

1

2 **Supplementary Information for**  
3 **Dynamic Cluster Formation Determines Viscosity and Diffusion**  
4 **in Dense Protein Solutions**

5 **Sören von Bülow, Marc Siggel, Max Linke and Gerhard Hummer**

6 **Gerhard Hummer.**

7 **E-mail: [gerhard.hummer@biophys.mpg.de](mailto:gerhard.hummer@biophys.mpg.de)**

8 **This PDF file includes:**

- 9     Supplementary text
- 10    Figs. S1 to S21
- 11    Tables S1 to S3
- 12    Caption for Movie S1
- 13    References for SI reference citations

14 **Other supplementary materials for this manuscript include the following:**

- 15     Movie S1

## 16 Supporting Information Text

### 17 Supplementary Methods

18 **MD Simulations of Dense Protein Solutions** Initial configurations for all systems were generated using Monte  
19 Carlo simulations with a coarse-grained protein model (1) with purely repulsive interactions ( $\epsilon = 0.1 k_B T$ ).  
20 For the coarse-grained and all-atom simulations we used cubic simulation boxes of equal size. For the  
21 atomistic simulations, all proteins were protonated with the Maestro modeling package (Schroedinger  
22 (2)) at pH 7.0. All Asp and Glu residues of UBQ, GB3, VIL were deprotonated, only Glu35 of LYZ was  
23 protonated. Lys and Arg residues of all proteins were protonated. His68 of UBQ was neutral, protonated  
24 at the  $\delta$ -carbon atom. His15 of LYZ was neutral, protonated at the  $\epsilon$ -carbon atom. N and C termini of all  
25 proteins were unblocked and charged. The structure of VIL is destabilized by the water model and unfolds  
26 within several nanoseconds simulation time. We therefore applied harmonic distance restraints (force  
27 constant =  $200 \text{ kJmol}^{-1} \text{ nm}^{-2}$ ) on atom pairs 51CE1–73CB, 51CE1–58CZ, 47CE1–51E2, 58CE1–69CB,  
28 42CD1–51CE2, 47CZ–55CG, 58CD2–66CG and 69CB–73CG.

29 MD simulations were performed using GROMACS (v2016.3) (3) with the Amber99SB\*-ILDN-Q (4–7)  
30 force field, TIP4P-D water (8), and NaCl ion concentrations (9) of 0.157 M for UBQ solutions, 0.200 M  
31 for GB3 solutions, 0.120 M for LYZ solutions and 0.200 M for VIL solutions, respectively. Different ion  
32 concentrations were used for the different proteins to be consistent with respective experiments. The  
33 structures were energy-minimized and equilibrated at 300 K (10) with temperature coupling time constant  
34  $\tau_T = 0.1 \text{ ps}$  for 100 ps in an NVT ensemble and for 5 ns at 1 bar (11) with pressure coupling time constant  
35  $\tau_p = 5 \text{ ps}$  and compressibility  $4.5 \times 10^{-5} \text{ bar}^{-1}$  in an NPT ensemble. Production runs in an NPT ensemble  
36 (12) ( $\tau_T = 1.0 \text{ ps}$ ,  $\tau_p = 5.0 \text{ ps}$ ) were carried out at each density for times as listed in Table S1. The leap-frog  
37 integrator with timestep 2 fs was used for all simulations. Atomic bonds to hydrogens were converted to  
38 constraints with LINCS (13). Short-range electrostatics and van der Waals cutoffs were set to 1.0 nm in all  
39 simulations. Long-range electrostatic interactions were accounted for using the Particle Mesh Ewald (14)  
40 method. Configurations were recorded every 100 ps. Every tenth snapshot (1 ns timestep) was used for all  
41 analyses.

42 **Protein Structure Fluctuations.** To calculate root-mean-square deviations (RMSD) and root-mean-square  
43 fluctuations (RMSF), we superimposed the structures onto a reference structure using the qcprot RMSD  
44 alignment algorithm (15, 16) implemented in MDAnalysis (17, 18). As reference we used the average  
45 simulation structure, obtained by RMSD alignment of the  $\alpha$ -carbon backbone atoms (UBQ: residues 1–70,  
46 GB3: residues 1–58, LYZ: 1–129, VIL: 41–76) and averaging the simulation structures iteratively until  
47 convergence (19). The crystal structure was used for the initial alignment. The RMSD was calculated as

$$48 \text{RMSD}(t) = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\vec{x}_i(t) - \langle \vec{x}_i \rangle\|^2}, \quad [\text{S1}]$$

49 with  $N$  the number of  $\alpha$ -carbon backbone atoms and  $\langle \vec{x}_i \rangle$  the position of atom  $i$  of the averaged simulation  
50 structure. The RMSF were calculated as

$$51 \text{RMSF}_i = \sqrt{\langle \|\vec{x}_i - \langle \vec{x}_i \rangle\|^2 \rangle}. \quad [\text{S2}]$$

52 **Protein Volume Fraction.** To determine the protein volume fraction  $\phi$  at different protein concentrations,  
53 we used the NPT simulations of  $N = 15$  (UBQ) and  $N = 20$  (GB3, LYZ, VIL) proteins at different box  
54 sizes and numbers of water molecules (Table S1). We found that the equilibrated volumes of the simulation  
55 boxes depend linearly on the number of water molecules (Fig. S12). From the intercepts of linear fits  
56 divided by  $N$ , we obtained the effective protein volumes,  $v_p(\text{UBQ}) = 10.407 \text{ nm}^3$ ,  $v_p(\text{GB3}) = 7.222 \text{ nm}^3$ ,  
57  $v_p(\text{LYZ}) = 17.228 \text{ nm}^3$  and  $v_p(\text{VIL}) = 5.118 \text{ nm}^3$ , respectively. The protein volume fraction in simulations  
58 with  $N$  proteins was then calculated as  $\phi = N v_p / \langle V \rangle$ , where  $\langle V \rangle$  is the box volume averaged over an NPT  
59 simulation.

60 **Viscosity Calculation.** The low-frequency, low-shear viscosity  $\eta$  of dense protein solutions differs from the  
61 viscosity  $\eta_0$  of the pure solvent consisting of only water and ions. We determined  $\eta(\phi)$  and  $\eta_0$  from MD  
62 simulations in an NVT ensemble by integration of the autocorrelation functions  $C_{ij}(t) = \langle P_{ij}(t)P_{ij}(0) \rangle$  of  
63 the pressure tensor fluctuations (20),

$$64 \quad \eta_{ij} = \frac{V}{k_B T} \int_0^\infty C_{ij}(t) dt, \quad [S3]$$

65 where  $C_{ij}(t)$  was determined for the three off-diagonal pressure tensor elements ( $P_{ij} = P_{xy}, P_{xz},$  and  $P_{yz}$ )  
66 and the three corresponding combinations of the diagonal pressure tensor elements [ $P_{ij} = (P_{xx} - P_{yy})/2,$   
67  $(P_{xx} - P_{zz})/2,$  and  $(P_{yy} - P_{zz})/2$ ].  $V$  is the (fixed) volume of the simulation box,  $T$  is the absolute  
68 temperature, and  $k_B$  is the Boltzmann constant. Starting structures for the NVT runs were taken from the  
69 long NPT simulations (see below).

70 Numerical integration of the tails of  $C_{ij}(t)$  is prone to large errors due to sampling noise. We therefore  
71 split the integration of  $C_{ij}(t)$  into a direct integration up to a cutoff time  $\tau_{\text{cut}}$  (see Table S2) and an  
72 analytical integration of a bi-exponential fit to the tail,

$$73 \quad C_{ij}(t) = a_0 e^{-t/\tau_0} + a_1 e^{-t/\tau_1} \text{ for } t > \tau_{\text{cut}}, \quad [S4]$$

74 where  $a_0, a_1, \tau_0,$  and  $\tau_1$  denote the fit parameters. The left ( $\tau_{\text{cut}}$ ) and right bounds of the fit are listed in  
75 Table S2. The fit function was integrated analytically from  $\tau_{\text{cut}}$  to  $\tau = \infty$ .

76 The viscosity of TIP4P-D water (8) with different ion concentrations was calculated by evaluating  $P_{ij}(t)$   
77 at 10 fs intervals from 100 ns simulations in the NVT ensemble.  $C_{ij}$  of  $P_{ij}$  was integrated to give  $\eta_{ij}$ .  $\eta$  was  
78 then calculated by averaging  $\eta_{ij}$  for the six  $P_{ij}$  with the standard error of the mean used as error estimate.  
79 To efficiently sample the viscosity of the dense protein solutions, we extracted fifty starting configurations  
80 from the respective trajectories and conducted NVT simulations from each of the starting configurations.  
81 Each extracted configuration was equilibrated for 1 ns (NPT) followed by 5 ns of simulation at constant  
82 volume (NVT), evaluating  $P_{ij}$  every 10 fs.  $C_{ij}$  was averaged over the fifty simulations and integrated to  
83 give  $\eta_{ij}$ .  $\eta_{ij}$  was then averaged as above to obtain the viscosity  $\eta$ .

**Translational Diffusion.** Mean squared displacements (MSD) were calculated for each density for time  
delays  $\tau = 1$  ns to  $\tau = 500$  ns. For each protein in the simulation box, long-time translational diffusion  
coefficients were obtained by fitting the Einstein relation

$$84 \quad \text{MSD}(\tau) = c + 6D_t^{\text{PBC}}\tau, \quad \text{for } \tau \rightarrow \infty \quad [S5]$$

85 to the MSD curves in the range 10 to 30 ns, where  $c$  is a fitted offset that accounts for short-time non-diffusive  
86 behavior. The diffusion coefficients of all proteins in the simulation box were then averaged. The error of  
87  $D_t$  for proteins in the dense solutions was estimated from the standard error of the mean of all proteins in  
88 the simulation box. The error of  $D_t$  in the dilute UBQ and GB3 system ( $N = 1$ ) was estimated from 1000  
89 one-dimensional random walks of lengths corresponding to the simulation trajectories (Table S1). From the  
90 random walks, MSDs were calculated and fitted to the Einstein relation in the range 0 to 5 ns to obtain  
91  $D_t$  (in analogy to the calculation of the protein's  $D_t$ ). The estimated error of  $D_t$  was then obtained from  
92  $\sigma/\sqrt{3-1}$ , with  $\sigma$  the standard deviation of the distribution of  $D_t$  from the random walks. The denominator  
93 corrects for the three dimensions in the real system compared to the one-dimensional random walk. For  
94 dilute LYZ and VIL solutions, two independent simulations were carried out, respectively, to determine  
95  $D_t^{\text{PBC}}$ . Finite-size corrected  $D_t$  was obtained from the y-intercept of fitting the finite-size correction formula  
96 (Eq. 7) to  $D_t^{\text{PBC}}(1/L)$ . The error of dilute LYZ and VIL solutions was obtained from the curvature of the  
97 likelihood in the fitting procedure, taking into account the individual errors of  $D_t^{\text{PBC}}$ , which were obtained  
98 from one-dimensional random walks, as for the dilute UBQ and GB3 solutions.

99 We compared the reduced translational diffusion coefficients  $D_t(\phi)/D_{t,\phi=0}$  to the approximate form for  
monodisperse non-interacting HS colloidal suspensions (21),

$$100 \quad D_{t,\text{coll}}(\phi)/D_{t,\phi=0} = \frac{(1-\phi)^3}{1+3/2\phi+2\phi^2+3\phi^3}, \quad [S6]$$

101 which was obtained as an extension of Batchelor’s quadratic approximation (22). We also compared reduced  
 102 translational diffusion coefficients to an alternative approximation of the diffusive behavior of monodisperse  
 103 non-interacting HS colloidal suspensions (including hydrodynamic interactions) (23),

$$104 \quad D_{t,\text{coll}}(\phi)/D_{t,\phi=0} = \frac{1 - 9\phi/32}{1 + H(\phi) + (\phi/\phi_0)/(1 - \phi/\phi_0)^2}, \quad [\text{S7}]$$

105 with  $\phi_0 \approx 0.5718$  and

$$106 \quad H(\phi) = \frac{2b(\phi)^2}{1 - b(\phi)} - \frac{c(\phi)}{1 + 2c(\phi)} - \frac{b(\phi)c(\phi)(2 + c(\phi))}{(1 + c(\phi))(1 - b(\phi) + c(\phi))}, \quad [\text{S8}]$$

107 with  $b(\phi) = (9\phi/8)^{1/2}$  and  $c(\phi) = 11\phi/16$ .

108 **Rotational Diffusion Coefficient from the Orientational Correlation Function.** The mean rotational diffusion  
 109 coefficient  $\bar{D}_r^{\text{PBC}}$  was calculated as

$$110 \quad \bar{D}_r^{\text{PBC}} = \frac{1}{3}(D_1 + D_2 + D_3), \quad [\text{S9}]$$

111 with  $D_1$ ,  $D_2$ , and  $D_3$  the rotational diffusion coefficients in the principal coordinate system of the rotation  
 112 diffusion tensor. The finite size effect of the mean rotation diffusion coefficient was corrected using (24)

$$113 \quad \bar{D}_r = \bar{D}_r^{\text{PBC}} + \frac{k_{\text{B}}T}{6\eta(\phi)V}, \quad [\text{S10}]$$

114 with  $V$  the mean box volume.

115 Additionally, the effective rotational diffusion coefficient was obtained from fits to the orientational  
 116 correlation function  $\langle\langle P_1(\cos \theta(t)) \rangle\rangle = \langle\langle \cos \theta(t) \rangle\rangle$  (25, 26), where  $\theta(t)$  is the angle traveled during time  $t$ ,  
 117 with  $\cos \theta(t) = \vec{v}(t) \cdot \vec{v}(0)$  and  $\vec{v}$  a unit vector associated with the protein (19). Here,  $\langle\langle \dots \rangle\rangle$  indicates  
 118 averaging over starting times and isotropic orientations of  $\vec{v}$ . We approximate the rotational behavior of the  
 119 system by assuming two isotropically tumbling species with distinct tumbling rates, one for fast-tumbling  
 120 free proteins and the other for slower-tumbling proteins in clusters. Therefore, a bi-exponentially decaying  
 121 function was fitted to the data

$$122 \quad \langle\langle P_1(\cos \theta(t)) \rangle\rangle = a_1 e^{-t/\tau_1} + (1 - a_1) e^{-t/\tau_2}, \quad [\text{S11}]$$

123 with fit parameters  $a_1$ ,  $\tau_1$ , and  $\tau_2$ . In the infinitely diluted system, Eq. S11 reduces to  $\langle\langle P_1(\cos \theta(t)) \rangle\rangle =$   
 124  $e^{-t/\tau_1}$  with  $\tau_1$  the only fitting parameter. Integration of  $\langle\langle P_1(\cos \theta(t)) \rangle\rangle$  yields

$$125 \quad \tau_c = a_1 \tau_1 + (1 - a_1) \tau_2 \quad [\text{S12}]$$

126 and  $\tau_c = \tau_1$  in the dilute system. An effective diffusion coefficient was obtained by averaging the contributions  
 127 from slow and fast rotational diffusion, weighting the individual correlation times  $\tau_1$  and  $\tau_2$  by the exponential  
 128 prefactors  $a_1$  and  $1 - a_1$

$$129 \quad \tilde{D}_r^{\text{PBC}} = \frac{a_1}{2\tau_1} + \frac{1 - a_1}{2\tau_2}. \quad [\text{S13}]$$

130  $\tilde{D}_r$  was obtained from  $\tilde{D}_r^{\text{PBC}}$  by correcting for finite-size effects using Eq. S10.

131 We compared the reduced rotational diffusion coefficients  $D_r(\phi)/D_{r,\phi=0}$  to predictions from colloidal  
 132 models of non-interacting HS (27):

$$133 \quad D_{r,\text{coll}}(\phi)/D_{r,\phi=0} = (1 - \phi) \frac{\eta_0}{\eta(\phi)}. \quad [\text{S14}]$$

134 **Hydropro Calculations.** Translational and rotational diffusion coefficients  $D_t$  and  $\bar{D}_r$  were also calculated  
 135 using the hydrodynamics program Hydropro (28) for  $T = 300$  K, the “atom” model, and viscosities  
 136  $\eta = 0.937$  mPa s, corresponding to the average pure-solvent values with ions (TIP4P-D water (8) with  
 137 0.120 M, 0.157 M and 0.200 M NaCl).

