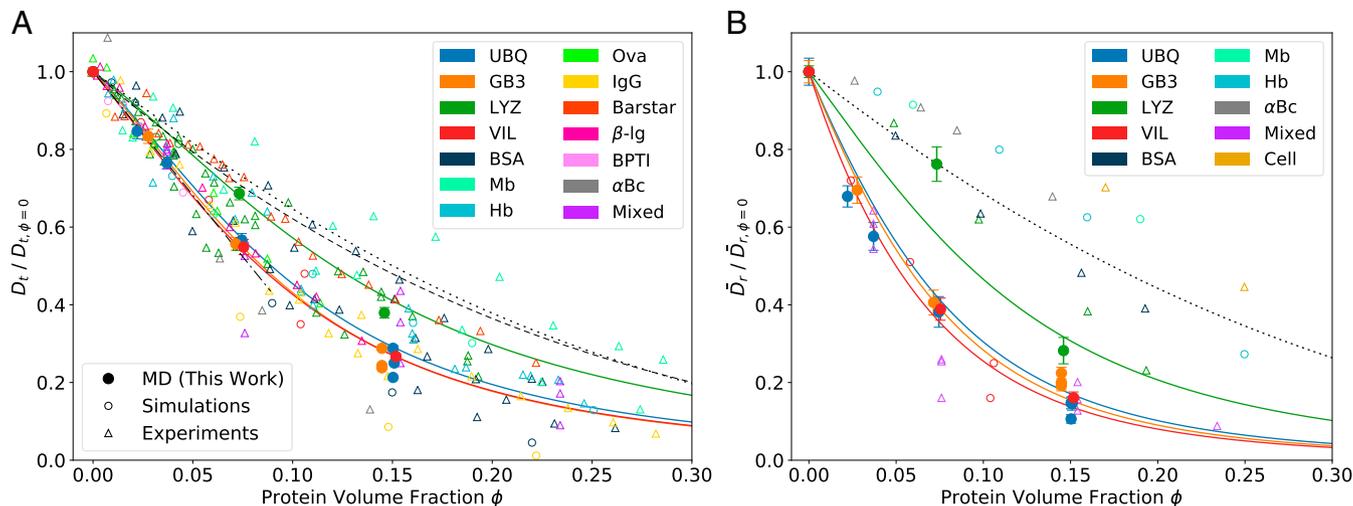
Fig. 5. Concentration-dependent protein diffusion. (*A*) Dependence of the normalized translational diffusion coefficient D_t^{red} on protein volume fraction ϕ. Solid circles show MD data from this study. Open circles show data from simulation studies (16, 26, 29, 87). Open triangles denote data from experimental studies (21–27, 75, 76, 88–93). Hb, hemoglobin; Mb, myoglobin; "Mixed," CI2 in different dense protein solutions; Ova, ovalbumin. The solid curves show the prediction of the dynamic cluster model $D_{t,clust}(\phi)$ (Eq. **8**, no adjustable parameters) for UBQ, GB3, LYZ, and VIL solutions, indicated by corresponding colors. The dashed and dotted curves show the slowdown of D_t^{red} predicted from colloid theory on noninteracting HSs by van Blaaderen et al. (94) and Tokuyama and Oppenheim (95), respectively (*SI Appendix*). The dashed-dotted line shows a linear fit to D_t^{red} of UBQ and GB3 at $\phi \leq 0.04$. A plot with experimental studies resolved by symbols is shown in *SI Appendix*, Fig. S20A. (*B*) Dependence of the normalized rotational diffusion coefficient D_r^{red} on protein volume fraction ϕ. Solid circles show MD data from this study. Open circles show data from simulation studies (16, 29). Open triangles denote data from experimental studies (22, 25, 96). "Cell": Hb or Mb in different cell types. The solid curves show the prediction of the dynamic cluster model $D_{r,clust}(\phi)$ (Eq. **9**, no adjustable parameters) for UBQ, GB3, LYZ, and VIL solutions, indicated by corresponding colors. The dashed curve shows the slowdown of D_r^{red} predicted from colloid theory for noninteracting HSs (97) (*SI Appendix*). A plot with experimental studies resolved by symbols is shown in *SI Appendix*, Fig. S20B.

Discussion

Relative Slowdown in Diffusion Is Consistent with Experiment and Cluster Model. The relative slowdown of translational and rotational diffusion of UBQ, GB3, LYZ, and VIL at increasing protein volume fraction is within the range of published experimental and simulation results. Literature data on the protein-concentration dependence of D_t were normalized to $D_t^{\text{red}} = D_t(\phi)/D_{t,\phi=0}$. If only protein mass concentrations were given, protein volume fractions were approximated according to the dependence of UBQ volume fraction on UBQ mass concentration.

The dependence of D_t^{red} on protein crowding is surprisingly consistent across the different studies (Fig. 5*A* and *SI Appendix*, Fig. S20*A*), considering that they cover both experiments and simulations and report results for proteins of different size, shape, and charge, for different model resolution and different experimental conditions (temperature, pH value). In particular, our simulation results for D_t^{red} of dense VIL solutions are in line with a recent study by Nawrocki et al. (29). Likewise, the slowdown of D_t^{red} of dense LYZ solutions agrees with the majority of experimental data on LYZ diffusion (green triangles in Fig. 5*A* and open green symbols in *SI Appendix*, Fig. S20*A*). We conclude that the slowdown of translational diffusion in concentrated protein solutions is a general feature of soluble proteins and that the extent of the slowdown is governed by the clustering propensity of the protein (notwithstanding additional shape effects, which may not be captured by our choice of globular proteins). We also conclude that the colloidal models of noninteracting HSs (94, 95) (*SI Appendix* and dashed and dotted lines in Fig. 5) significantly underestimate the slowdown of translational diffusion of most proteins studied.

