



HAL
open science

Optimized OPEP Force Field for Simulation of Crowded Protein Solutions

Stepan Timr, Simone Melchionna, Philippe Derreumaux, Fabio Sterpone

► **To cite this version:**

Stepan Timr, Simone Melchionna, Philippe Derreumaux, Fabio Sterpone. Optimized OPEP Force Field for Simulation of Crowded Protein Solutions. *Journal of Physical Chemistry B*, 2023, 127 (16), pp.3616-3623. 10.1021/acs.jpcc.3c00253 . hal-04234930

HAL Id: hal-04234930

<https://hal.science/hal-04234930>

Submitted on 10 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimized OPEP Force Field for Simulation of Crowded Protein Solutions

Stepan Timr,* Simone Melchionna, Philippe Derreumaux, and Fabio Sterpone*



Cite This: *J. Phys. Chem. B* 2023, 127, 3616–3623



Read Online

ACCESS |



Metrics & More

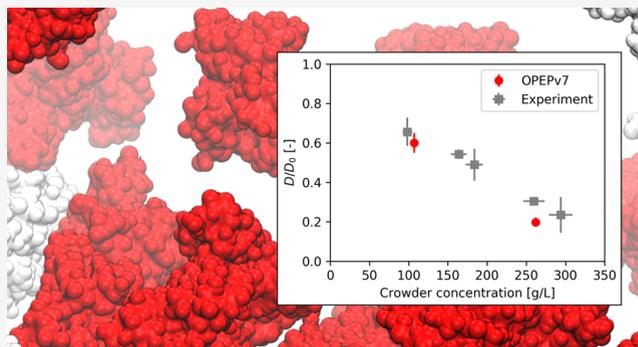


Article Recommendations



Supporting Information

ABSTRACT: Macromolecular crowding has profound effects on the mobility of proteins, with strong implications on the rates of intracellular processes. To describe the dynamics of crowded environments, detailed molecular models are needed, capturing the structures and interactions arising in the crowded system. In this work, we present OPEPv7, which is a coarse-grained force field at amino-acid resolution, suited for rigid-body simulations of the structure and dynamics of crowded solutions formed by globular proteins. Using the OPEP protein model as a starting point, we have refined the intermolecular interactions to match the experimentally observed dynamical slowdown caused by crowding. The resulting force field successfully reproduces the diffusion slowdown in homogeneous and heterogeneous protein solutions at different crowding conditions. Coupled with the lattice Boltzmann technique, it allows the study of dynamical phenomena in protein assemblies and opens the way for the *in silico* rheology of protein solutions.



INTRODUCTION

In living organisms as well as in numerous engineering applications, proteins are immersed in a crowded environment, which can modulate their structural and dynamical properties.^{1–3} In particular, crowding can strongly impact protein diffusion,⁴ with profound effects on the rates of metabolic reactions and mechanisms of intracellular signaling.⁵

By providing detailed microscopic insights, molecular dynamics (MD) simulations have contributed significantly to our understanding of protein function.⁶ Due to their high computational costs, all-atomistic MD simulations have conventionally focused on the behavior of an isolated protein or that of a few of such proteins under conditions approaching those in a dilute solution (that is, in a situation which does not accurately reflect the actual working environment of most proteins). Despite recent progress,^{7,8} the wide range of temporal and spatial scales involved in crowding still poses a challenge to all-atom MD simulations. On the other hand, coarse-grained simulations have been used to explore large model systems at longer time scales using different levels of molecular resolution and theory.^{9,10} Interestingly, from these works it emerged that fine-tuning of the intermolecular interactions was needed in order to reproduce the experimentally observed transport properties of the crowded solutions. It now seems clear that the engagement in local clusters controls the protein mobility in crowded protein solutions.^{8,11}

Combining a detailed description of the protein backbone with a coarse-grained description of amino-acid side chains,

Optimized Potential for Efficient Protein structure Prediction (OPEP) is an implicit-solvent protein force field designed to characterize processes involving the folding and unfolding of peptides and proteins.¹² To capture the effects of solvent-mediated hydrodynamic interactions on protein motions and structures, OPEP was coupled to a lattice Boltzmann description of solvent kinetics.¹³ This hybrid particle–lattice approach has proven successful in numerous applications, including simulating the kinetics of amyloid aggregation and investigating protein mechanical unfolding under shear.^{14–19}

In this paper, we present OPEPv7, which is a streamlined and optimized version of the OPEP force field targeted at large-scale simulation of crowded protein solutions. The aim is to provide a model that would adequately describe protein mobility in crowded mixtures, including the characterization of protein–protein contacts, clusters, and their lifetimes. To reduce computational costs and to expand accessible time scales, we simplify the description of the backbone. Moreover, we tweak the side chain–side chain interactions in order to obtain a good agreement with experimental measurements of protein diffusion in model protein systems. The resulting force

Received: January 12, 2023

Revised: April 5, 2023

Published: April 18, 2023



field allows one to investigate protein diffusion in crowded protein solutions as well as inquire with regard to the structure of such crowded systems.