138 **Protein Cluster Formation.** Protein clusters with size  $m$  were defined based on a distance criterion. A  
 139 cluster encompasses all proteins with minimal  $\alpha$ -carbon distance  $< 0.65$  nm to at least one other protein  
 140 in the cluster. Only pairs staying in contact for at least 1 ns were considered, which is the time interval  
 141 between consecutive structures in the analyzed trajectories. To assess the influence of slight changes in  
 142 the distance cutoff criterion, the cutoff distance was varied from 0.6 to 0.75 nm (Fig. S11), finding that a  
 143 cutoff of 0.65 nm leads to the good agreement of our cluster model with diffusion data also at the highest  
 144 concentration (200 mg/ml). At this concentration, the mean cluster size  $\bar{m}$  depends significantly on the  
 145 distance cutoff. Therefore, the calculation of cluster sizes based on a simple  $\alpha$ -carbon distance criterion  
 146 employed here is not reliable above 100 mg/ml.

147 The lifetime of protein pairs was analyzed to determine if cluster formation is transient, dynamic, or  
 148 permanent (following Liu's definition (29)). We analyzed the distributions of protein pair lifetimes, which  
 149 we defined as the mean time in which the minimal distance between at least one pair of  $\alpha$ -carbon atoms of  
 150 two proteins does not exceed 0.65 nm.

151 We calculated the preferred contact interfaces of proteins by counting the number of heavy atom contacts.  
 152 A three-tiered approach was employed to reduce the computational time needed for the analysis. First, the  
 153 distances of the protein centers of mass (COM) were calculated. If the proteins were close (as identified by  
 154 a COM distance  $\leq 6$  nm), then the  $\alpha$ -carbon distances were computed. If at least one  $\alpha$ -carbon distance  
 155 was below the cutoff 0.65 nm, then all heavy-atom distances in the protein pair were calculated. The  
 156 heavy-atoms involved in intermolecular heavy-atom pairs with distance  $\leq 0.5$  nm were counted per residue.  
 157 Figs. 4 and S17 show the proteins color-coded by the counts.

158 **Relation Between Cluster Size and Protein Binding Affinity.** Consider a system of identical proteins at a  
 159 number concentration  $\rho$  (i.e., proteins per unit volume). We assume that the protein clusters  $C_m$  of size  $m$   
 160 are in equilibrium with each other,



162 with concentrations  $c_m = [C_m]$  that satisfy binding equilibria

$$163 \quad \frac{c_{m'}c_{m''}}{c_{m'+m''}} = K_d \quad [S16]$$

164 for all  $m', m'' \geq 1$  with  $K_d$  an effective dissociation constant. For simplicity, we assume  $K_d$  to be independent  
 165 of  $m'$  and  $m''$ . In effect, this model assumes that a cluster of size  $m$  is held together by  $m - 1$  energetically  
 166 equivalent interactions as, e.g., in a Cayley tree (Bethe lattice). We show that in our simulations of  
 167 concentrated protein solutions the number of protein-protein connections (a connection between two  
 168 proteins is established if at least one  $\alpha$ -carbon distance is  $< 0.65$  nm) indeed increases as  $m - 1$  with cluster  
 169 size  $m$  up to intermediate (100 mg/ml) protein concentration (Fig. S21). The concentrations  $c_m$  of the  
 170 different cluster sizes then satisfy

$$171 \quad c_m = c_1^m / K_d^{m-1} . \quad [S17]$$

172 It is easily verified that for these  $c_m$ , with an as yet unknown monomer concentration  $c_1$ , Eq. S16 is satisfied.  
 173 Using this expression for  $c_m$ , we obtain a relation between the total concentration of proteins  $\rho$  and the  
 174 concentration of monomers  $c_1$ ,

$$175 \quad \sum_{m=1}^{\infty} m c_m = \frac{c_1}{(1 - c_1/K_d)^2} = \rho , \quad [S18]$$

176 which can be solved to express the monomer concentration  $c_1$  in terms of  $\rho$  and  $K_d$ . The mean cluster size  
 177 is

$$178 \quad \bar{m} = \frac{\sum_{m=1}^{\infty} m c_m}{\sum_{m=1}^{\infty} c_m} = \frac{1}{1 - c_1/K_d} . \quad [S19]$$

179 Substituting the solution of Eq. S18, we obtain for the mean cluster size

$$180 \quad \bar{m} = \frac{1 + \sqrt{1 + 4\rho/K_d}}{2} \approx 1 + \rho/K_d , \quad [S20]$$

181 where the last expression is the approximation for low protein concentrations  $\rho$ . We thus find that the mean  
 182 cluster size grows approximately as  $\bar{m} \approx 1 + \rho/K_d = 1 + \phi/(v_p K_d) = 1 + \zeta\phi$  at low protein concentrations,  
 183 with  $v_p$  the protein volume,  $\phi$  the protein volume fraction, and  $\zeta = 1/(v_p K_d)$  the clustering propensity.

184 **Colloidal Suspension Model.** Colloidal particles with attractive interactions are often modeled as sticky  
 185 hard spheres (30) of radius  $a$  and diameter  $\sigma = 2a$  with interaction potentials that depend on the pair  
 186 distance  $r$  as

$$187 \quad \beta u(r) = \begin{cases} \infty, & \text{for } r < \sigma \\ \ln [12\tau(d/\sigma - 1)], & \text{for } \sigma < r < d \\ 0, & \text{for } r > d \end{cases} \quad [S21]$$

188 in the limit of  $d \rightarrow \sigma$ . In the Monte Carlo simulations of the mean cluster size reported in the main text,  
 189 we used a finite  $d = 1.05\sigma$ . At low concentrations, the radial distribution function can be approximated as  
 190 (31)  $g(r) = \theta(r - \sigma) + \delta(r - \sigma)\sigma/12\tau$  where  $\theta(x)$  is the Heaviside step function and  $\delta(x)$  is Dirac's delta  
 191 function. In this limit, the fraction of bound particles is obtained by integration over the spherical shell,  
 192  $f_{\text{bound}} = \lim_{\epsilon \rightarrow 0} 2\pi\rho \int_{2a-\epsilon}^{2a+\epsilon} r^2 dr g(r) = \rho v_{\text{HS}}/\tau$  with  $v_{\text{HS}} = 4\pi a^3/3$  the hard-sphere volume and  $\phi = v_{\text{HS}}\rho$ .  
 193 In an equilibrium binding model, this fraction is  $K_a\rho$ . We thus arrive at a relation between the association  
 194 constant and Baxter parameter  $\tau = v_{\text{HS}}/K_a = v_{\text{HS}}K_d$ .

195 **Displacement Pair Correlation Function.** We calculated the displacement pair correlation function  $H_{ij}$  for  
 196 the motion of proteins  $i$  and  $j$  following the method of Ando and Skolnick (32),

$$197 \quad H_{ij}(\tau) = \frac{\sum \Delta \vec{r}_i(\tau) \cdot \Delta \vec{r}_j(\tau)}{\sqrt{\sum \|\Delta \vec{r}_i(\tau)\|^2} \sqrt{\sum \|\Delta \vec{r}_j(\tau)\|^2}}, \quad [S22]$$

198 with  $\Delta \vec{r}_i(\tau)$  the vector traveled by protein  $i$  in time delay  $\tau$ . We evaluated  $H_{ij}$  for varying delays  
 199  $\tau = 2, 4, \dots, 40$  ns and 0.1 nm-wide minimum  $\alpha$ -carbon distance windows from 0.6–3.0 nm. Here, only proteins  
 200 are considered if their minimum  $\alpha$ -carbon distance falls in the respective distance window both at time  $t$   
 201 and  $t + \tau$ .

## 202 Supplementary Results

203 **Effect of Concentration on Protein Stability.** Proteins in commonly employed protein force fields are too  
 204 sticky, causing fast and irreversible aggregation in crowded solutions (33). In our atomistic MD simulations  
 205 of dense and dilute protein solutions, we therefore used the TIP4P-D water model (8). An increase in the  
 206 water-protein dispersion interaction attenuates the aggregation propensity. However, the TIP4P-D water  
 207 model slightly destabilizes native protein structures (8).

208 UBQ  $\alpha$ -carbon atoms show a mean RMSD of 1.2–1.5 Å to the average simulation structure throughout  
 209 all simulations (Fig. S1, left panels). The RMSD of the average simulation structure to the crystal structure  
 210 is likewise in the range 1.3–1.5 Å. RMSF confirm a very similar flexibility of dilute and dense UBQ solutions  
 211 (Fig. S4), showing that the major contribution to the RMSD is from the C-terminal flexible tail (residues  
 212 71–76). In the simulations, the internal structure and dynamics of UBQ is unperturbed by protein crowding  
 213 and UBQ remains folded.

214 GB3 is slightly destabilized in the dilute solution and its  $\alpha$ -helix begins to unfold at ca. 2  $\mu\text{s}$  simulation  
 215 time. The mean RMSD values of GB3  $\alpha$ -carbon atoms to the average simulation structure in the dense  
 216 solutions are in the range 1.0–1.2 Å and thus consistently smaller than the RMSD values of the dilute  
 217 protein (0–2  $\mu\text{s}$ ) at 1.4 Å (Fig. S2, right panels), suggesting a stabilizing effect of macromolecular crowding  
 218 on the native state of GB3. The RMSD values of the mean simulation structures to the crystal structure are  
 219 0.9–1.1 Å for the dense solutions and 1.6 Å for the dilute solution (0–2  $\mu\text{s}$ ). RMSF calculations confirm that  
 220 the residues 28–42 of the  $\alpha$ -helix fluctuate more in dilute solution than in the dense GB3 solutions (Fig. S4).  
 221 This finding, although in accordance with the theory of entropic stabilization of the native state due to  
 222 excluded volume, contrasts with a previous study (34), in which non-specific, shape-driven interactions led

223 to native state destabilization in crowded conditions. Although partial unfolding of the  $\alpha$ -helix has only a  
224 minor effect on the translational and rotational diffusional behavior of the protein (not shown), we only  
225 considered the first  $2 \mu\text{s}$  of the dilute GB3 simulation for further analysis.

226 LYZ is stable throughout the simulation, with mean RMSD values to the average simulation structure of  
227  $1.0\text{--}1.1 \text{ \AA}$  in all simulations. The RMSD values of the mean simulation structures to the crystal structure  
228 are  $1.5\text{--}1.9 \text{ \AA}$  and RMSF values of the dilute systems indicate high flexibility in the loop of residues 100–105,  
229 which is reduced in the dense solutions (Fig. S4).

230 VIL is the least stable protein model of the proteins studied. The unrestrained model readily unfolds in  
231 TIP4P-D water. When applying restraints to key residues in the hydrophobic core of the protein, the model  
232 remains stable throughout the simulation with mean RMSD values to the average simulation structure of  
233  $1.2\text{--}1.7 \text{ \AA}$  in all simulations. The RMSD values of the mean simulation structures to the crystal structure  
234 are  $1.6\text{--}1.8 \text{ \AA}$ . One of the two simulations of the dilute VIL systems shows somewhat higher flexibility than  
235 all other simulations (Fig. S4). Nevertheless, the restrained protein remains folded in all cases.

**Table S1. Specifications of simulated dense protein solutions.<sup>a</sup>**

| Protein | N   | Concentration<br>[mg/ml] | $\phi$      | Atoms     | L<br>[nm] | $t_{\text{total}}$<br>[ns] |
|---------|-----|--------------------------|-------------|-----------|-----------|----------------------------|
| UBQ     | 1   | $\approx 0$              | $\approx 0$ | 36,000    | 6.464     | 5000                       |
|         | 15  | 30                       | 0.0221      | 933,000   | 19.178    | 2000                       |
|         | 15  | 50                       | 0.0370      | 556,000   | 16.152    | 2000                       |
|         | 15  | 100                      | 0.0745      | 276,000   | 12.797    | 2000                       |
|         | 15  | 200                      | 0.1503      | 135,000   | 10.127    | 2000                       |
|         | 120 | 200                      | 0.1509      | 1,079,000 | 20.227    | 1000                       |
|         | 405 | 200                      | 0.1503      | 3,657,000 | 30.382    | 1000                       |
| GB3     | 1   | $\approx 0$              | $\approx 0$ | 23,000    | 5.552     | 3000                       |
|         | 20  | 39                       | 0.0277      | 690,000   | 17.345    | 2000                       |
|         | 20  | 100                      | 0.0714      | 266,000   | 12.643    | 2000                       |
|         | 20  | 200                      | 0.1447      | 130,000   | 9.994     | 2000                       |
|         | 160 | 200                      | 0.1447      | 1,043,216 | 19.986    | 1000                       |
|         | 540 | 200                      | 0.1447      | 3,520,854 | 29.979    | 1000                       |
| LYZ     | 1   | $\approx 0$              | $\approx 0$ | 46,990    | 7.092     | 1105                       |
|         | 1   | $\approx 0$              | $\approx 0$ | 69,688    | 8.082     | 1825                       |
|         | 20  | 100                      | 0.0733      | 617,202   | 16.753    | 1001                       |
|         | 20  | 200                      | 0.1459      | 306,738   | 13.315    | 1083                       |
| VIL     | 1   | $\approx 0$              | $\approx 0$ | 18,346    | 5.180     | 906                        |
|         | 1   | $\approx 0$              | $\approx 0$ | 49,106    | 7.186     | 1328                       |
|         | 20  | 100                      | 0.0754      | 178,266   | 11.070    | 1042                       |
|         | 20  | 200                      | 0.1517      | 87,830    | 8.770     | 1251                       |

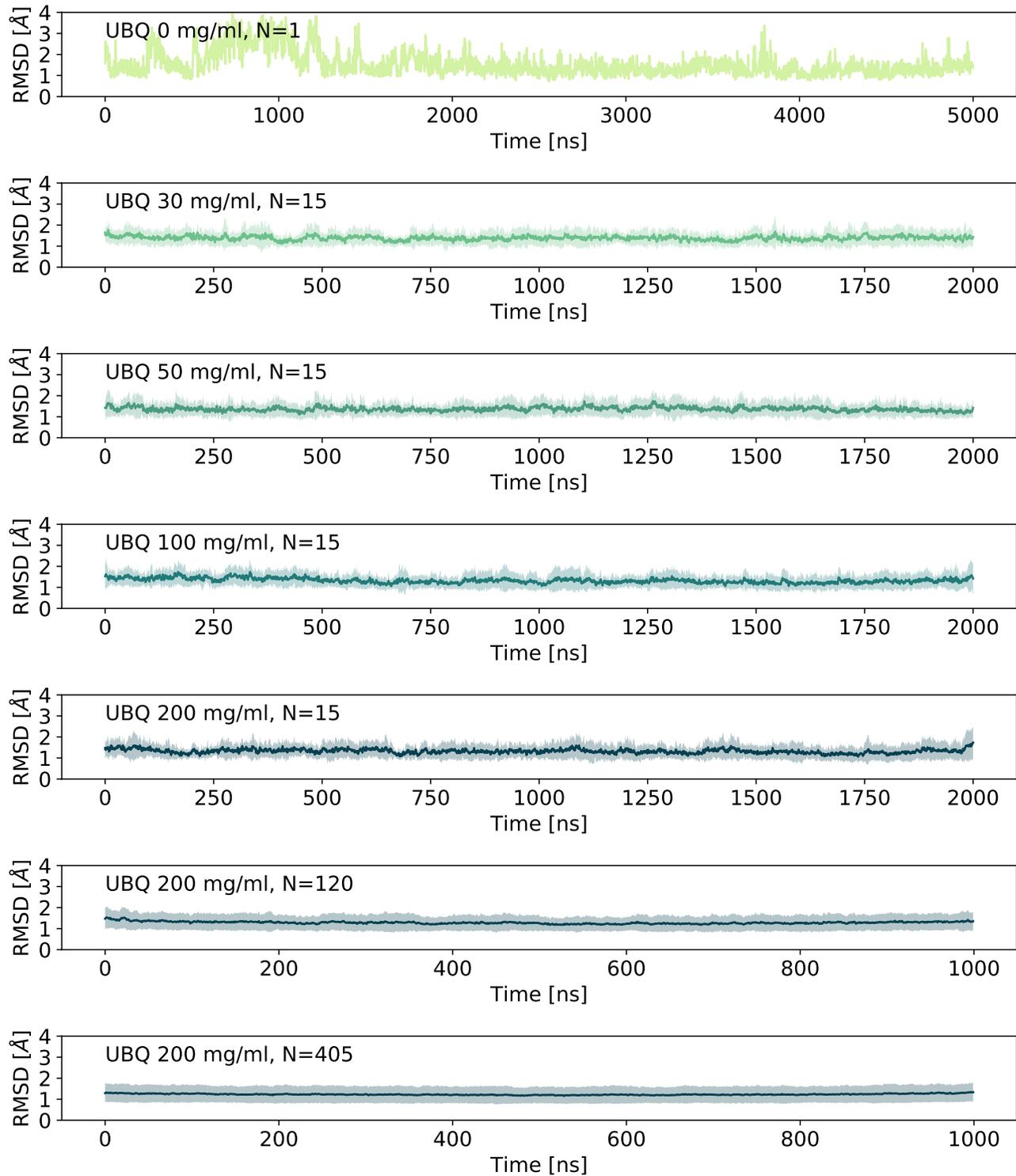
<sup>a</sup> The concentration of proteins is approximate. N: Number of proteins in the simulation box.  $\phi$ : Mean protein volume fraction. Atoms: Total number of atoms in the simulation box. L: Edge length of cubic simulation box.  $t_{\text{total}}$ : Total simulation time.