We collected data on the slowdown of the rotational diffusion from experimental and simulation studies and normalized the diffusion coefficients to $D_r^{\text{red}} = \bar{D}_r(\phi)/\bar{D}_{r,\phi=0}$. The slowdown of D_r^{red} is less consistent in the considered studies (Fig. 5*B* and *SI Appendix*, Fig. S20*B*) than the slowdown of D_t^{red}. We suggest that the greater variability in the slowdown of D_r^{red} is due to the R_h^3 dependence of D_r^{red} compared with the R_h dependence of D_t^{red} in the Stokes–Einstein relations. Up to intermediate concentration (100 mg/mL), the slowdown of D_r^{red} in LYZ solutions, although in line with experiment (25), does not abide by the cluster model prediction, even though the model takes into account the reduced clustering propensity of the LYZ protein, possibly from its high positive charge (+9) (Fig. 3*C* and ζ in Table 1). Indeed, the slowdown of D_r^{red} at low LYZ concentrations appears to be better represented by a noninteracting HS model. In our simulations, LYZ forms very few clusters at intermediate (100 mg/mL) concentration. At higher concentration (200 mg/mL), we find a significant contribution of higher cluster numbers (>5) (*SI Appendix*, Fig. S10). From small-angle X-ray scattering data, Stradner et al. (79) inferred slightly larger LYZ clusters (green crosses in Fig. 3*C*) compared with our simulations. However, the higher pH in their experiments may have reduced the net protein charge and increased cluster formation. Scattering data were subsequently found to be compatible with lysozyme being largely repulsive (98).

We stress that our dynamic cluster model (Eqs. **8** and **9** and solid lines in Figs. 3 *A* and *B* and 5) predicts the relative slowdown in diffusion based on protein cluster size and viscosity. In light of this, the prediction of the model represents the observed and calculated slowdown in diffusion strikingly well. Our cluster model relates differences in the relative diffusivities of different proteins to differences in their nonspecific interactions. This observation provides a physical basis for the protein-dependent concentration scaling factor ξ introduced by ref. 21 to establish consistency with a HS diffusion model (95).

Separating the Effects of Clustering and Hydrodynamics. Using an elegant MD simulation setup, Nawrocki et al. (30) found that direct protein interactions are the dominant contributors to the slowdown of rotational diffusivity, whereas hydrodynamics play only a minor role. Here, we could show that the effective viscosity accounts for the indirect, hydrodynamic effects of dense solutions on the diffusivity slowdown (*SI Appendix*, Fig. S15). In dense UBQ, GB3, and VIL solutions at 200 mg/mL, the viscosity is ∼2.5-fold increased compared to the solvent viscosity (Fig. 2). The translational diffusion decreases by a factor of ∼4, whereas the rotational diffusion decreases by a factor of ∼6 (Fig. 3 *A* and *B*). The direct effect of protein clustering (i.e., the increase in the effective hydrodynamic radius) accounts for an additional factor ∼4/2.5 = 1.6 on translational diffusion and a factor ∼6/2.5 = 2.4 on rotational diffusion. We therefore find that direct clustering interactions contribute to ∼40% and 50% of the total slowdown in translational and rotational diffusivity, respectively, of dense UBQ, GB3, and VIL solutions.

The displacement pair correlation function assesses concerted protein motion and is used to distinguish short-range and long-range interactions in simulations of crowded systems (19). At short distances, the proteins form clusters, and their motion is highly correlated (*SI Appendix*, Fig. S16). The pair correlation decreases at larger protein pair distances, because the short-range interactions fade out and only contributions from long-range effects (electrostatics, hydrodynamics) remain. Interestingly, at low protein concentrations (30–50 mg/mL), the protein motions are correlated up to high distances, whereas the correlation decreases quickly with increasing distance at high protein concentration. This suggests that the motion of proteins at low concentration is more effectively coupled by hydrodynamic and electrostatic forces. At high concentrations, the coupled motion of a protein pair at short distances appears to be effectively quenched by interference of competing proteins. This trend is surprising, because proteins may serve as bridging intermediates at high concentrations, which should effectively increase the pair correlation for the distances considered.

Protein Solutions as Colloidal Suspensions with Attractions. The effective dissociation constants K_d and Baxter parameters τ were derived according to the low-concentration behavior of the sticky-HS model for attractive colloidal particles. The K_d values for nonspecific protein–protein binding derived from the structure, dynamics, and viscosity of the protein solutions are fairly consistent with each other (Table 1). As listed in Table 1, K_d estimates from $\overline{m}(\phi)$, $g(r)$, $D_t(\phi)$, k_{off}, and $\eta(\phi)$ vary within about one order of magnitude for a given protein. The corresponding values of τ also agree well with each other. Interestingly, the decreased clustering propensity of LYZ solutions compared with UBQ, GB3, and VIL solutions is reflected only by a higher τ, but not by an increase in the effective dissociation constant $K_d = \tau/v_p$, because the increased τ value of LYZ is compensated by its larger volume v_p. In turn, this results in a relatively small $\zeta = 1/(K_d v_p)$ for LYZ. For our limited set of four proteins, protein stickiness decreases with size, thereby ameliorating the aggregation problem for large proteins pointed out by Ando and Skolnick (19).

The formation of 3D clusters explains the strong increase of the mean cluster size $\overline{m}(\phi)$ in sticky-HS solutions at large ϕ (Fig. 3*C*). Sticky spheres have no orientational preference, which allows the formation of compact clusters. By contrast, protein–protein interactions are directional and only some protein orientations result in favorable interactions (*SI Appendix*, Fig. S17). The orientational preference disfavors the formation of compact clusters, as can be seen in the formation of only $m-1$ protein connections for a cluster of size m up to intermediate (100 mg/mL) protein concentration (*SI Appendix*, Fig. S21).