METHODS

Functional Form of the Force Field. The standard version of the OPEP force field comes with an atomistic description of the backbone, comprising bond, angle, and torsion potentials, as well as specific hydrogen bonding terms.¹² While such a detailed representation of the backbone allows simulating protein conformational transitions and folding, it is not necessary for the description of stable proteins that encounter each other in a crowded solution, a process primarily governed by intermolecular interactions of side chains on protein surfaces. Therefore, as a first approximation, the protein can be modeled by an elastic network of tunable flexibility, and using a reduced resolution of the backbone.

OPEPv7 comes with a simplified description of the backbone, consisting exclusively of $C\alpha$ beads with radii increased by 50%, in comparison to the original force field to recover the original volume of the backbone. The protein conformation is maintained with the help of an elastic network interconnecting backbone- and side-chain beads within a cutoff, typically set to 6–8 Å. This simplified description allows an order-of-magnitude increase in the simulation time step while reducing significantly (about 3-fold) the number of beads in a simulation. The explicit presence of $C\alpha$ beads eases the task of back-mapping to the atomistic resolution when a multiscale strategy is deployed.^{14,20}

Both in standard OPEP and OPEPv7, the side chain–side chain interactions are described by short-range pairwise interaction potentials. Depending on whether they are purely repulsive or whether they contain an attractive part, these pairwise interaction potentials assume the form of a $1/r^8$ potential or a more-complex expression combining an e^{-2r/r^6} repulsion with an attractive term described by the tanh function (see the [Supporting Methods](#) section in the Supporting Information for more details). The short-range nature of these potentials reflects the screening by physiological ion concentrations in cytoplasm-like environments. The strength of the repulsion or attraction is controlled by the ϵ_{ij} parameter, and the position of the minimum of the attractive potential is governed by the distance parameter r_{ij}^0 (see the [Supporting Information](#)). As a tunable parameter, an empirical scaling factor is uniformly applied to all ϵ_{ij} in OPEP. Values between 1.6 and 1.8 were shown to yield a good agreement with experimental data on protein folding energetics.^{12,21}

Derivation of OPEPv7. As a starting point for further optimization, we used version 4 of the OPEP¹² with the empirical scaling factor set to 1.0. We adjusted the interaction potentials of oppositely charged residue pairs by tweaking the ϵ parameters ($\epsilon_{\text{ARG-ASP}} = 2.43$ kcal/mol, $\epsilon_{\text{ARG-GLU}} = 2.51$ kcal/mol, $\epsilon_{\text{LYS-ASP}} = 1.86$ kcal/mol, $\epsilon_{\text{LYS-GLU}} = 1.92$ kcal/mol). As a consequence, ARG-involving ion pairs became stronger relative to LYS-containing ion pairs, qualitatively reproducing a tendency observed in potential-of-mean-force profiles derived from all-atom simulations.²¹ All other side chain–side chain nonbonded interactions were kept the same as in the original OPEPv4 force field except for the CYS–CYS pair, where we replaced the original high ϵ parameter, implicitly describing disulfide bridging, with the lower value $\epsilon_{\text{CYS-CYS}} = 1.43304$ kcal/mol. This value was transferred from the CYS–MET pair,

that is, a pair involving another sulfur-containing residue, yet without the capacity to form disulfide bridges. We note that given the low abundance of CYS residues on protein surfaces,²² the overall dynamics of the crowded protein solution is likely to exhibit little sensitivity to the precise value of the CYS–CYS interaction parameter if the interaction strength is not excessively high.

Model of an Intrinsically Disordered Protein. To test the applicability of OPEPv7 to flexible, intrinsically disordered proteins (IDP), we created an IDP version of OPEPv7 in which each amino acid is represented by a unique bead. The approach was inspired by previous models introduced to describe IDP.²³ For glycine, alanine, and proline, the interacting potential terms are those associated with their $C\alpha$ beads; for the other amino acids, we used the side-chain potential interaction terms. The individual beads were connected by harmonic potentials with a force constant $K = 2.39$ kcal/mol and an equilibrium distance $b_0 = 3.8$ Å, forming a flexible polymer chain.