**Table S2. Time windows used to fit  $C_{ij}(t)$ .**

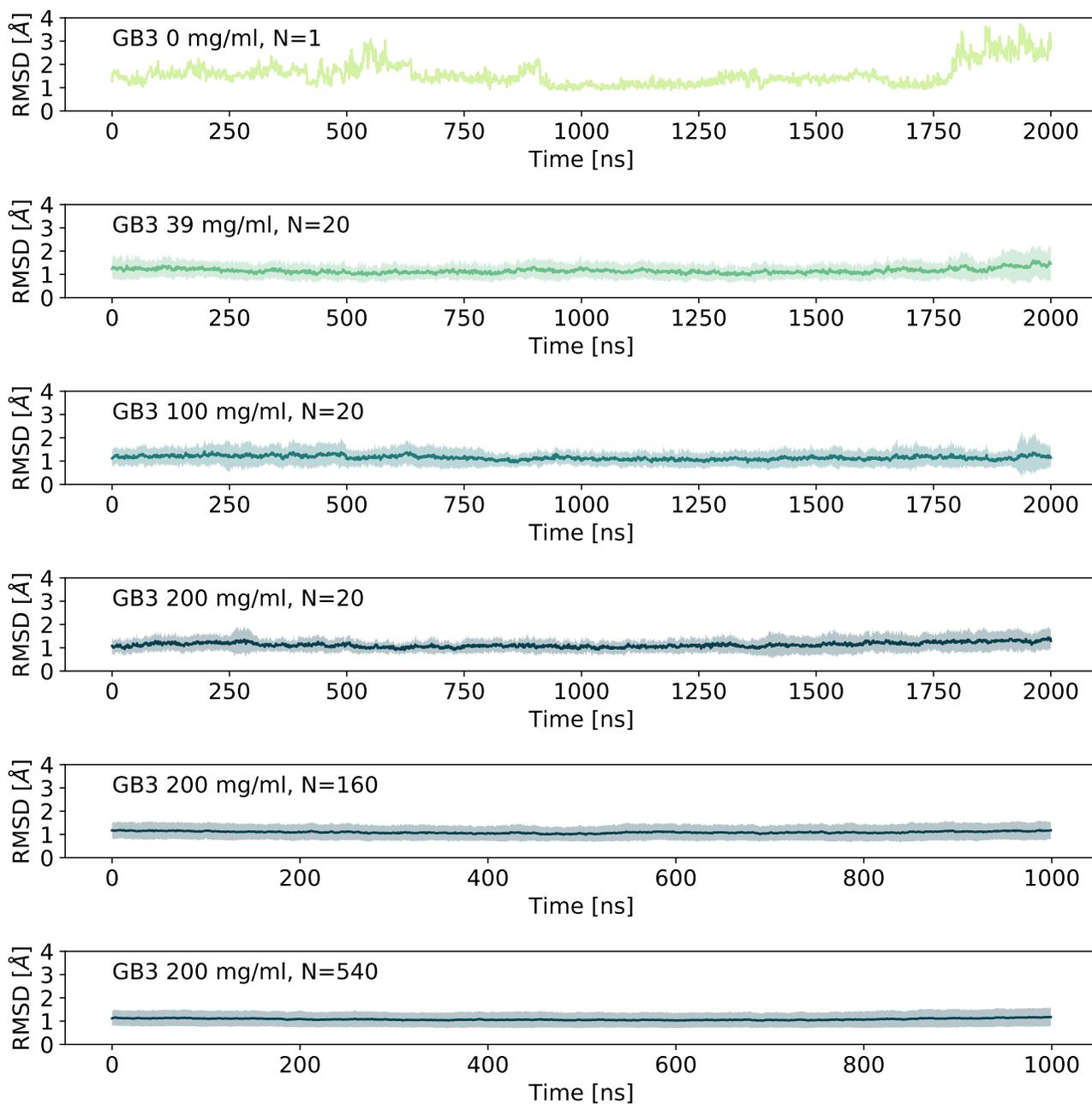
| System  |             | Left bound $\tau_{\text{cut}}$ | Right bound |
|---------|-------------|--------------------------------|-------------|
| TIP4P-D | 0 mM NaCl   | 1.0 ps                         | 4.0 ps      |
|         | 157 mM NaCl | 1.0 ps                         | 4.0 ps      |
|         | 200 mM NaCl | 1.0 ps                         | 4.0 ps      |
| UBQ     | 30 mg/ml    | 1.5 ps                         | 7.0 ps      |
|         | 50 mg/ml    | 1.5 ps                         | 6.5 ps      |
|         | 100 mg/ml   | 1.5 ps                         | 8.0 ps      |
|         | 200 mg/ml   | 1.5 ps                         | 10.0 ps     |
| GB3     | 39 mg/ml    | 1.5 ps                         | 5.5 ps      |
|         | 100 mg/ml   | 1.5 ps                         | 6.0 ps      |
|         | 200 mg/ml   | 1.5 ps                         | 12.0 ps     |
| LYZ     | 100 mg/ml   | 1.5 ps                         | 6.0 ps      |
|         | 200 mg/ml   | 1.5 ps                         | 9.0 ps      |
| VIL     | 100 mg/ml   | 1.5 ps                         | 7.0 ps      |
|         | 200 mg/ml   | 1.5 ps                         | 15.0 ps     |

**Table S3. Translational and rotational diffusion coefficients before finite-size correction.**

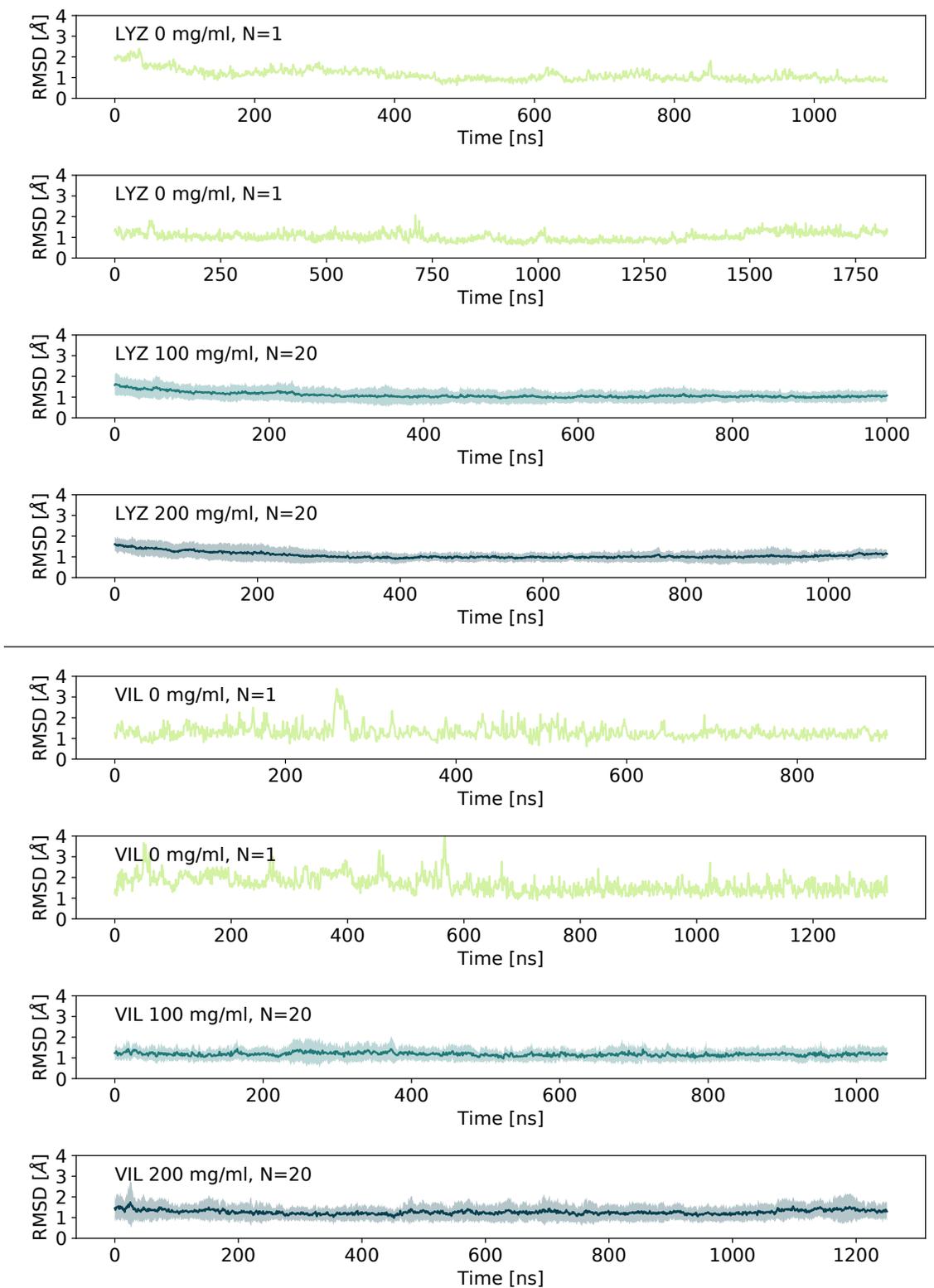
| System |           | N   | $D_t^{\text{PBC}}$<br>$10^{-7} \text{ cm}^2 \text{ s}^{-1}$ | $\bar{D}_r^{\text{PBC}}$<br>$10^7 \text{ s}^{-1}$ |
|--------|-----------|-----|---|---|
| UBQ    | 0 mg/ml   | 1   | 4.56  | 3.70  |
|        | 30 mg/ml  | 15  | 8.81  | 2.69  |
|        | 50 mg/ml  | 15  | 7.32  | 2.27  |
|        | 100 mg/ml | 15  | 4.55  | 1.49  |
|        | 200 mg/ml | 15  | 1.58  | 0.56  |
|        | 200 mg/ml | 120 | 2.30  | 0.58  |
|        | 200 mg/ml | 405 | 2.19  | 0.42  |
| GB3    | 0 mg/ml   | 1   | 4.60  | 4.30  |
|        | 39 mg/ml  | 20  | 9.70  | 3.28  |
|        | 100 mg/ml | 20  | 5.27  | 1.90  |
|        | 200 mg/ml | 20  | 2.05  | 0.88  |
|        | 200 mg/ml | 160 | 2.59  | 1.06  |
|        | 200 mg/ml | 540 | 2.92  | 0.94  |
| LYZ    | 0 mg/ml   | 1   | 3.66  | 1.76  |
|        | 0 mg/ml   | 1   | 4.39  | 2.30  |
|        | 100 mg/ml | 20  | 5.47  | 1.67  |
|        | 200 mg/ml | 20  | 2.21  | 0.61  |
| VIL    | 0 mg/ml   | 1   | 6.13  | 6.25  |
|        | 0 mg/ml   | 1   | 8.84  | 6.59  |
|        | 100 mg/ml | 20  | 5.78  | 2.60  |
|        | 200 mg/ml | 20  | 1.92  | 1.05  |



**Fig. S1.** RMSD of UBQ  $\alpha$ -carbon atoms to the average simulation structure in simulations of dilute and dense solutions. Colored curves show mean RMSD values, averaged for all proteins  $N$  in the simulation box. Shades indicate the standard deviation of the distribution of RMSD values of individual proteins.



**Fig. S2.** RMSD of GB3  $\alpha$ -carbon atoms to the average simulation structure in simulations of dilute and dense solutions. Colored curves show mean RMSD values, averaged for all proteins  $N$  in the simulation box. Shades indicate the standard deviation of the distribution of RMSD values of individual proteins.



**Fig. S3.** RMSD of LYZ and VIL  $\alpha$ -carbon atoms to the average simulation structure in simulations of dilute and dense solutions. Colored curves show mean RMSD values, averaged for all proteins  $N$  in the simulation box. Shades indicate the standard deviation of the distribution of RMSD values of individual proteins.