In consideration of these differences, τ should be considered an effective parameter whose value depends on the property that is fitted and on the fit range. Therefore, we emphasize that the sticky HS model is only a rough approximation of a complex protein solution, which nonetheless explains the behavior of the complex protein systems surprisingly well up to intermediate protein concentration.

From light-scattering experiments, Scherer et al. (57) estimated $\tau \approx 0.6$ for concentrated monoclonal antibody mAB2 solution (150 mM NaCl concentration), which is of the same order of magnitude as our simulation results (Table 1). They observed a strong dependence of τ on the ionic strength, ranging from $\sim$1,000 to $\sim$0.5 when increasing the ion concentration from 40 mM to 600 mM. It will be interesting to study the effect of varying ionic strengths on τ by MD simulations. In contrast to mAB2 solutions, the scattering data for antibody mAB1 solutions could not be represented by Baxter's sticky HS model due to stronger (and possibly long-range) interactions, whereas an alternative model based on two oligomeric species (57, 58) fitted the data well. This suggests that only weakly interacting proteins are well described by the attractive HS model (57).

Limitations. The generality of our findings on the connection of cluster size, viscosity, and diffusion is subject to several limitations.

All four proteins studied here are small and globular. This justifies the use of averaged rotational diffusion coefficients $\bar{D}_r$ and facilitates the comparison with colloid theory, but does not consider the effects of different shapes on the concentration dependence of diffusion. Dense solutions of larger proteins and proteins with varying degree of anisotropy will have to be simulated atomistically to address these issues.

Slight finite-size effects on protein clustering and diffusivity were observed. In the smallest simulation of dense protein solutions (UBQ at 200 mg/mL with $N = 15$ proteins), occasional formation of a box-spanning cluster was observed, as seen in the contribution of cluster sizes $m = 12-15$ to the cluster size distribution for this system (*SI Appendix*, Fig. S10). This artifact was not seen in any of the other simulations and does not appear to strongly affect the diffusivity and viscosity of the UBQ solution (Figs. 2 and 3). The translational and rotational diffusion coefficients at high concentration (200 mg/mL) vary slightly for higher protein numbers in the simulation box. The absolute deviations in D_t and D_r between small and large systems are small. However, the relative deviations in D_t and D_r are substantial, overestimating the viscosity calculated from the Stokes–Einstein relations (Fig. 2, circles). On the one hand, we attribute these deviations to incomplete sampling of the extremely large simulation systems. We expect that the simulations with large system size ($N \geq 120$) take more time to equilibrate, thus introducing a bias on the diffusion coefficients, which is not accounted for by the current finite-size corrections (Eq. **7** and *SI Appendix*, Eq. **S10**). On the other hand, we observe a significant tail to high cluster numbers in the cluster distribution in these systems (*SI Appendix*, Fig. S10). Whereas the mean cluster sizes in the small and large systems are similar (Fig. 3*C*), occasional large clusters in the large systems appear to slightly suppress diffusion in a way not captured by the smaller simulation systems. Assessing and quantifying both effects in detail would merit a separate study.

Based on the clustering data for UBQ and GB3 solutions, we assume that the cluster size increases linearly with protein concentration up to intermediate (100 mg/mL) protein concentration. Although this simple model works surprisingly well in explaining our calculated diffusion data even at higher concentrations (Fig. 3 *A* and *B*), the actual functional dependence of cluster size on protein concentration may be nonlinear and depend on the specifics of the system (protein type, pH, ionic strength, temperature). Much longer simulations would be needed to precisely determine the cluster distribution and mean cluster size at high protein concentration.

The representation of dense protein solutions by Baxter's sticky HS suspensions is limited to weak, short-range protein–protein interactions. As scattering data show (57), the sticky HS model does not represent experimental data well for strongly interacting particles or particles with significant attractive long-range (electrostatic) interactions. In our simulations and in cellular conditions, these interactions are effectively shielded by the ions in the solution, making this limitation less relevant in vivo. Nevertheless, it would be interesting to test the applicability of the cluster model for these cases.

Conclusions

By performing all-atom molecular dynamics simulations of dense protein solutions, we found an increase in the viscosity of the solutions at higher protein volume fractions, consistent with experimental results (39, 61, 62). This increase is considerably higher than predicted by colloidal models of noninteracting HSs, stressing the importance of measuring or calculating rather than approximating the viscosity at protein volume fractions approaching cellular crowding conditions. We calculated translational and rotational diffusion coefficients and corrected them for finite-size effects using the respective viscosity of the solution. Translational diffusion and rotational diffusion are strongly affected by protein crowding. For LYZ solutions, experimentally measured diffusion coefficients are available also at high concentration and are in excellent agreement with our simulation data (Fig. 3 *A* and *B*). We calculated effective hydrodynamic radii using the Stokes–Einstein relations and found that a similar increase in the effective hydrodynamic radius can be inferred from the slowdown of translational and rotational diffusion caused by the formation of dynamic protein clusters. Indeed, establishing consistency with the Stokes–Einstein relations requires accounting for protein cluster formation (42) as a result of attractive interactions (60). Overall, we conclude that the concentration dependence of protein cluster size, the translational and rotational diffusion coefficient, and viscosity are consistent with each other (exception: LYZ at 100 mg/mL) and—for the proteins studied here—are explained well by Baxter's sticky HS model of colloidal suspensions.