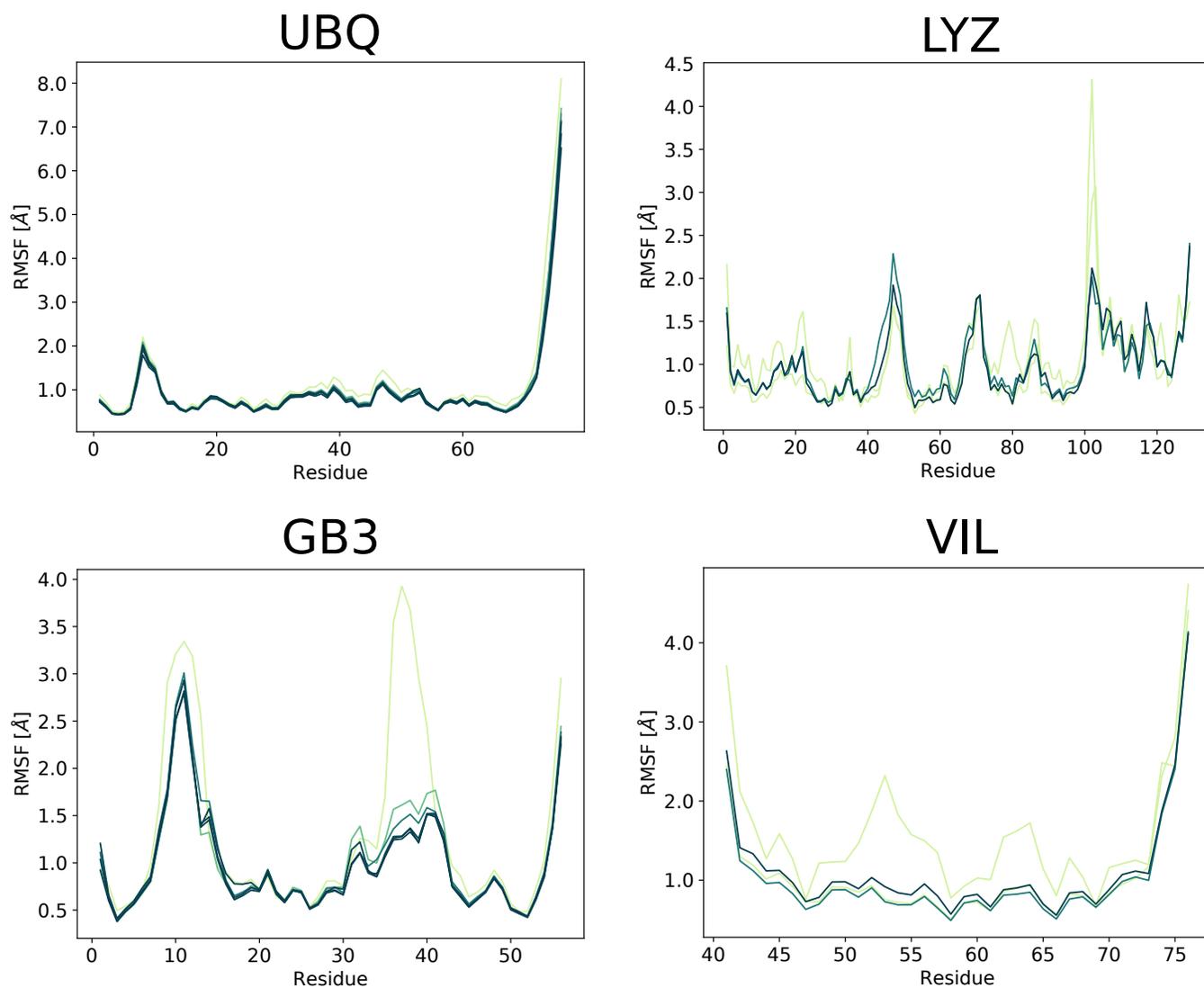
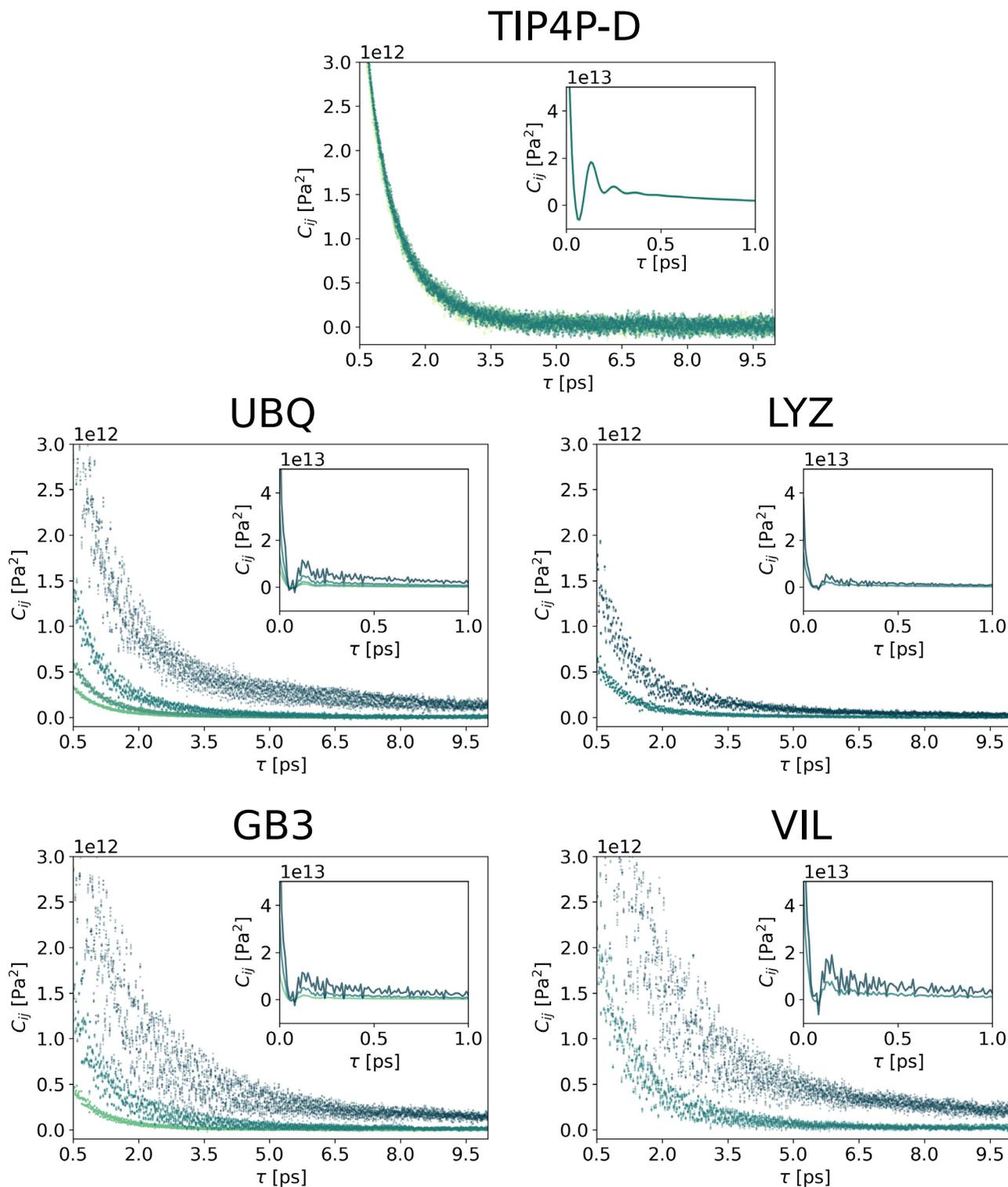
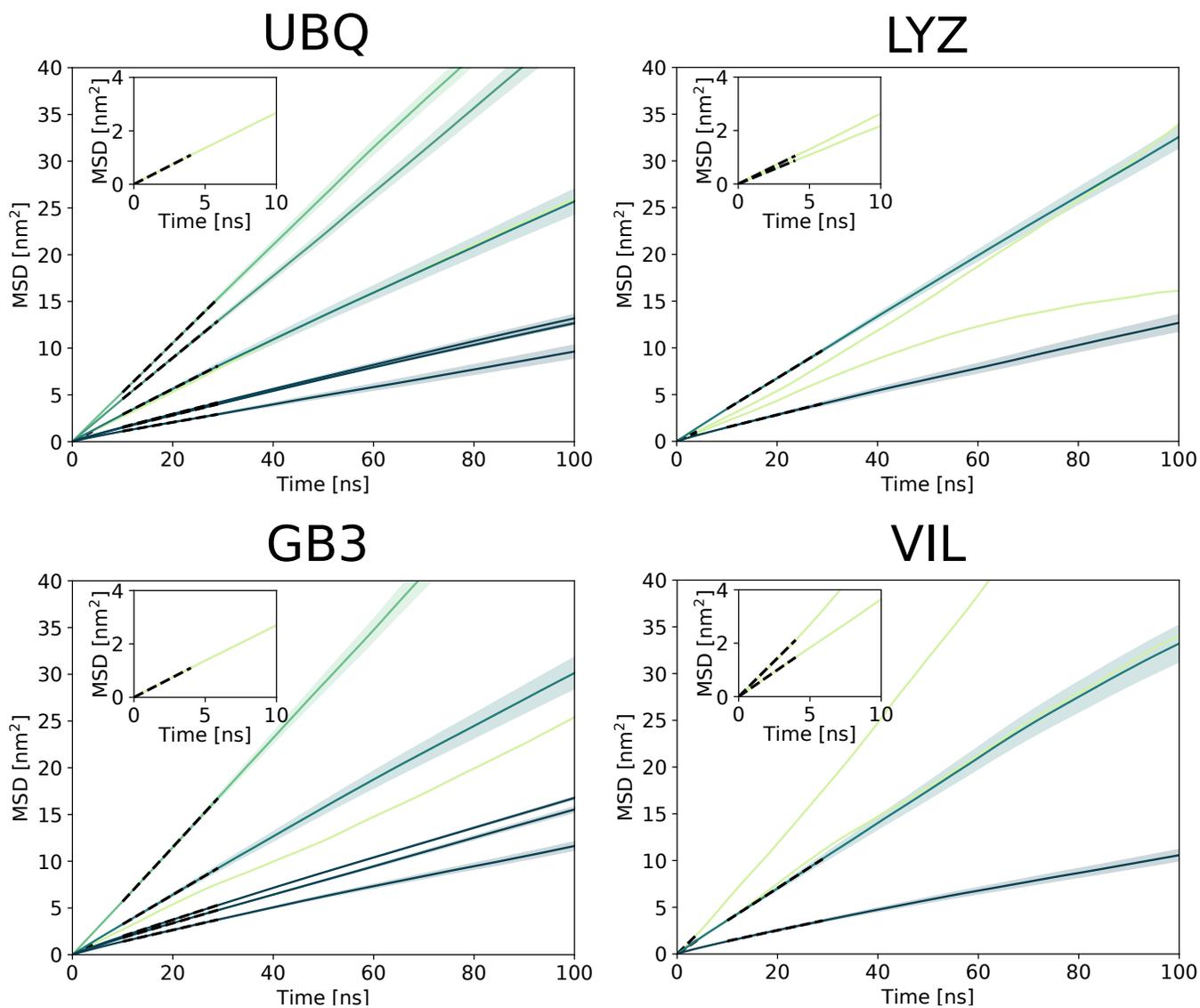


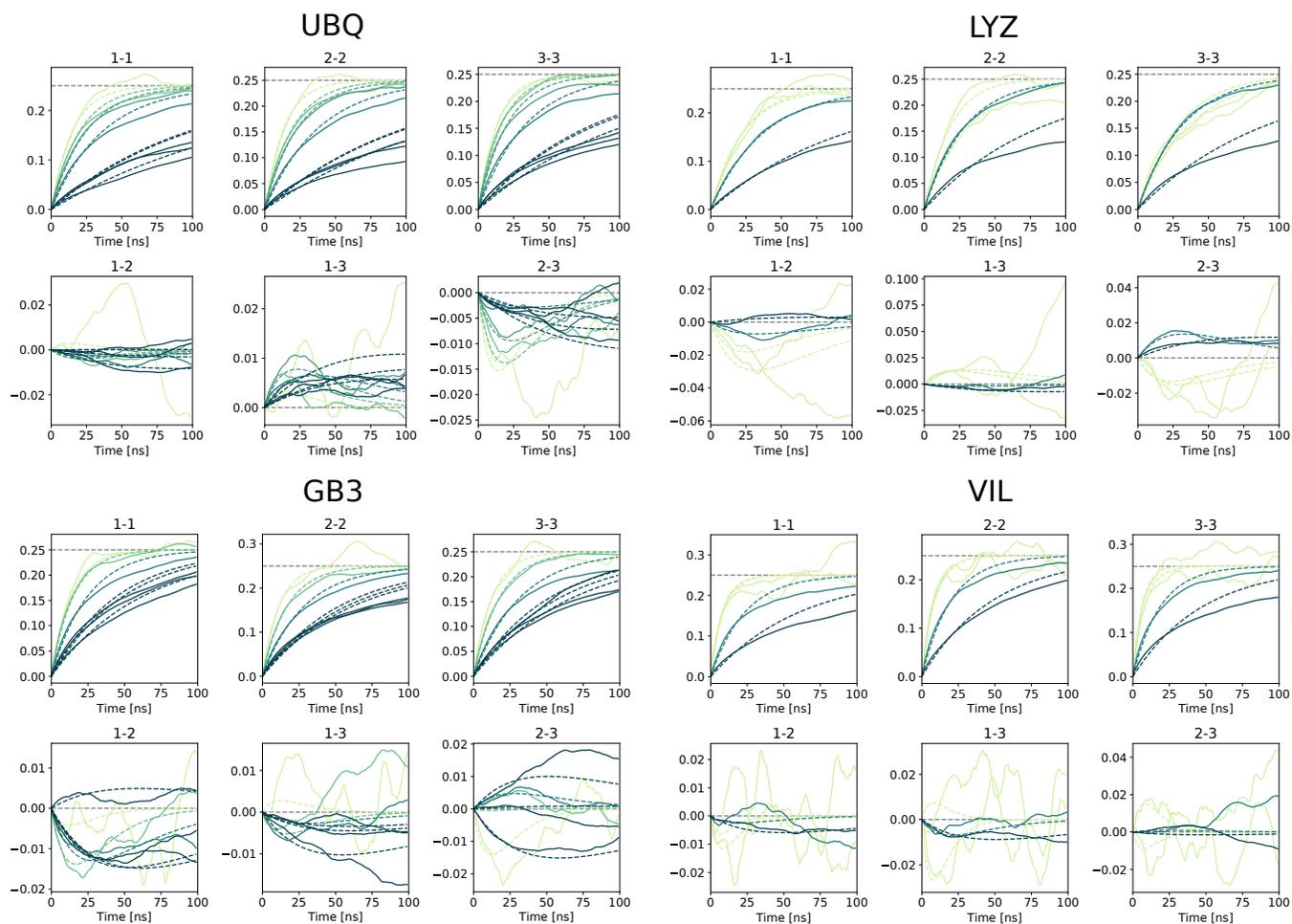
Fig. S4. RMSF of  $\alpha$ -carbon atoms to the average simulation structure in simulations of dilute and dense solutions. Darker colors indicate higher protein concentration.



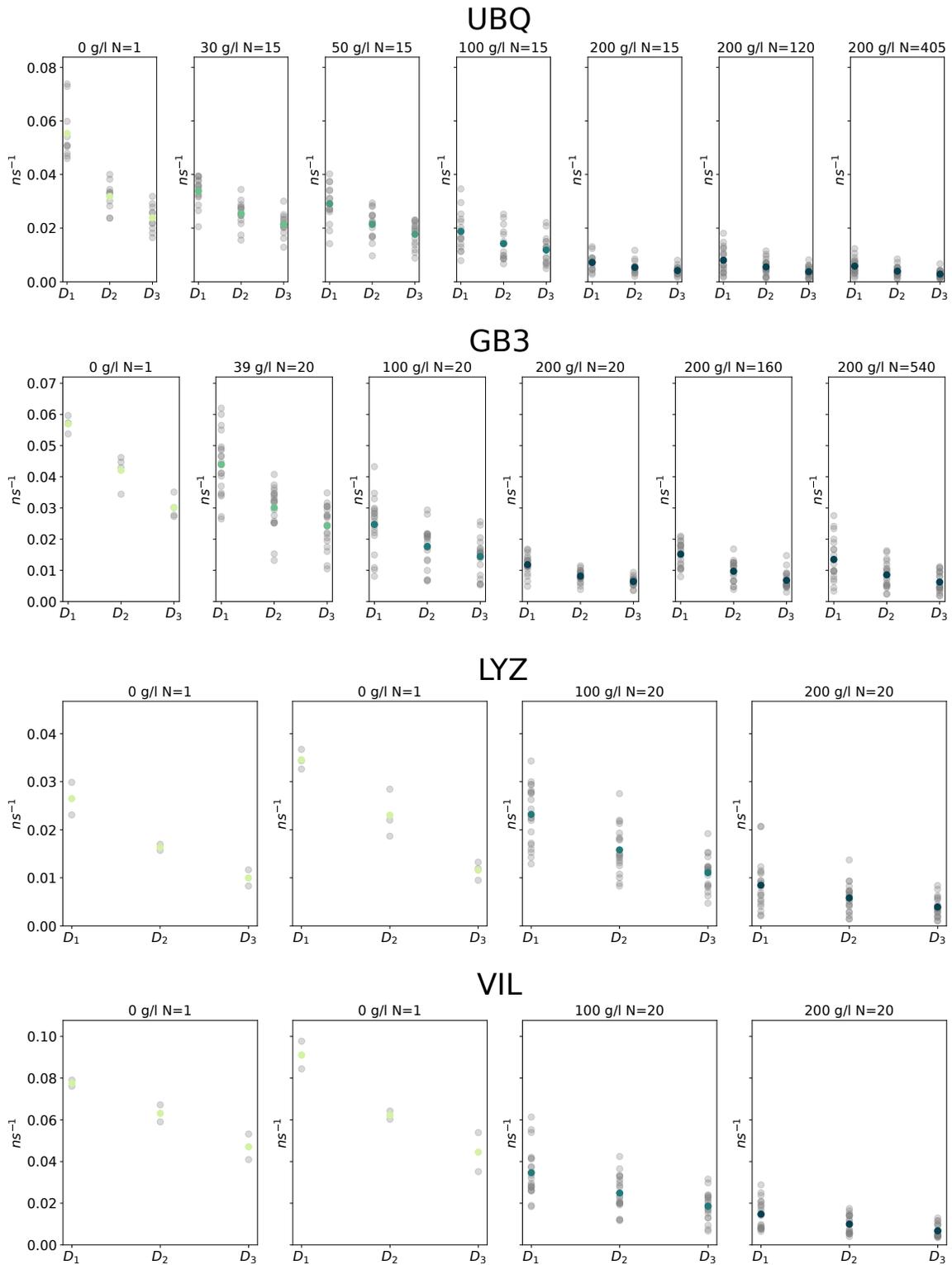
**Fig. S5.** Autocorrelation functions  $C_{ij}$  of the three off-diagonal pressure tensor elements  $P_{xy}$ ,  $P_{xz}$ ,  $P_{yz}$  and three combinations of the diagonal pressure tensor elements  $\frac{1}{2}(P_{xx} - P_{yy})$ ,  $\frac{1}{2}(P_{xx} - P_{zz})$  and  $\frac{1}{2}(P_{yy} - P_{zz})$  of TIP4P-D water, dense UBQ solutions and dense GB3 solutions. Darker colors indicate higher ion concentration (top) or protein concentration (bottom). The insets show  $C_{ij}$  at short times.