Representing the diffusion data in reduced form as a function of protein volume fraction showed that the relative slowdown in translational diffusion is consistent with results from previous studies. The relative slowdown in rotational diffusion shows a larger spread, consistent with the notion that rotational diffusivity depends more sensibly on clustering propensity and thus on the specifics of the protein interactions. Dynamic cluster formation has recently been observed also for membrane proteins (59) and shown to slow down rotational diffusion. In light of our analysis in terms of the cluster and colloidal models, we would expect similar affinities $K_d \approx 10$ mM for other abundant, soluble proteins with similar size and shape. It will be interesting to explore the limits of our cluster model, i.e., if the cluster model remains valid for large, anisotropic, and more sticky proteins. In this context, the effect of liquid–liquid phase separation (99) on protein diffusivity is an exciting question to address.

We find that the proteins favor certain orientations for interactions and our findings on UBQ contact interfaces are consistent with experiments (82). The protein interactions lead to highly correlated motion at short distances and the correlation is sustained up to larger distances at low concentration. At high concentration, despite increased protein cluster formation, the pair correlation (19) at similar distances is decreased.

In the cellular environment, the situation is complicated by molecular heterogeneity, reactions, partitioning in microenvironments by phase separation, interactions with membranes

and structural proteins, and other factors (99–102). Nevertheless, the findings here and in earlier work (103), as well as the observation that diffusion in cell lysates is similar to diffusion in crowded protein solutions (22), suggest that both in concentrated solution and in cells, proteins appear to diffuse not as isolated particles, but as members of dynamic clusters between which they constantly exchange. From the consistency of our diffusivity results with experiments in solution and on the basis of our cluster model, we conclude that—in crowded conditions corresponding to the cellular concentration—the strength of nonspecific protein–protein interactions for abundant proteins such as UBQ should correspond to low-millimolar binding.

We can now carry out atomistic MD simulations of crowded simulations at an unprecedented scale (29), here with up to 540 proteins and 3.6 million atoms in the box simulated over microseconds. Atomistic simulations of solutions of protein mixtures, possibly reflecting the distribution of proteins in the cell, no longer seem out of reach (52). Developments in nucleic acid force fields (104–109) will make it attractive to test the above findings on dense nucleic acid solutions and dense protein–nucleic acid mixtures. Ultimately, the macromolecular diversity in the cell will have to be considered (101) to predict passive diffusion in vivo.

SI Appendix. *SI Appendix* contains supplementary text; *SI Appendix*, Figs. S1–S21; *SI Appendix*, Tables S1–S3; Movie S1; and SI references.

Movie S1. Shown is an atomistic MD simulation of 540 GB3 proteins in concentrated solution (200 mg/mL) at simulation time 0–500 ns. The fully flexible proteins are shown in surface representation and differentiated by color. For clarity, water and ions are omitted. Proteins that seem to appear and disappear traverse the periodic boundaries.

Materials and Methods

MD Simulations of Dense Protein Solutions. We performed all-atom MD simulations of solutions of human UBQ [PDB code 1UBQ (110)], GB3 [PDB code 1P7F (111)], LYZ [PDB code 1E8L (112)], and VIL [PDB code 1VII (113)] at up to five different densities with $N = 15$ (UBQ) and $N = 20$ (GB3, LYZ, VIL) proteins in the simulation box. In addition, the most concentrated UBQ and GB3 solutions (200 mg/mL) were simulated with large simulation boxes containing from $N = 120$ to $N = 540$ proteins (*SI Appendix*, Table S1). To mimic an infinitely dilute system, MD simulations with a single protein copy were carried out. The simulation procedures are detailed in *SI Appendix*.

Viscosity Calculation and Approximations. The low-frequency, low-shear viscosity $\eta(\phi)$ of dense protein solutions differs from the viscosity η_0 of the pure solvent consisting of only water and ions. We determined $\eta(\phi)$ and η_0 from MD simulations by integration of the autocorrelation functions of the pressure tensor fluctuations (114), as detailed in *SI Appendix*. The dependence of η on the protein volume fraction ϕ is well captured by a quadratic function,

$$\eta(\phi) = (1 + 2.5\phi + b\phi^2)\eta_0,$$ [3]

with parameter b fitted to the calculated viscosities, η_0 the solvent viscosity averaged over the NaCl concentrations used in this study, and the coefficient 2.5 adopted from Einstein's colloid theory (66). The experiments of Woutersen and De Kruif (69) have shown that the coefficient b increases with the strength of attractions between colloidal particles.

We compared the calculated viscosities to predictions from colloid theory. At low solute volume fractions ϕ, Einstein (66) predicted a linear dependence of the viscosity of HS suspensions on ϕ,

$$\eta(\phi) = (1 + 2.5\phi)\eta_0.$$ [4]

For higher solute volume fractions, this expression was modified by Ladd (63, 64) to

$$\eta(\phi) = \frac{1 + 1.5\phi(1 + S(\phi))}{1 - \phi(1 + S(\phi))}\eta_0,$$ [5]

with $S(\phi) = \phi + \phi^2 - 2.3\phi^3$.

In addition, we also estimated the viscosity by assuming that the Stokes–Einstein relations for rotational diffusion, $D_r = k_B T/(8\pi\eta R_h^3)$, and for translational diffusion, $D_t = k_B T/(6\pi\eta R_h)$, are exactly satisfied, such that

$$\eta(\phi) = \frac{k_B T}{\pi}\sqrt{\frac{\bar{D}_r(\phi)}{27 D_t^3(\phi)}},$$ [6]

where the bar indicates that we average over orientational asymmetries.

Translational Diffusion. Translational diffusion coefficients D_t^{PBC} were obtained for each protein density by fitting the Einstein relation to the MSD, as detailed in *SI Appendix*. The diffusion coefficients were corrected for finite-size effects using (115)

$$D_t = D_t^{PBC} + \frac{k_B T\xi}{6\pi\eta(\phi)L},$$ [7]

with $\xi = 2.837297$ and L the edge length of the cubic simulation box. In the simulations of dilute proteins, where the protein radius R_p is comparable to the box dimension L, we used the expanded correction, $\xi = 2.837297 - 4\pi R_p^2/3L^2$ (115).

We compared the calculated translational diffusion coefficients to the predictions of a dynamic cluster model without any free parameters,

$$D_{t,\text{clust}}(\phi) = D_{t,\phi=0}\frac{\eta_0}{\eta(\phi)(1 + \zeta\phi)^{1/3}},$$ [8]

where $1 + \zeta\phi$ is the mean cluster size, as defined in *SI Appendix* and calculated directly from the MD structures. This model is based on the assumption that the slowdown in translational diffusion is linked to the increase in the effective hydrodynamic radius and in the viscosity via the Stokes–Einstein relation for translational diffusion. We also compared the reduced translational diffusion coefficients $D_t(\phi)/D_{t,\phi=0}$ to the approximate form for monodisperse noninteracting HS colloidal suspensions (94, 95) (*SI Appendix*).

Rotational Diffusion. Rotational diffusion coefficients were calculated following the procedure by Linke et al. (80, 116), as detailed in *SI Appendix*. Additionally, an effective rotational diffusion coefficient was obtained from fits to the orientational correlation function $\langle\langle P_1(\cos\theta(t))\rangle\rangle = \langle\langle\cos\theta(t)\rangle\rangle$ (117, 118), as detailed in *SI Appendix*.

We compared the calculated rotational diffusion coefficients to the predictions of the dynamic cluster model,

$$D_{r,\text{clust}}(\phi) = D_{r,\phi=0}\frac{\eta_0}{\eta(\phi)(1 + \zeta\phi)},$$ [9]

where we again assumed the Stokes–Einstein relation to apply with a hydrodynamic radius cubed proportional to the mean cluster size $1 + \zeta\phi$. We also compared the reduced rotational diffusion coefficients $D_r(\phi)/D_{r,\phi=0}$ to predictions from colloidal models of noninteracting HSs (97) (*SI Appendix*).

Hydrodynamic Radius. We solved the Stokes–Einstein relations for translational and rotational diffusion to define effective hydrodynamic radii for translation,

$$R_{h,t}(\phi) = \frac{k_B T}{6\pi\eta(\phi)D_t(\phi)},$$ [10]

and for rotation,

$$R_{h,r}(\phi) = \left(\frac{k_B T}{8\pi\eta(\phi)\bar{D}_r(\phi)}\right)^{1/3}.$$ [11]

Gaussian error propagation of $D_t(\phi)$, $\bar{D}_r(\phi)$, and $\eta(\phi)$ was employed to estimate the errors of $R_{h,t}$ and $R_{h,r}$.