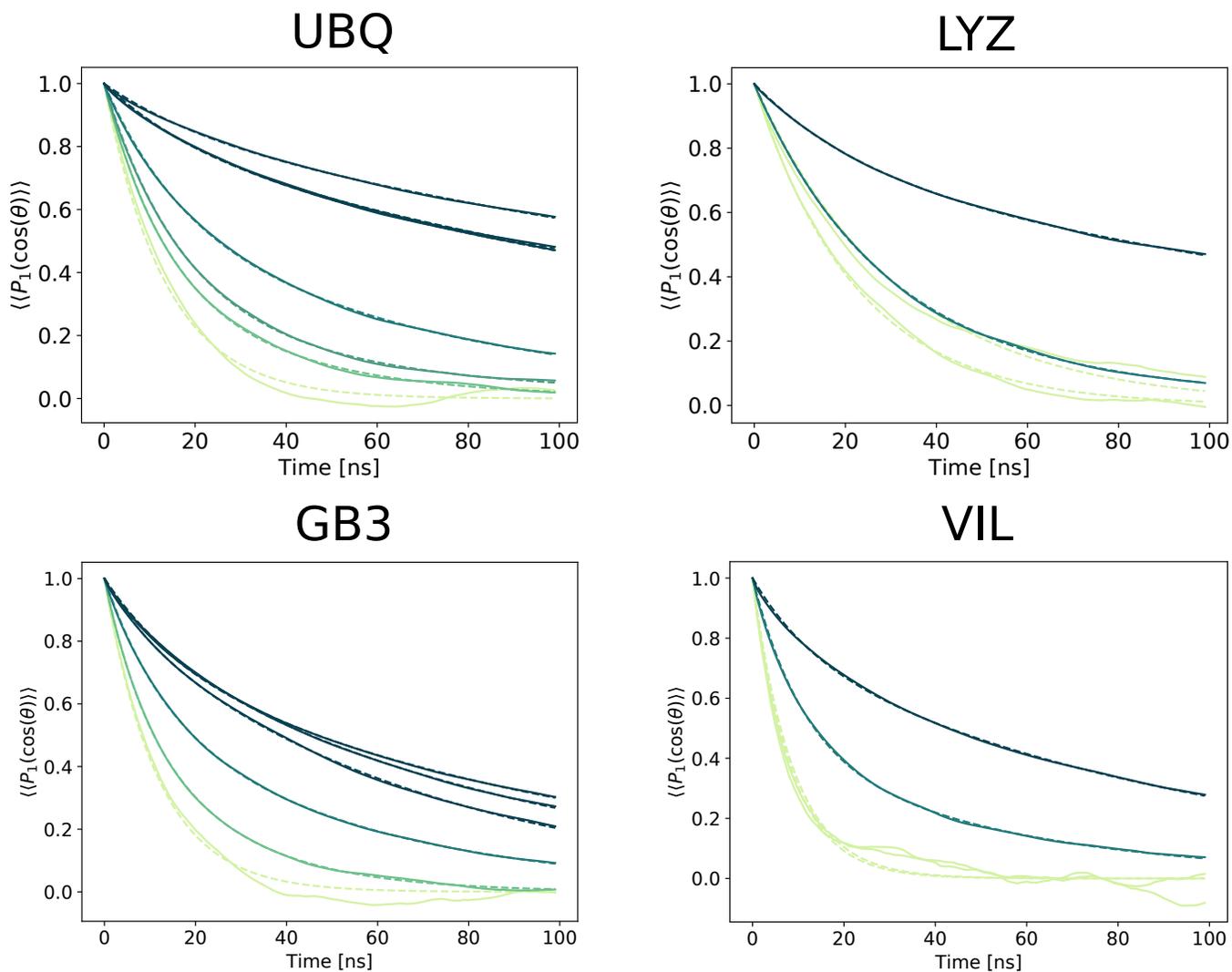
**Fig. S6.** MSD of UBQ, GB3, LYZ and VIL in dense protein solutions. Lines show the MSD averaged over all proteins in the solution. Respective standard errors of the mean are indicated as transparent surfaces. Black dashed lines indicate the linear fits in the region 10–30 ns (0–5 ns for dilute systems) used to extract the translational diffusion coefficients via Eq. S5. Darker colors indicate higher protein concentration.



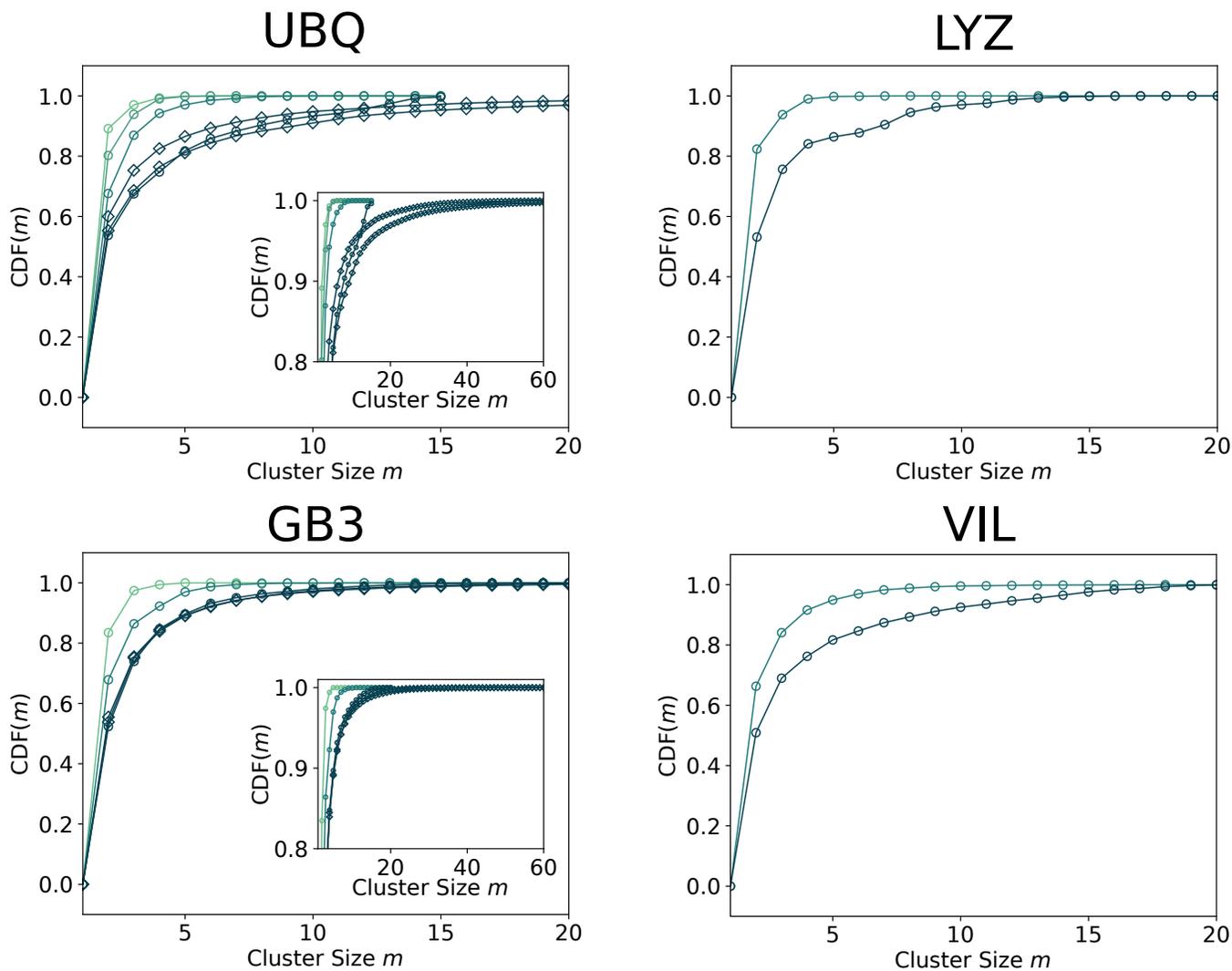
**Fig. S7.** Fits to the time-dependence of the six distinct quaternion covariances  $u_1 u_1 - u_2 u_3$  of UBQ, GB3, LYZ and VIL in dense protein solutions. Solid lines are the results of the MD simulations. Dashed lines are the fits corresponding to rigid-body rotational diffusion. Darker colors indicate higher protein concentration. Numbers  $i-j$  above the panels indicate the quaternion-covariance  $u_i u_j$ .



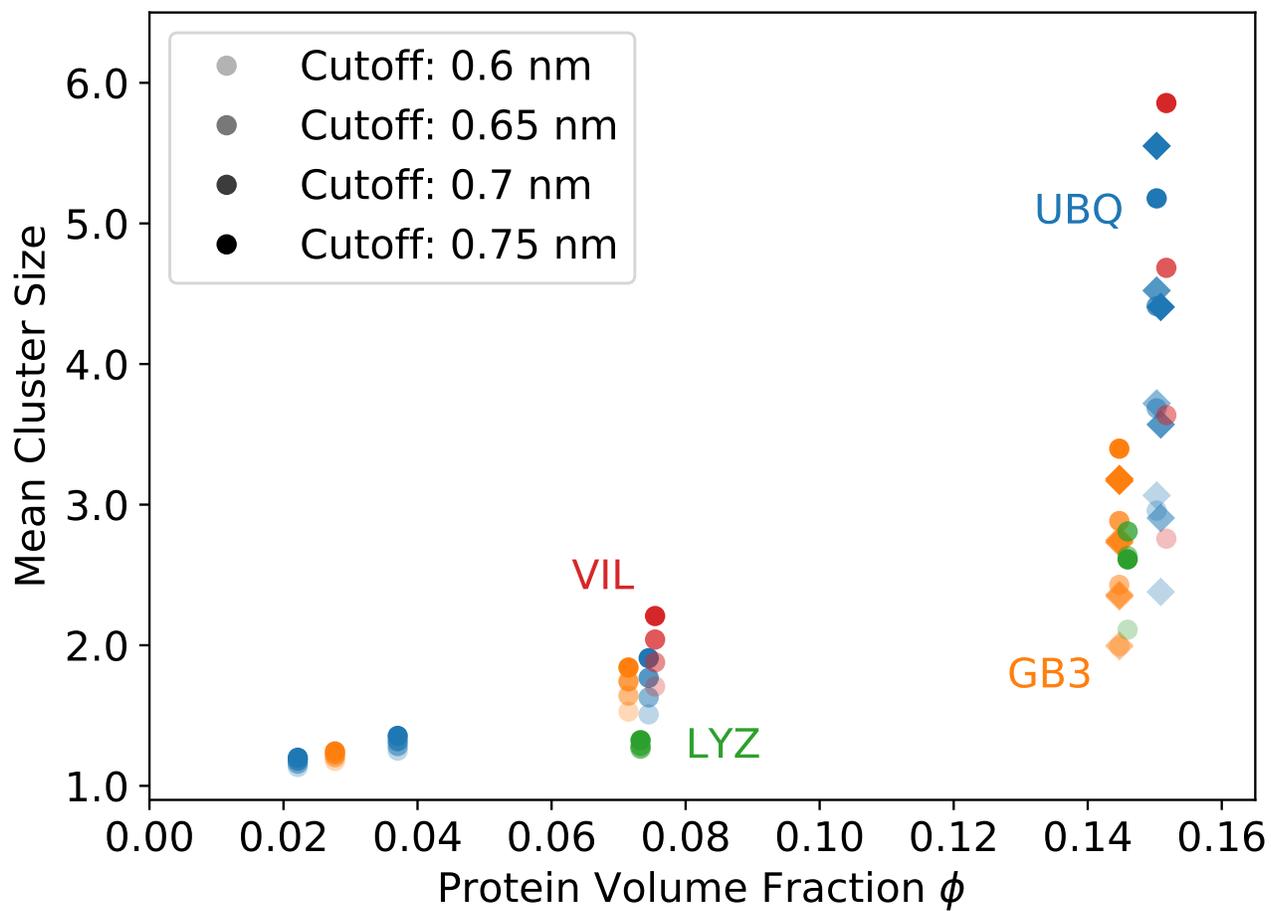
**Fig. S8.** Rotational diffusion coefficients  $D_1$ ,  $D_2$ , and  $D_3$  of dense protein solutions before correction for finite-size effects. Rotational diffusion coefficients of individual proteins are shown in gray. Mean rotational diffusion coefficients are shown in colors. Darker colors indicate higher protein concentration.



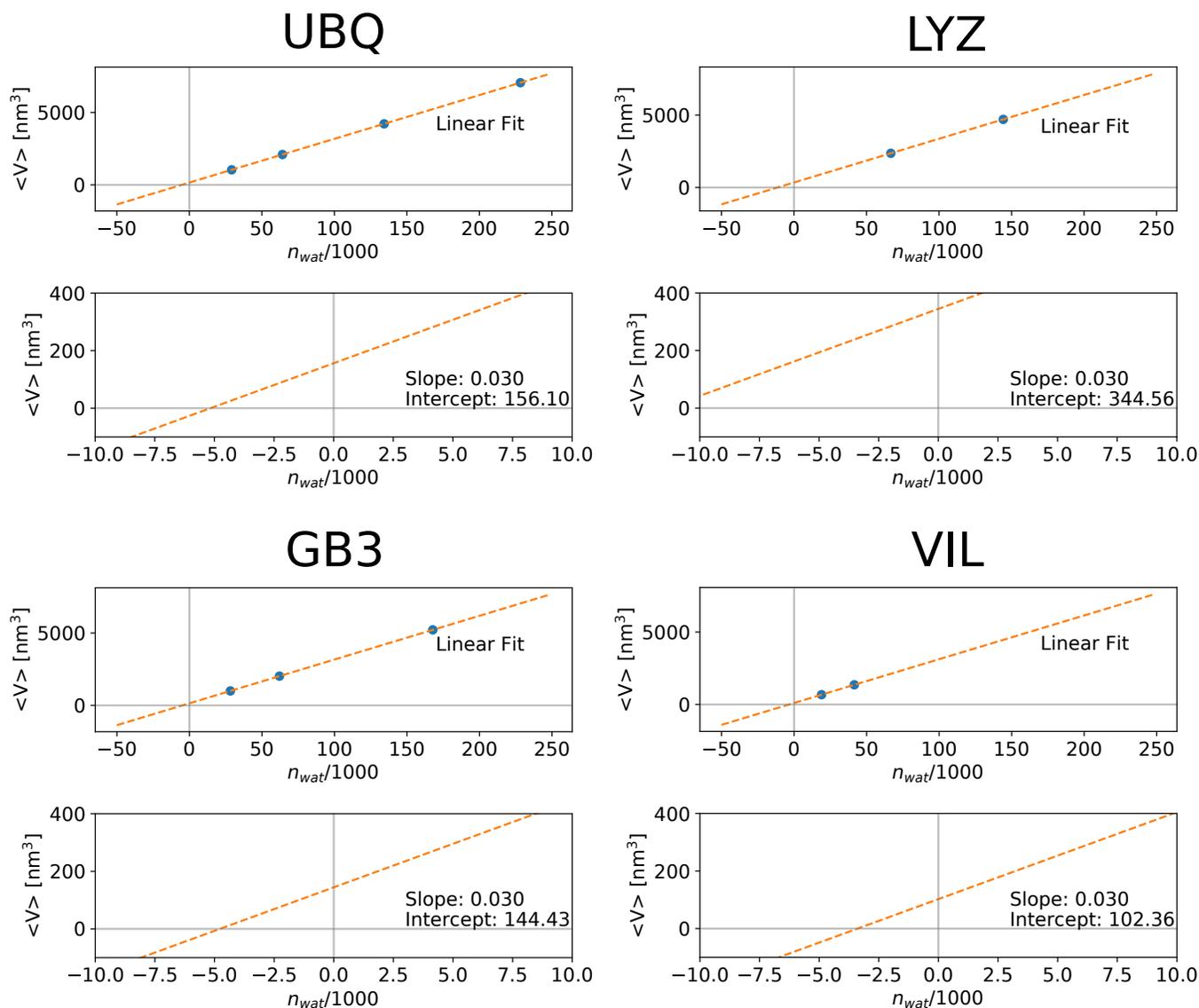
**Fig. S9.** Orientational correlation functions of UBQ, GB3, LYZ and VIL in concentrated solutions. Solid lines:  $\langle\langle P_1(\cos \theta(t)) \rangle\rangle$  calculated from  $u_0 u_0$ . Dashed lines: Bi-exponential fit to  $\langle\langle P_1(\cos \theta(t)) \rangle\rangle$  (single exponential fit for dilute system). Darker colors indicate higher protein concentration.