1. Zimmerman SB, Trach SO (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J Mol Biol* 222:599–620.
2. Ellis RJ (2001) Macromolecular crowding: Obvious but underappreciated. *Trends Biochem Sci* 26:597–604.
3. Zimmerman SB, Minton AP (1993) Macromolecular crowding: Biochemical, biophysical, and physiological consequences. *Annu Rev Biophys Biomol Struct* 22:27–65.
4. Minton AP (2001) The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J Biol Chem* 276:10577–10580.
5. Feig M, Sugita Y (2012) Variable interactions between protein crowders and biomolecular solutes are important in understanding cellular crowding. *J Phys Chem B* 116:599–605.
6. Mittal S, Chowhan RK, Singh LR (2015) Macromolecular crowding: Macromolecules friend or foe. *Biochim Biophys Acta Gen Subj* 1850:1822–1831.
7. Ellis RJ, Minton AP (2006) Protein aggregation in crowded environments. *Biol Chem* 387:485–497.
8. Zhou HX (2013) Influence of crowded cellular environments on protein folding, binding, and oligomerization: Biological consequences and potentials of atomistic modeling. *FEBS Lett* 587:1053–1061.
9. Park S, Agmon N (2008) Theory and simulation of diffusion-controlled Michaelis-Menten kinetics for a static enzyme in solution. *J Phys Chem B* 112:5977–5987.
10. Guseman AJ, Perez Goncalves GM, Speer SL, Young GB, Pielak GJ (2018) Protein shape modulates crowding effects. *Proc Natl Acad Sci USA* 115:10965–10970.
11. Blanco MA, Perevozchikova T, Martorana V, Manno M, Roberts CJ (2014) Protein-protein interactions in dilute to concentrated solutions: α-Chymotrypsinogen in acidic conditions. *J Phys Chem B* 118:5817–5831.
12. Kekenes-Huskey PM, Eun C, McCammon JA (2015) Enzyme localization, crowding, and buffers collectively modulate diffusion-influenced signal transduction: Insights from continuum diffusion modeling. *J Chem Phys* 143:94103.
13. Benton LA, Smith AE, Young GB, Pielak GJ (2012) Unexpected effects of macromolecular crowding on protein stability. *Biochemistry* 51:9773–9775.
14. Muramatsu N, Minton AP (1988) Tracer diffusion of globular proteins in concentrated protein solutions. *Proc Natl Acad Sci USA* 85:2984–2988.
15. Dix JA, Verkman A (2008) Crowding effects on diffusion in solutions and cells. *Annu Rev Biophys* 37:247–263.
16. Mereghetti P, Wade RC (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.
17. Ansari A, Jones CM, Henry ER, Hofrichter J, Eaton WA (1992) The role of solvent viscosity in the dynamics of protein conformational changes. *Science* 256:1796–1798.
18. Elcock AH (2010) Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Curr Opin Struct Biol* 20:196–206.
19. Ando T, Skolnick J (2010) Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. *Proc Natl Acad Sci USA* 107:18457–18462.
20. Trovato F, Tozzini V (2014) Diffusion within the cytoplasm: A mesoscale model of interacting macromolecules. *Biophys J* 107:2579–2591.
21. Nesmelova IV, Skirda VD, Fedotov VD (2002) Generalized concentration dependence of globular protein self-diffusion coefficients in aqueous solutions. *Biopolymers* 63:132–140.
22. Wang Y, Li C, Pielak GJ (2010) Effects of proteins on protein diffusion. *J Am Chem Soc* 132:9392–9397.
23. Porcar L, et al. (2010) Formation of the dynamic clusters in concentrated lysozyme protein solutions. *J Phys Chem Lett* 1:126–129.
24. Liu Y, et al. (2011) Lysozyme protein solution with an intermediate range order structure. *J Phys Chem B* 115:7238–7247.
25. Roos M, et al. (2016) Coupling and decoupling of rotational and translational diffusion of proteins under crowding conditions. *J Am Chem Soc* 138:10365–10372.
26. Balbo J, Mereghetti P, Herten DP, Wade RC (2013) The shape of protein crowders is a major determinant of protein diffusion. *Biophys J* 104:1576–1584.
27. Grimaldo M, Roosen-Runge F, Zhang F, Seydel T, Schreiber F (2014) Diffusion and dynamics of γ-globulin in crowded aqueous solutions. *J Phys Chem B* 118:7203–7209.
28. Feig M, Sugita Y (2013) Reaching new levels of realism in modeling biological macromolecules in cellular environments. *J Mol Graph Model* 45:144–156.
29. Nawrocki G, Wang PH, Yu I, Sugita Y, Feig M (2017) Slow-down in diffusion in crowded protein solutions correlates with transient cluster formation. *J Phys Chem B* 121:11072–11084.
30. Nawrocki G, Karaboga A, Sugita Y, Feig M (2019) Effect of protein–protein interactions and solvent viscosity on the rotational diffusion of proteins in crowded environments. *Phys Chem Chem Phys* 21:876–883.
31. Phillies GD, Ullmann GS, Ullmann K, Lin TH (1985) Phenomenological scaling laws for "semidilute" macromolecule solutions from light scattering by optical probe particles. *J Chem Phys* 82:5242–5246.
32. Gorti S, Ware BR (1985) Probe diffusion in an aqueous polyelectrolyte solution. *J Chem Phys* 83:6449–6456.
33. Konopka MC, et al. (2009) Cytoplasmic protein mobility in osmotically stressed Escherichia coli. *J Bacteriol* 91:231–237.
34. Verkman AS (2003) Diffusion in cells measured by fluorescence recovery after photobleaching. *Methods Enzymol* 360:635–648.
35. Mastro AM, Babich MA, Taylor WD, Keith AD (1984) Diffusion of a small molecule in the cytoplasm of mammalian cells. *Proc Natl Acad Sci USA* 81:3414–3418.
36. Jin S, Verkman AS (2007) Single particle tracking of complex diffusion in membranes: Simulation and detection of barrier, raft, and interaction phenomena. *J Phys Chem B* 111:3625–3632.
37. Dauty E, Verkman AS (2004) Molecular crowding reduces to a similar extent the diffusion of small solutes and macromolecules: Measurement by fluorescence correlation spectroscopy. *J Mol Recognit* 17:441–447.