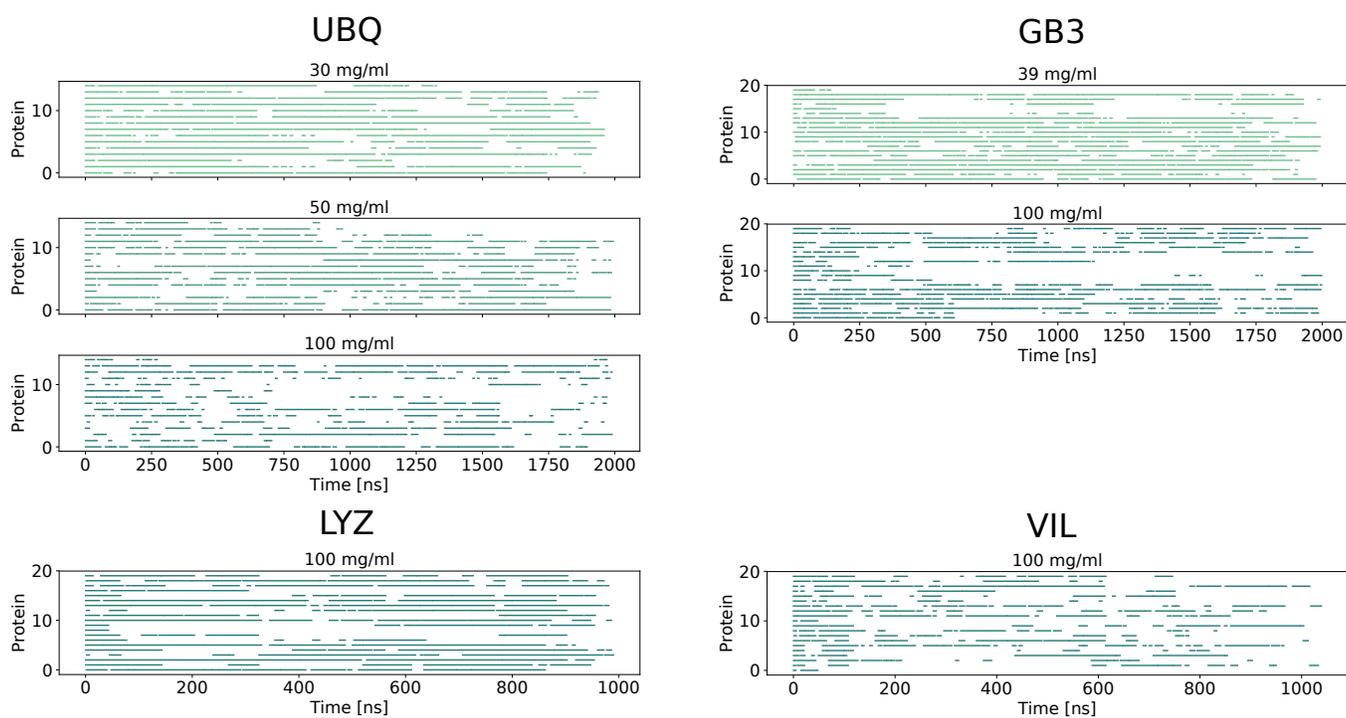
**Fig. S10.** Cumulative cluster size distributions of dense protein solutions of small simulation systems  $N \leq 20$  (circles) and large simulation systems  $N \geq 120$  (diamonds). The insets show the significant contribution of large cluster sizes to the cluster distribution of the large systems. Darker colors indicate higher protein concentration.



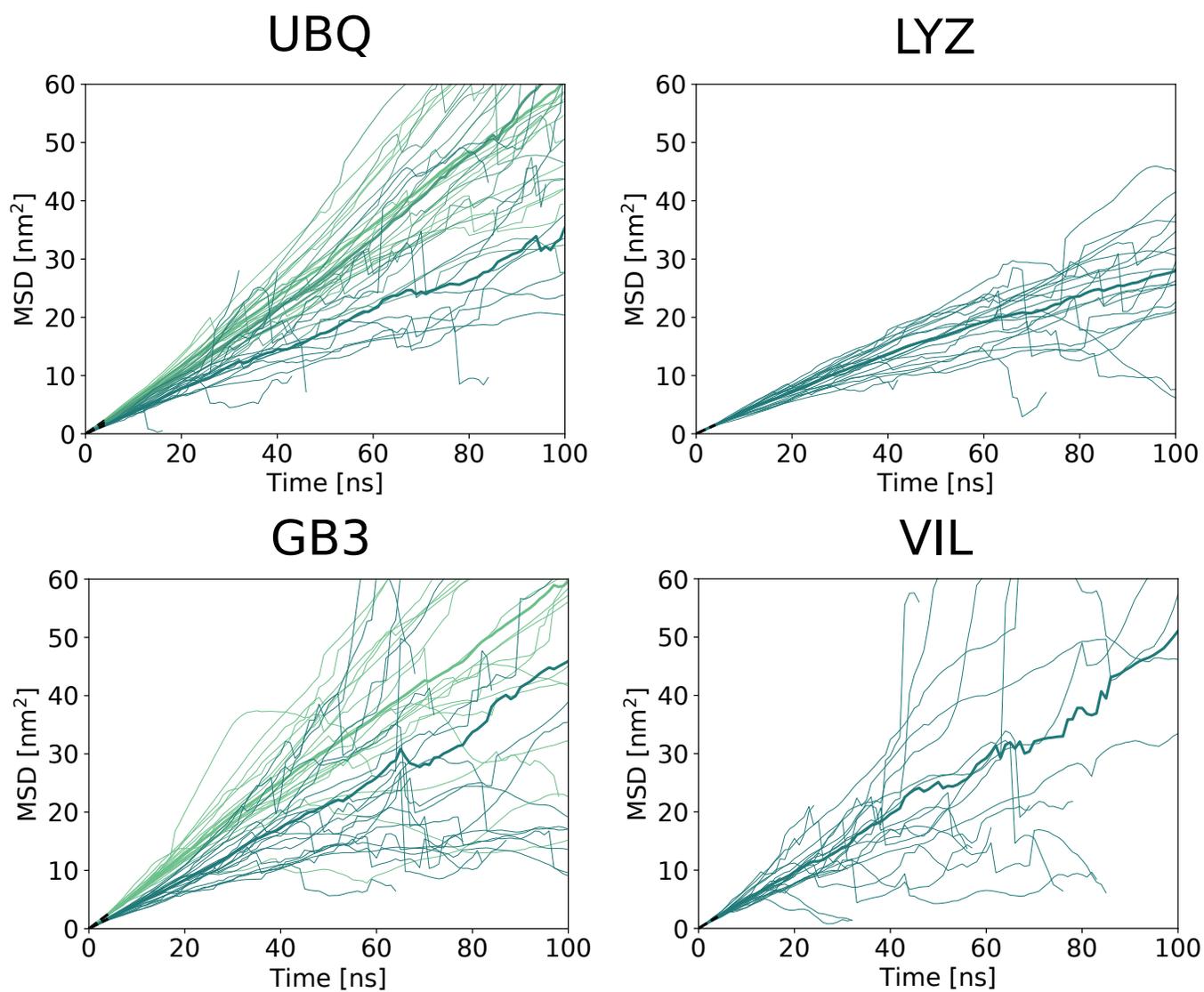
**Fig. S11.** Dependence of cluster size  $m$  on protein volume fraction  $\phi$  for  $\alpha$ -carbon distance cutoffs from 0.6 to 0.75 nm. Circles show results for small systems  $N \leq 20$ , diamonds show results for large systems  $N \geq 120$ .



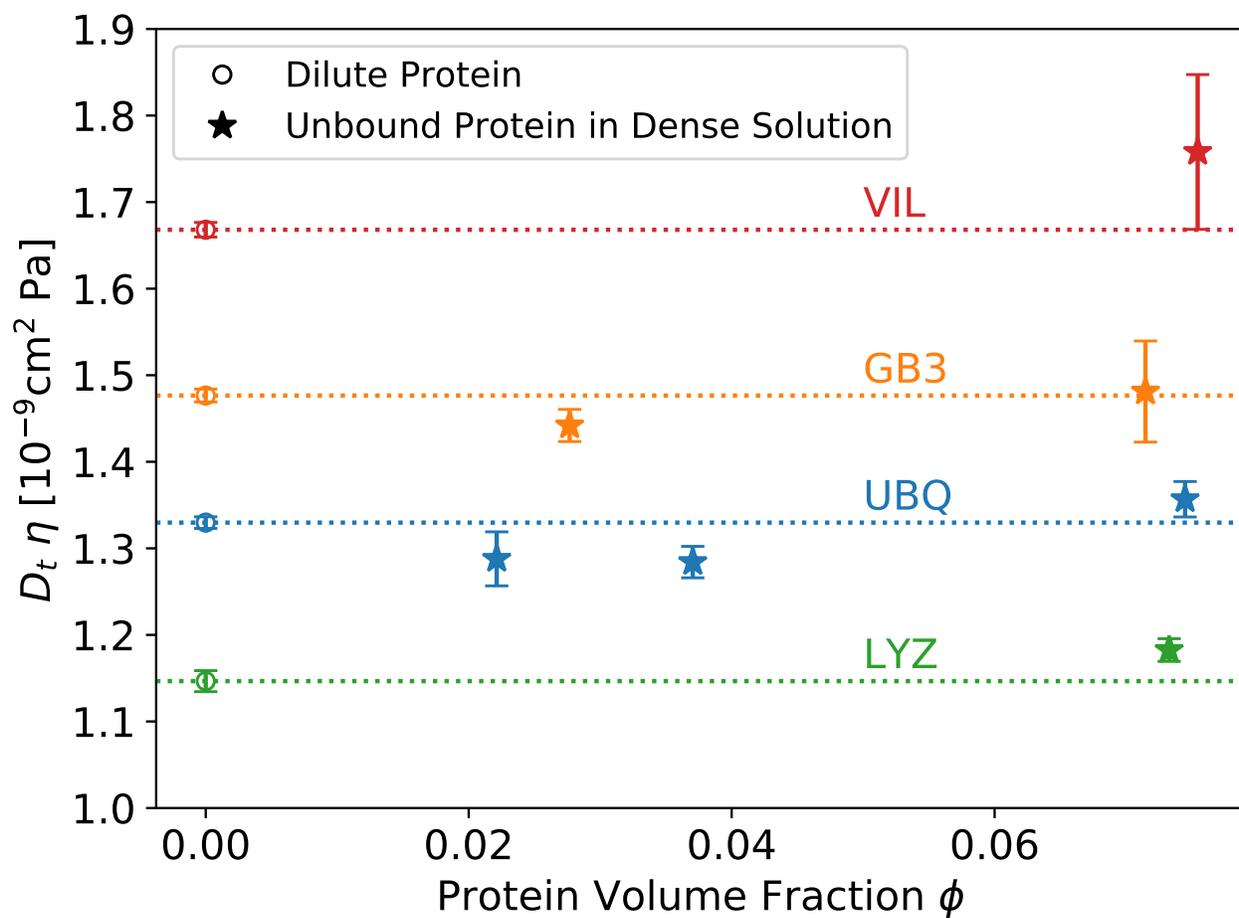
**Fig. S12.** Dependence of the mean volume of the simulation box on the number of water molecules  $n_{\text{wat}}$ . (Individual lower panels) Zoom-ins showing intersections of fit with axes.



**Fig. S13.** Times spent unbound. Simulation times in which the protein has not been part of a cluster for at least 5 ns are indicated by colored lines.



**Fig. S14.** MSD curves of unbound proteins in concentrated protein solutions. For each protein, MSDs were recorded for the times spent unbound (i.e. not in a cluster). Darker colors indicate higher protein concentration.



**Fig. S15.** Dependence of the product of viscosity and diffusion coefficient  $D_t^{\text{free}}$  of 'free' (unbound) protein (stars) and the product of viscosity and diffusion coefficient  $D_t$  of all proteins (circles) on protein volume fraction  $\phi$ . Dotted lines show the values of the dilute solutions.