38. Szymański J, Patkowski A, Wilk A, Garstecki P, Holyst R (2006) Diffusion and viscosity in a crowded environment: From nano- to macroscale. *J Phys Chem B* 110:25593–25597.
39. Zorrilla S, Hink MA, Visser AJ, Lillo MP (2007) Translational and rotational motions of proteins in a protein crowded environment. *Biophys Chem* 125:298–305.
40. Roosen-Runge F, et al. (2011) Protein self-diffusion in crowded solutions. *Proc Natl Acad Sci USA* 108:11815–11820.
41. Li C, Liu M (2013) Protein dynamics in living cells studied by in-cell NMR spectroscopy. *FEBS Lett* 587:1008–1011.
42. Rothe M, et al. (2016) Transient binding accounts for apparent violation of the generalized Stokes-Einstein relation in crowded protein solutions. *Phys Chem Chem Phys* 18:18006–18014.
43. Geyer T (2012) Mixing normal and anomalous diffusion. *J Chem Phys* 137:115101.
44. Kondrat S, Zimmermann O, Wiechert W, Lieres EV (2015) The effect of composition on diffusion of macromolecules in a crowded environment. *Phys Biol* 12:46003.
45. Kim YC, Best RB, Mittal J (2010) Macromolecular crowding effects on protein-protein binding affinity and specificity. *J Chem Phys* 133:205101.
46. Kim YC, Mittal J (2013) Crowding induced entropy-enthalpy compensation in protein association equilibria. *Phys Rev Lett* 110:208102.
47. Kang H, Pincus PA, Hyeon C, Thirumalai D (2015) Effects of macromolecular crowding on the collapse of biopolymers. *Phys Rev Lett* 114:68303.
48. McGuffee SR, Elcock AH (2006) Atomically detailed simulations of concentrated protein solutions: The effects of salt, pH, point mutations, and protein concentration in simulations of 1000-molecule systems. *J Am Chem Soc* 128:12098–12110.
49. Frembgen-Kesner T, Elcock AH (2009) Striking effects of hydrodynamic interactions on the simulated diffusion and folding of proteins. *J Chem Theory Comput* 5:242–256.
50. Cossins BP, Jacobson MP, Guallar V (2011) A new view of the bacterial cytosol environment. *PLoS Comput Biol* 7:e1002066.
51. Lu C, Prada-Gracia D, Rao F (2014) Structure and dynamics of water in crowded environments slows down peptide conformational changes. *J Chem Phys* 141:45101.
52. Yu I, et al. (2016) Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *Elife* 5:e19274.
53. Minton AP (1980) Excluded volume as a determinant of protein structure and stability. *Biophys J* 32:77–79.
54. Minton AP (1983) The effect of volume occupancy upon the thermodynamic activity of proteins: Some biochemical consequences. *Mol Cell Biochem* 55:119–140.
55. Muramatsu N, Minton AP (1989) Hidden self-association of proteins. *J Mol Recognit* 1:166–171.
56. Stradner A, et al. (2004) Equilibrium cluster formation in concentrated protein solutions and colloids. *Nature* 432:492–495.
57. Scherer TM, Liu J, Shire SJ, Minton AP (2010) Intermolecular interactions of IgG1 monoclonal antibodies at high concentrations characterized by light scattering. *J Phys Chem B* 114:12948–12957.
58. Lilyestrom WG, Yadav S, Shire SJ, Scherer TM (2013) Monoclonal antibody self-association, cluster formation, and rheology at high concentrations. *J Phys Chem B* 117:6373–6384.
59. Paulino J, et al. (2019) Influenza A M2 channel clustering at high protein/lipid ratios: Viral budding implications. *Biophys J* 116:1075–1084.
60. Turton DA, Wynne K (2014) Stokes-Einstein-Debye failure in molecular orientational diffusion: Exception or rule? *J Phys Chem B* 118:4600–4604.
61. Godfrin PD, et al. (2015) Short-time glassy dynamics in viscous protein solutions with competing interactions. *Phys Rev Lett* 115:228302.
62. Heinen M, et al. (2012) Viscosity and diffusion: Crowding and salt effects in protein solutions. *Soft Matter* 8:1404–1419.
63. Ladd AJC (1990) Hydrodynamic transport coefficients of random dispersions of hard spheres. *J Chem Phys* 93:3484–3494.
64. Bedeaux D (1987) The effective viscosity for a suspension of spheres. *J Colloid Interf Sci* 118:80–86.
65. Piana S, Donchev AG, Robustelli P, Shaw DE (2015) Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J Phys Chem B* 119:5113–5123.
66. Einstein A (1906) Eine neue bestimmung der moleküldimensionen [A new determination of molecule dimensions]. *Ann Phys* 19:289–306.
67. Russel WB (1984) The Huggins coefficient as a means for characterizing suspended particles. *J Chem Soc Faraday Trans 2 Mol Chem Phys* 80:31–41.
68. Cichocki B, Felderhof BU (1990) Diffusion coefficients and effective viscosity of suspensions of sticky hard spheres with hydrodynamic interactions. *J Chem Phys* 93:4427–4432.
69. Woutersen ATJM, De Kruif CG (1991) The rheology of adhesive hard sphere dispersions. *J Chem Phys* 94:5739–5750.
70. Ortega A, et al. (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys J* 101:892–898.
71. Altieri AS, Byrd RA, Hinton DP (1995) Association of biomolecular systems via pulsed field gradient NMR self-diffusion measurements. *J Am Chem Soc* 117:7566–7567.
72. Mayo KH, Ilyina E, Park H (1996) A recipe for designing water-soluble, beta-sheet-forming peptides. *Protein Sci* 5:1301–1315.
73. Dubin SB, Clark NA, Benedek GB (1971) Measurement of the rotational diffusion coefficient of lysozyme by depolarized light scattering: Configuration of lysozyme in solution. *J Chem Phys* 54:5158–5164.
74. Mikol V, Hirsch E, Giegé R (1990) Diagnostic of precipitant for biomacromolecule crystallization by quasi-elastic light-scattering. *J Mol Biol* 213:187–195.

75. Price WS, Tsuchiya F, Arata Y (1999) Lysozyme aggregation and solution properties studied using PGSE NMR diffusion measurements. *J Am Chem Soc* 121:11503–11512.
76. Coffman J, Lightfoot E, Root T (1997) Protein diffusion in porous chromatographic media studied by proton and fluorine PFG-NMR. *J Phys Chem B* 101:2218–2223.
77. Tjandra N, Bax A, Feller SE, Pastor RW (1995) Rotational diffusion anisotropy of human ubiquitin from 15N NMR relaxation. *J Am Chem Soc* 117:12562–12566.
78. Hall JB, Fushman D (2003) Characterization of the overall and local dynamics of a protein with intermediate rotational anisotropy: Differentiating between conformational exchange and anisotropic diffusion in the B3 domain of protein G. *J Biomol NMR* 27:261–275.
79. Stradner A, Cardinaux F, Schurtenberger P (2006) A small-angle scattering study on equilibrium clusters in lysozyme solutions. *J Phys Chem B* 110:21222–21231.
80. Linke M, Köfinger J, Hummer G (2018) Fully anisotropic rotational diffusion tensor from molecular dynamics simulations. *J Phys Chem B* 122:5630–5639.
81. Linke M, Köfinger J, Hummer G (2018) Rotational diffusion depends on box size in molecular dynamics simulations. *J Phys Chem Lett* 9:2874–2878.
82. Liu Z, et al. (2012) Noncovalent dimerization of ubiquitin. *Angew Chem Int Ed* 51:469–472.
83. Ermakova E (2005) Lysozyme dimerization: Brownian dynamics simulation. *J Mol Model* 12:34–41.
84. Baxter RJ (1968) Percus-Yevick equation for hard spheres with surface adhesion. *J Chem Phys* 49:2770–2774.
85. Kranendonk WGT, Frenkel D (1988) Simulation of the adhesive-hard-sphere model. *Mol Phys* 64:403–424.
86. Verduin H, Dhont JK (1995) Phase diagram of a model adhesive hard-sphere dispersion. *J Colloid Interf Sci* 172:425–437.
87. Mereghetti P, Gabdoulline RR, Wade RC (2010) Brownian dynamics simulation of protein solutions: Structural and dynamical properties. *Biophys J* 99:3782–3791.
88. Keller KH, Canales ER, Yam S (1971) Tracer and mutual diffusion coefficients of proteins. *J Phys Chem* 75:379–387.
89. Gibbs SJ, Chu AS, Lightfoot EN, Root TW (1991) Ovalbumin diffusion at low ionic strength. *J Phys Chem* 95:467–471.
90. Le Bon C, Nicolai T, Kuil ME, Hollander JG (1999) Self-diffusion and cooperative diffusion of globular proteins in solution. *J Phys Chem B* 103:10294–10299.
91. Longeville S, Doster W, Diehl M, Gähler R, Petry W (2003) Neutron resonance spin echo: Oxygen transport in crowded protein solutions. *Neutron Spin Echo Spectroscopy*: Basics, Trends and Applications, eds Mezei F, Pappas C, Gutberlet T (Springer, Berlin), pp 325–335.
92. Doster W, Longeville S (2007) Microscopic diffusion and hydrodynamic interactions of hemoglobin in red blood cells. *Biophys J* 93:1360–1368.
93. Longeville S, Stingaciu LR (2017) Hemoglobin diffusion and the dynamics of oxygen capture by red blood cells. *Sci Rep* 7:1–10.
94. van Blaaderen A, Peetermans J, Maret G, Dhont JKG (1992) Long-time self-diffusion of spherical colloidal particles measured with fluorescence recovery after photobleaching. *J Chem Phys* 96:4591–4603.
95. Tokuyama M, Oppenheim I (1994) Dynamics of hard-sphere suspensions. *Phys Rev E–Stat Nonlinear Soft Matter Phys* 50:R16–R19.
96. Wang D, Kreutzer U, Chung Y, Jue T (1997) Myoglobin and hemoglobin rotational diffusion in the cell. *Biophys J* 73:2764–2770.
97. Zuzovsky M, Adler PM, Brenner H (1983) Spatially periodic suspensions of convex particles in linear shear flows. III. Dilute arrays of spheres suspended in Newtonian fluids. *Phys Fluids* 26:1714.
98. Shukla A, et al. (2008) Absence of equilibrium cluster phase in concentrated lysozyme solutions. *Proc Natl Acad Sci USA* 105:5075–5080.
99. Hyman AA, Weber CA, Jülicher F (2014) Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol* 30:39–58.
100. Luby-Phelps K (2013) The physical chemistry of cytoplasm and its influence on cell function: An update. *Mol Biol Cell* 24:2593–2596.
101. Feig M, Yu I, Wang PH, Nawrocki G, Sugita Y (2017) Crowding in cellular environments at an atomistic level from computer simulations. *J Phys Chem B* 121:8009–8025.
102. Rivas G, Minton AP (2018) Toward an understanding of biochemical equilibria within living cells. *Biophys Rev* 10:241–253.
103. Johnson ME, Hummer G (2011) Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks. *Proc Natl Acad Sci USA* 108:603–608.
104. Ivani I, et al. (2015) Parmbsc1: A refined force field for DNA simulations. *Nat Methods* 13:55–58.
105. Zgarbová M, et al. (2015) Refinement of the sugar-phosphate backbone torsion beta for AMBER force fields improves the description of Z- and B-DNA. *J Chem Theory Comput* 11:5723–5736.
106. Pérez A, et al. (2007) Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys J* 92:3817–3829.
107. Bana P, et al. (2010) Performance of molecular mechanics force fields for RNA simulations: Stability of UUCG and GNRA hairpins. *J Chem Theory Comput* 6:3836–3849.
108. Zgarbová M, et al. (2011) Refinement of the Cornell, et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J Chem Theory Comput* 7:2886–2902.
109. Tan D, Piana S, Dirks RM, Shaw DE (2018) RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc Natl Acad Sci USA* 115:201713027.
110. Vijay-Kumar S, Bugg CE, Cook WJ (1987) Structure of ubiquitin refined at 1.8 A resolution. *J Mol Biol* 194:531–544.
111. Ulmer TS, et al. (2003) Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. *J Am Chem Soc* 125:9179–9191.
112. Schwalbe H, et al. (2001) A refined solution structure of hen lysozyme determined using residual dipolar coupling data. *Protein Sci* 10:677–688.
113. McKnight CJ, Matsudaira PT, Kim PS (1997) NMR structure of the 35-residue villin headpiece subdomain. *Nat Struct Biol* 4:180–184.
114. Hess B (2002) Determining the shear viscosity of model liquids from molecular dynamics simulations. *J Chem Phys* 116:209.
115. Yeh IC, Hummer G (2004) System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *J Phys Chem B* 108:15873–15879.
116. Linke M (2018) Pydiffusion–A python software implementation to calculate the quaternion covariances from simulations, fit the rotational diffusion tensor, and run rotational Brownian dynamics simulations. Available at https://github.com/bio-phys/. Accessed April 10, 2018.
117. Favro LD (1960) Theory of the rotational Brownian motion of a free rigid body. *Phys Rev* 119:53–62.
118. Woessner DE (1962) Nuclear spin relaxation in ellipsoids undergoing rotational Brownian motion. *J Chem Phys* 37:647–654.