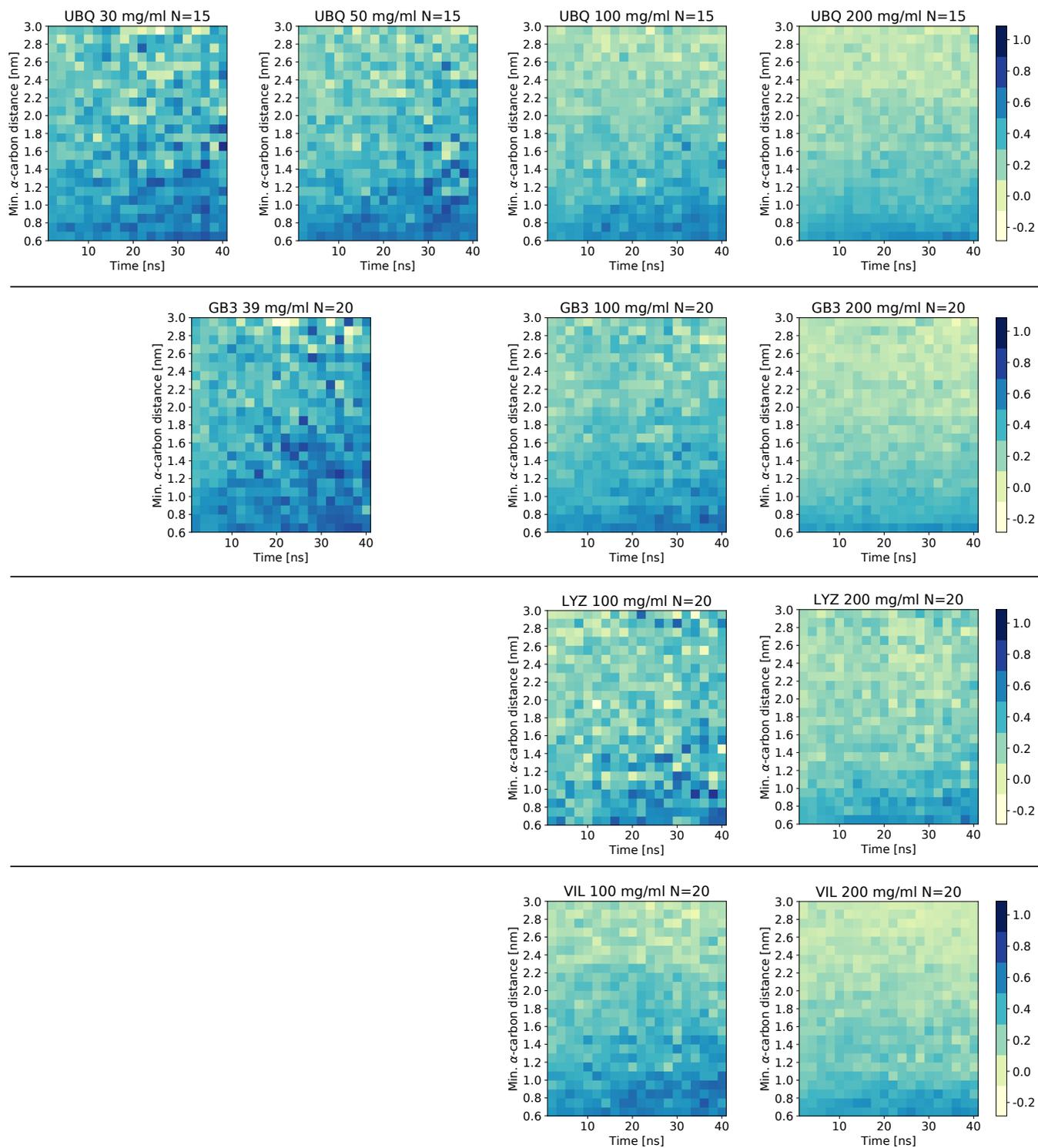
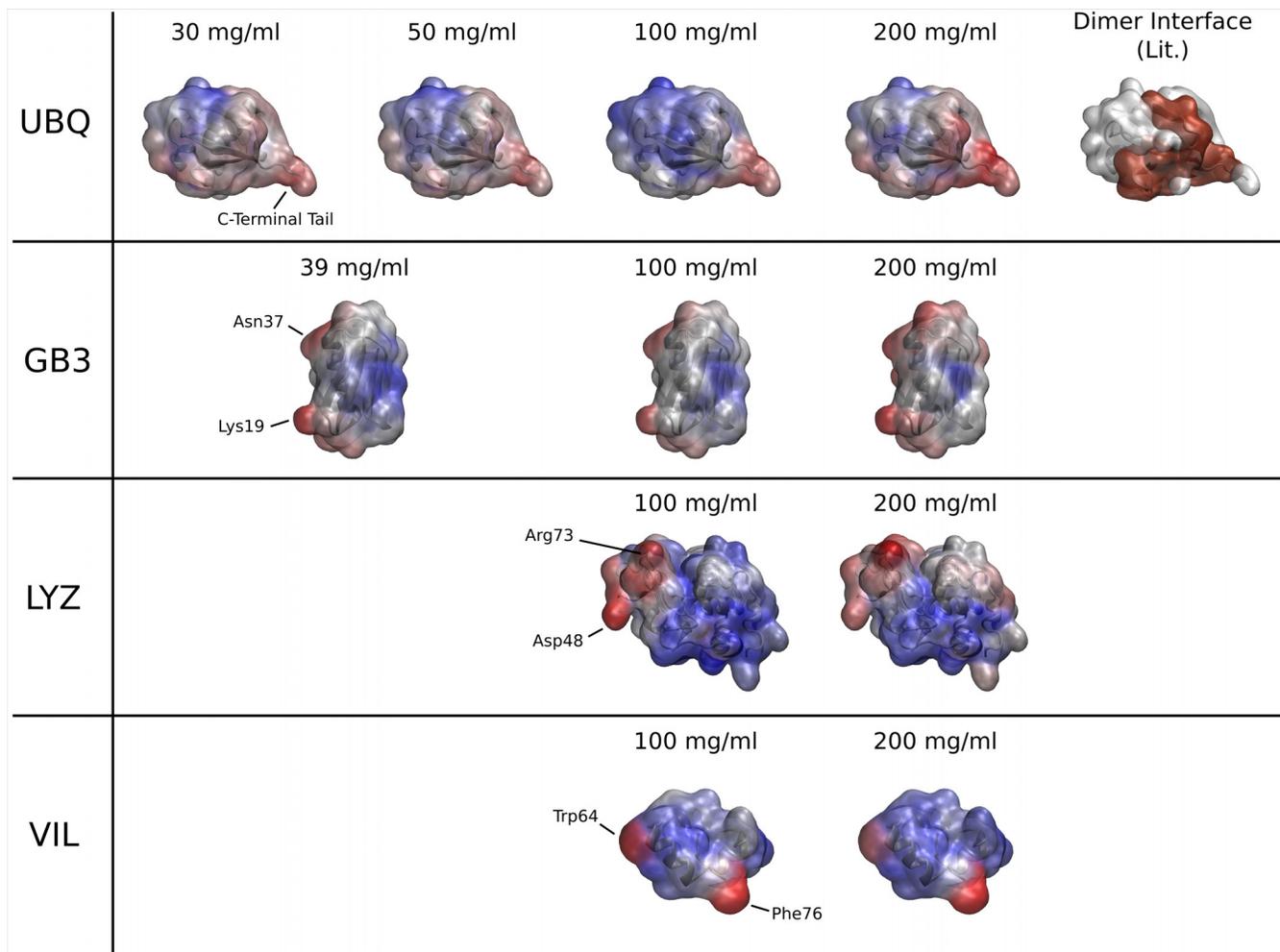
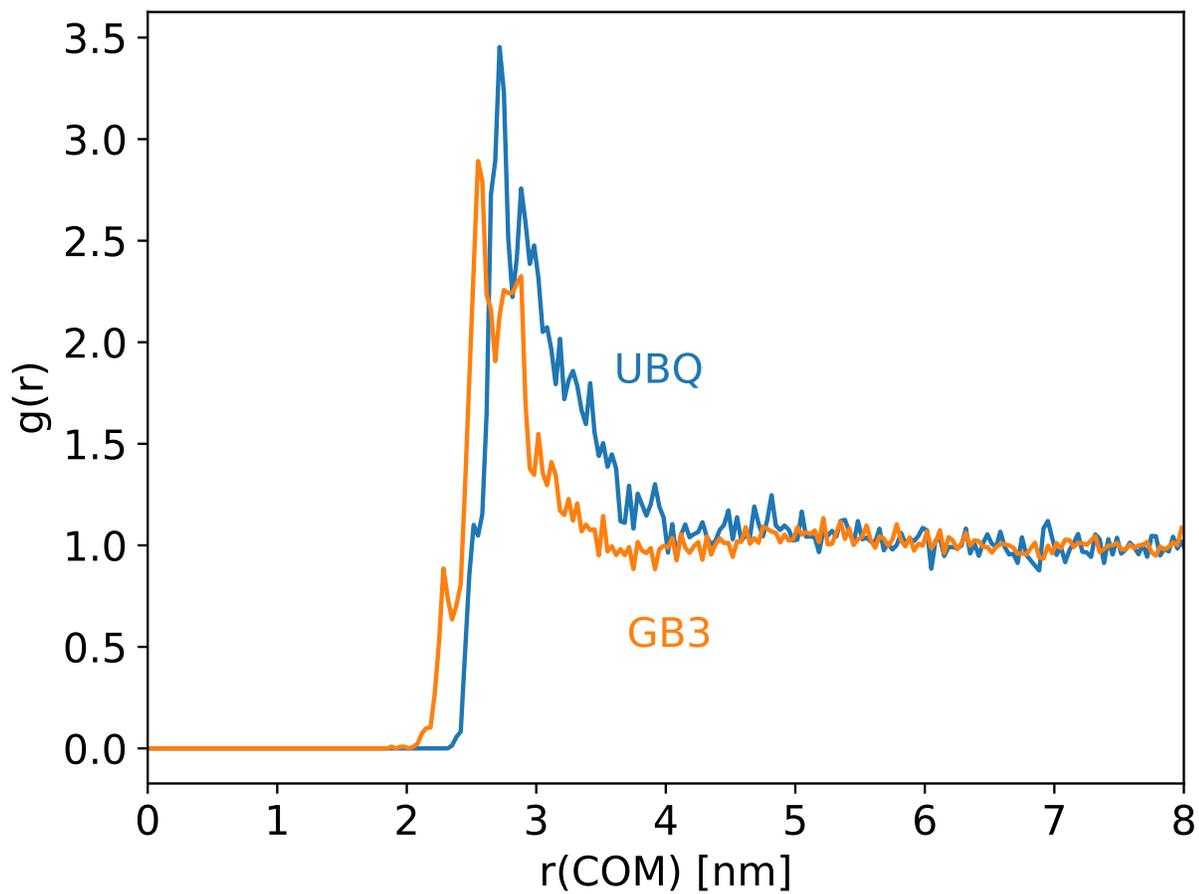


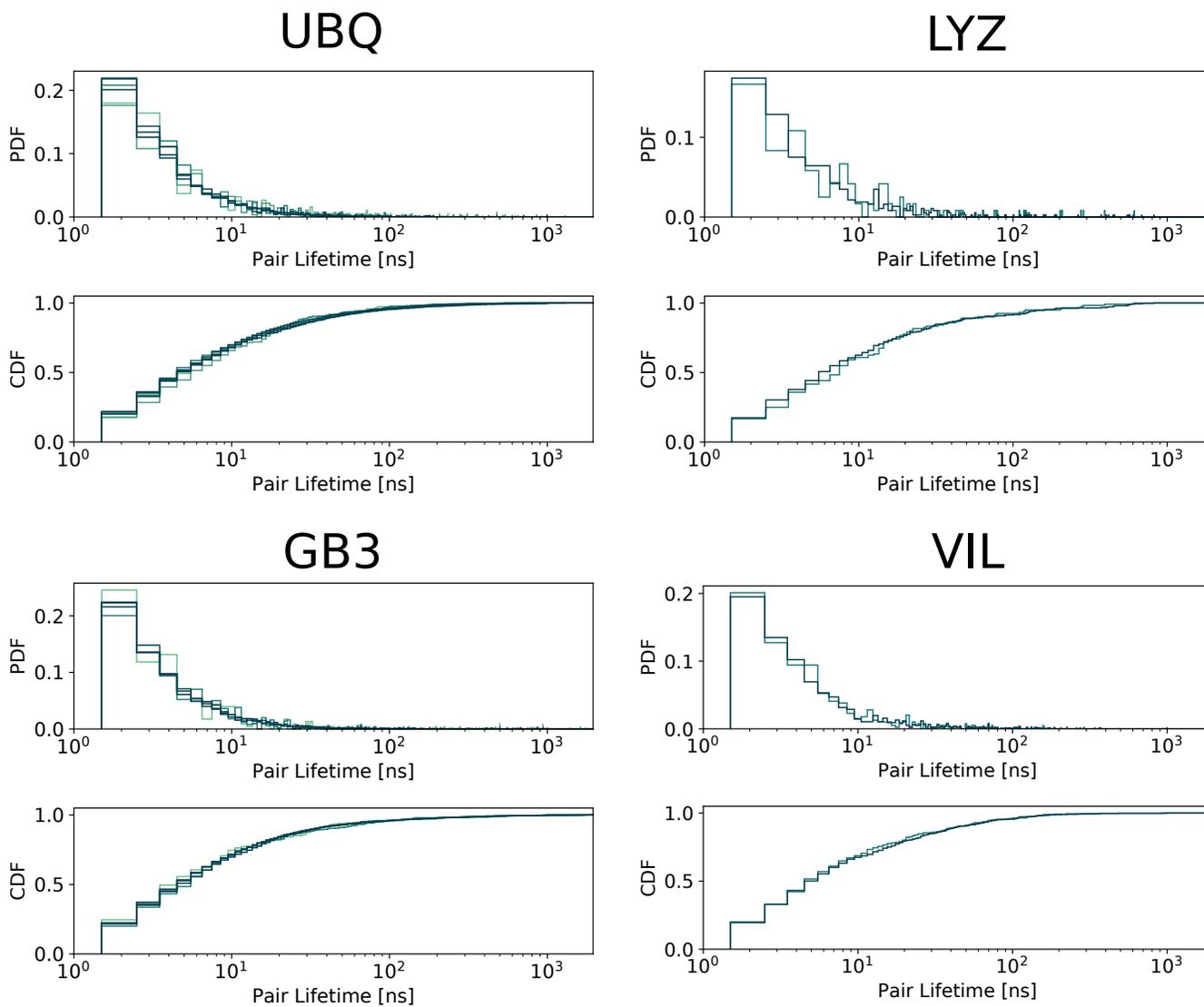
Fig. S16. Dependence of the displacement pair correlation on the minimum  $\alpha$ -carbon distance of the protein pairs (y-axis) and on the time delay (x-axis).



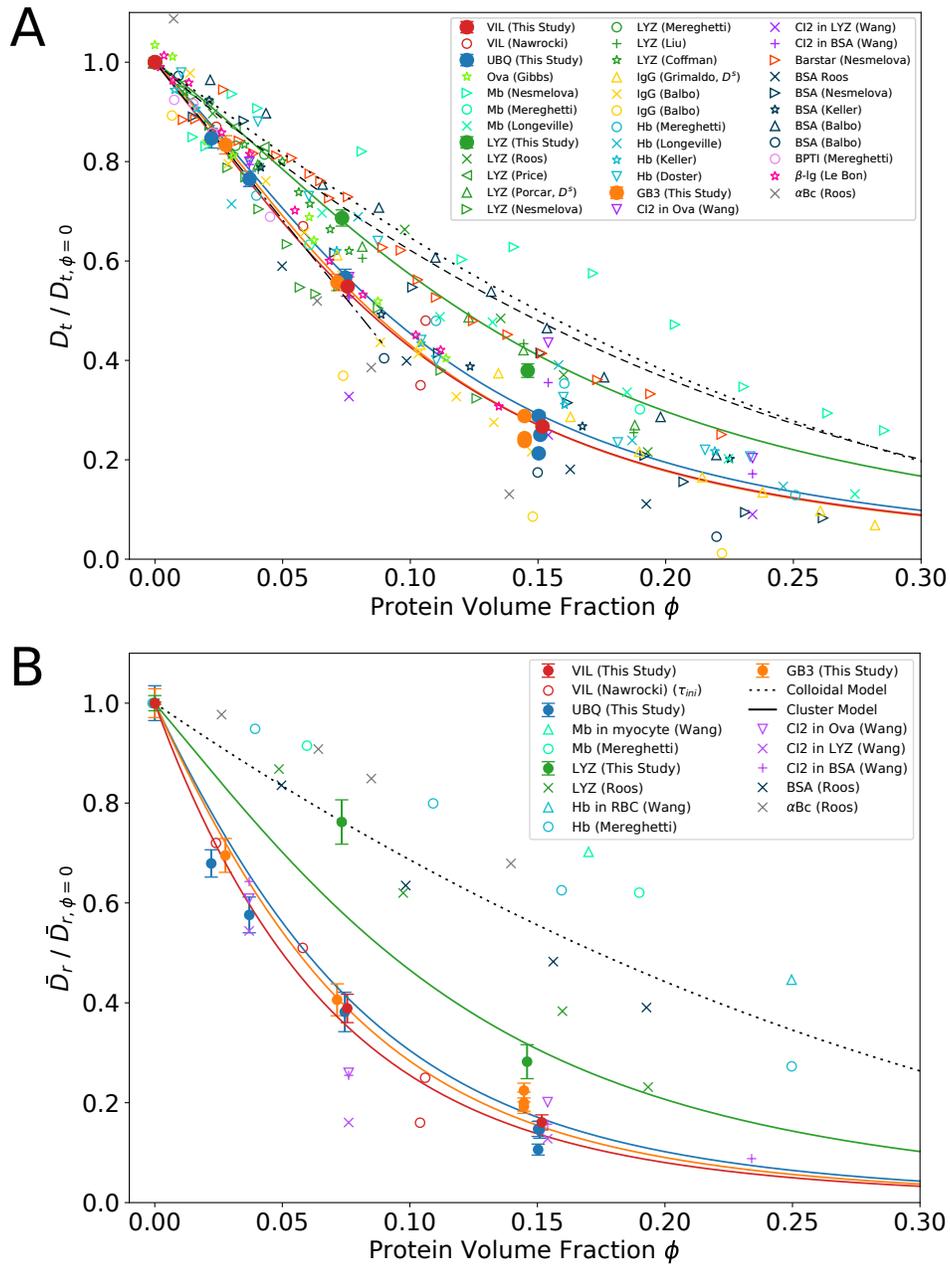
**Fig. S17.** Contributions to protein-protein interaction. The proteins are colored from no contribution (blue), intermediate contribution (white) to strong contribution (red). Residues that contribute most to interactions are labeled. The experimental dimer interface for UBQ is derived from (35).



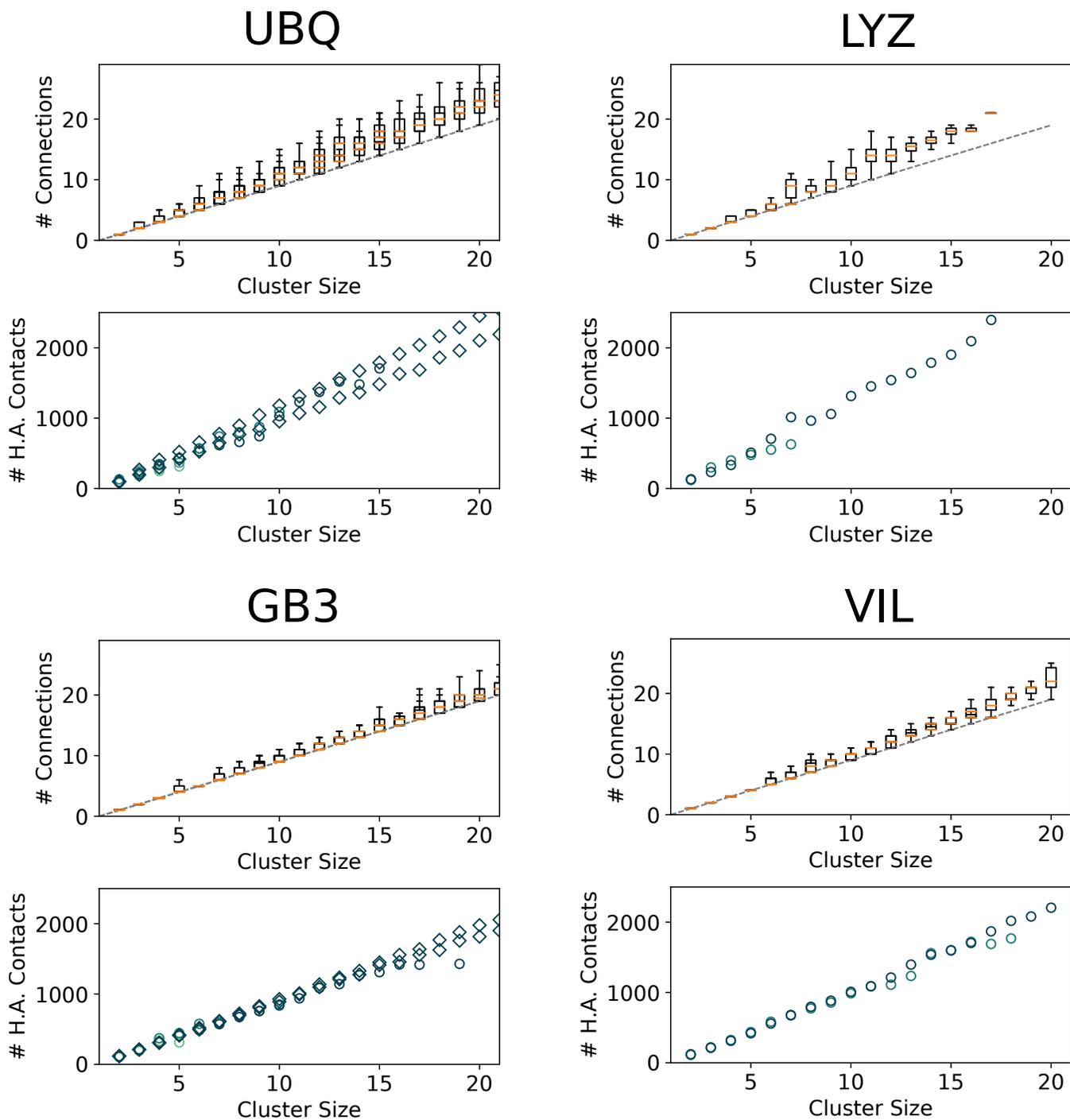
**Fig. S18.** Radial distribution functions  $g(r)$  of protein centers of mass of UBQ (30 mg/ml) and GB3 (39 mg/ml) solutions, which are, respectively, the lowest finite concentrations simulated. For the calculation of  $B_2$ ,  $g(r) - 1$  was integrated from  $r = 0$  to  $r = 4$  nm.



**Fig. S19.** Distribution of protein pair lifetimes. Darker colors indicate higher protein concentration.



**Fig. S20.** Concentration-dependent protein diffusion. (A) Dependence of the normalized translational diffusion coefficient  $D_t^{\text{red}}$  on protein volume fraction  $\phi$ . Filled circles show MD data from this study. Open circles show data from simulation studies (36–39). Other open symbols denote data from experimental studies (29, 38, 40–52). Mb: Myoglobin, Hb: Hemoglobin, Ova: Ovalbumin, "Mixed": CI2 in different dense protein solutions. All data refer to the long-time translational diffusion coefficient, with the exception of (49, 50), which refer to the short-time diffusion coefficient, indicated by  $D^s$ . The solid curves show the prediction of the dynamic cluster model  $D_{t,\text{clust}}(\phi)$  for UBQ, GB3, LYZ and VIL solutions, indicated by corresponding colors. The dashed and dotted curves show the slowdown of  $D_t^{\text{red}}$  predicted from colloid theory on non-interacting HS by van Blaaderen (21) and Tokuyama (23), respectively. The dash-dotted line shows a linear fit to  $D_t^{\text{red}}$  of UBQ and GB3 at  $\phi \leq 0.04$ . (B) Dependence of the normalized rotational diffusion coefficient  $\bar{D}_r^{\text{red}}$  on protein volume fraction  $\phi$ . Open circles show data from simulation studies (37, 39). Other open symbols denote data from experimental studies (48, 51, 53). "Cell": Hb or Mb in different cell types. The solid curves show the prediction of the dynamic cluster model  $\bar{D}_{r,\text{clust}}(\phi)$  for UBQ, GB3, LYZ and VIL solutions, indicated by corresponding colors. The dashed curve shows the slowdown of  $\bar{D}_r^{\text{red}}$  predicted from colloid theory on non-interacting HS (27).



**Fig. S21.** Dependence of the number of protein-protein connections (1 if  $C_\alpha$ -distance  $\leq 0.65$  nm, 0 otherwise) and number of heavy-atom pairs (distance  $\leq 0.50$  nm) on the cluster size. The grey line in the upper panels shows  $y = m - 1$ .

237 **Movie S1. MovieS1.avi: Atomistic MD simulation of 540 GB3 proteins in concentrated**  
238 **solution (200 mg/ml) at simulation time 0–500 ns. The fully flexible proteins are shown in**  
239 **surface representation and differentiated by color. For clarity, water and ions are omitted.**  
240 **Proteins that seem to appear and disappear traverse the periodic boundary in the direction**  
241 **of sight.**

## 242 **References**

- 243 1. Kim YC, Hummer G (2008) Coarse-grained models for simulations of multiprotein complexes: Appli-  
244 cation to ubiquitin binding. *J. Mol. Biol.* 375(5):1416–1433.
- 245 2. (2018) Maestro, Schrödinger, LLC, New York, NY.
- 246 3. Abraham MJ, et al. (2015) Gromacs: High performance molecular simulations through multi-level  
247 parallelism from laptops to supercomputers. *SoftwareX* 1-2:19–25.
- 248 4. Lindorff-Larsen K, et al. (2010) Improved side-chain torsion potentials for the amber ff99sb protein  
249 force field. *Proteins Struct. Funct. Bioinforma.* 78(8):1950–1958.
- 250 5. Best RB, Hummer G (2009) Optimized molecular dynamics force fields applied to the helix-coil  
251 transition of polypeptides. *J. Phys. Chem. B* 113(26):9004–9015.
- 252 6. Hornak V, et al. (2006) Comparison of multiple amber force fields and development of improved protein  
253 backbone parameters. *Proteins Struct. Funct. Bioinforma.* 65(3):712–725.
- 254 7. Best RB, De Sancho D, Mittal J (2012) Residue-specific  $\alpha$ -helix propensities from molecular simulation.  
255 *Biophys. J.* 102(6):1462–1467.
- 256 8. Piana S, Donchev AG, Robustelli P, Shaw DE (2015) Water dispersion interactions strongly influence  
257 simulated structural properties of disordered protein states. *J. Phys. Chem. B* 119(16):5113–5123.
- 258 9. Joung IS, Cheatham TE, Suk Joung I, Cheatham TE (2008) Determination of alkali and halide  
259 monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*  
260 112(30):9020–9041.
- 261 10. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J. Chem.*  
262 *Phys.* 126(1):14101.
- 263 11. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics  
264 with coupling to an external bath. *J. Chem. Phys.* 81(8):3684–3690.
- 265 12. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics  
266 method. *J. Appl. Phys.* 52(12):7182–7190.
- 267 13. Sridharan S, et al. (2019) Proteome-wide solubility and thermal stability profiling reveals distinct  
268 regulatory roles for atp. *Nat. Commun.* 10(1):1155.
- 269 14. Darden T, York D, Pedersen L (1993) Particle Mesh Ewald: An Nlog(N) method for Ewald sums in  
270 large systems. *J. Chem. Phys.* 98(12):10089–10092.
- 271 15. Theobald DL (2005) Rapid calculation of RMSDs using a quaternion-based characteristic polynomial.  
272 *Acta Crystallogr. A* 61(4):478–480.
- 273 16. Kneller GR (2010) Letter to the editor comment on “ fast determination of the optimal rotational  
274 matrix for macromolecular superpositions ”. *J. Comput. Chem.* 1561(7):31–32.
- 275 17. Gowers RJ, et al. (2016) MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics  
276 Simulations in *Proceedings 15th Python Science Conference*, eds. Benthall S, Rostrup S. pp. 98–105.
- 277 18. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O (2011) Mdanalysis: A toolkit for the analysis  
278 of molecular dynamics simulations. *J. Comput. Chem.* 32(10):2319–2327.
- 279 19. Linke M, Köfinger J, Hummer G (2018) Fully anisotropic rotational diffusion tensor from molecular  
280 dynamics simulations. *J. Phys. Chem. B* 122(21):5630–5639.
- 281 20. Hess B (2002) Determining the shear viscosity of model liquids from molecular dynamics simulations.  
282 *J. Chem. Phys.* 116(1):209.
- 283 21. 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(6):4591–4603.
22. Batchelor GK (1983) Diffusion in a dilute polydisperse system of interacting spheres. *J. Fluid Mech.* 131(-1):155.
23. Tokuyama M, Oppenheim I (1994) Dynamics of hard-sphere suspensions. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* 50(1):R16–R19.
24. Linke M, Köfinger J, Hummer G (2018) Rotational diffusion depends on box size in molecular dynamics simulations. *J. Phys. Chem. Lett.* 9(11):2874–2878.
25. Favro LD (1960) Theory of the rotational Brownian motion of a free rigid body. *Phys. Rev.* 119(1):53–62.
26. Woessner DE (1998) Nuclear spin relaxation in ellipsoids undergoing rotational Brownian motion. *J. Chem. Phys.* 647(1962):647–654.
27. 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(7):1714.
28. Ortega A, et al. (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* 101(4):892–898.
29. Liu Y, et al. (2011) Lysozyme protein solution with an intermediate range order structure. *J. Phys. Chem. B* 115(22):7238–7247.
30. Baxter RJ (1968) Percus-Yevick equation for hard spheres with surface adhesion. *J. Chem. Phys.* 49(6):2770–2774.
31. Cichocki B, Felderhof BU (1990) Diffusion coefficients and effective viscosity of suspensions of sticky hard spheres with hydrodynamic interactions. *J. Chem. Phys.* 93(6):4427–4432.
32. Ando T, Skolnick J (2010) Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. *Proc. Natl. Acad. Sci.* 107(43):18457–18462.
33. Petrov D, Karplus M (2014) Are current atomistic force fields accurate enough to study proteins in crowded environments. *PLoS Comput. Biol.* 10(5):1–11.
34. Yu I, et al. (2016) Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *Elife* 5:e19274.
35. Liu Z, et al. (2012) Noncovalent dimerization of ubiquitin. *Angew. Chemie Int. Ed.* 51(2):469–472.
36. Mereghetti P, Gabdoulline RR, Wade RC (2010) Brownian dynamics simulation of protein solutions: Structural and dynamical properties. *Biophys. J.* 99(11):3782–3791.
37. 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(29):8523–8533.
38. Balbo J, Mereghetti P, Herten DP, Wade RC (2013) The shape of protein crowders is a major determinant of protein diffusion. *Biophys. J.* 104(7):1576–1584.
39. 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(49):11072–11084.
40. Keller KH, Canales ER, Yam S (1971) Tracer and mutual diffusion coefficients of proteins. *J. Phys. Chem.* 75(3):379–387.
41. Gibbs SJ, Chu AS, Lightfoot EN, Root TW (1991) Ovalbumin diffusion at low ionic strength. *J. Phys. Chem.* 95(1):467–471.
42. 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(12):2218–2223.
43. 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(46):10294–10299.
44. Price WS, Tsuchiya F, Arata Y (1999) Lysozyme aggregation and solution properties studied using PGSE NMR diffusion measurements. *J. Am. Chem. Soc.* 121(49):11503–11512.
45. Nesmelova IV, Skirda VD, Fedotov VD (2002) Generalized concentration dependence of globular protein self-diffusion coefficients in aqueous solutions. *Biopolymers* 63(2):132–140.

- 333 46. Longeville S, Doster W, Diehl M, Gähler R, Petry W (2003) Neutron resonance spin echo: Oxygen  
334 transport in crowded protein solutions in *Neutron Spin Echo Spectrosc.* pp. 325–335.
- 335 47. Doster W, Longeville S (2007) Microscopic diffusion and hydrodynamic interactions of hemoglobin in  
336 red blood cells. *Biophys. J.* 93(4):1360–1368.
- 337 48. Wang Y, Li C, Pielak GJ (2010) Effects of proteins on protein diffusion. *J. Am. Chem. Soc.* 132(16):9392–  
338 9397.
- 339 49. Porcar L, et al. (2010) Formation of the dynamic clusters in concentrated lysozyme protein solutions.  
340 *J. Phys. Chem. Lett.* 1(1):126–129.
- 341 50. Grimaldo M, Roosen-Runge F, Zhang F, Seydel T, Schreiber F (2014) Diffusion and dynamics of  
342  $\gamma$ -globulin in crowded aqueous solutions. *J. Phys. Chem. B* 118(25):7203–7209.
- 343 51. Roos M, et al. (2016) Coupling and decoupling of rotational and translational diffusion of proteins  
344 under crowding conditions. *J. Am. Chem. Soc.* 138(32):10365–10372.
- 345 52. Longeville S, Stingaciu LR (2017) Hemoglobin diffusion and the dynamics of oxygen capture by red  
346 blood cells. *Sci. Rep.* 7(1):1–10.
- 347 53. Wang D, Kreutzer U, Chung Y, Jue T (1997) Myoglobin and hemoglobin rotational diffusion in the  
348 cell. *Biophys. J.* 73(5):2764–2770.