Fits with a Simplified Functional Form. The functional form of the OPEPv7 force field may not be available in all MD simulation packages. Inspired by the recent development of a highly scalable software for the MD simulations of coarse-grained systems,²⁴ we derived a simplified version of the side chain–side chain interaction potential using the standard 12–6 Lennard-Jones (LJ) form for attractive pairs while maintaining the $1/r^8$ OPEPv7 term for purely repulsive pairs. To obtain the parameters of the LJ potential, we performed a least-squares fit of the LJ radial distribution function to the original OPEPv7 radial distribution function in the 3–10 Å interval for each pair, while fixing the position of the minimum of the interaction potential to the OPEPv7 value.

LBMD Simulations. Lattice Boltzmann molecular dynamics (LBMD) simulations^{13,25} were performed using the MUPHY software²⁶ (version 2). The conformations of the proteins were restrained by an elastic network (distance cutoff 6–8 Å, force constant 5 kcal mol⁻¹ Å⁻²). The simulations were done in an NVT ensemble at $T = 300$ K, using a time step of 10 fs for bonded interactions and a time step of 20 fs for nonbonded and hydrodynamic interactions. Solvent-mediated hydrodynamic interactions were described using the lattice Boltzmann (LB) technique,²⁷ employing the BGK (Bhatnagar-Gross-Krook) collisional operator,²⁸ with a lattice grid spacing of 3 Å. The LB kinematic viscosity ν_0 was set to reproduce bulk water behavior under ambient conditions. The parameter γ , governing the coupling with the fluid, was set separately for each protein species by matching the translational diffusion coefficient in dilute conditions with a prediction obtained from the HYDROPRO 10 software²⁹ for the temperature of 25 °C (298.15 K). For this purpose, each isolated protein molecule, described using the OPEPv4 1.8/2.1 model with the same backbone treatment as in OPEPv7, was simulated in a cubic box with an edge length of 150 Å, using the MUPHY software and the same simulation parameters as described above. The diffusion coefficient was computed by averaging the results of three independent trajectories (at least 200 ns long). For each trajectory, the diffusion coefficient was obtained by fitting the mean square displacement (MSD) curve of the protein center-of-mass in the 1–5 ns regime. The value of γ was changed iteratively until a satisfactory agreement with the HYDROPRO prediction was reached. The resulting γ values, used with the optimized OPEPv7 force field, together with the corresponding

diffusion coefficients in dilute conditions are listed in Table S1 in the Supporting Information.

The homogeneous BSA systems were formed by 16 (107 g/L) or 8 (262 g/L) BSA molecules in cubic boxes of edge length 255 or 150 Å, respectively. The binary CI2 systems contained 10 CI2 molecules placed in 255 Å cubic boxes, together with 15 BSA molecules or 70 lysozyme molecules for 100 g/L systems and 45 BSA molecules or 209 lysozyme molecules for 300 g/L systems. The 300 g/L α -synuclein/BSA system contained 10 α -synuclein molecules in the presence of 45 BSA proteins inside a 255 Å cubic box. Initial protein positions and orientations in the crowded simulation boxes were generated by the Packmol software³⁰ (version 18.169) or were adopted from previous trajectories. The trajectory lengths were 1 μ s. The initial 200 ns of each trajectory were discarded from the analysis.

Protein translational diffusion coefficients (D) were calculated from linear fits of the protein center-of-mass MSD curves (see Figure S1 in the Supporting Information). For a given protein species, the diffusion coefficients of all the individual proteins in the simulation box were averaged, and the standard error of the mean was calculated to estimate the uncertainty of the mean. The uncertainty of the reported diffusion slowdown was estimated by error propagation, combining the standard errors of the mean of the diffusion coefficient determined in crowded and dilute conditions, respectively. As described in the Supporting Information, we verified that the reported values of the diffusion slowdown were not impacted strongly by finite-size artifacts due to hydrodynamic interactions acting in the periodic system.

Two residues/proteins were considered to be in contact if the minimum distance between their beads in the given trajectory frame was below 7.34 Å, that is, a value corresponding to the largest r_{ij}^0 parameter for any pair of beads in OPEP. A group of proteins interconnected with contacts was considered a cluster.

RESULTS

The Problem of Modeling Protein Mobility in Crowded Solutions. Previous results obtained from relatively short trajectories (<100 ns), including one simulation system containing more than 18 000 proteins, suggested that the OPEP force field correctly captured the experimentally observed decrease in protein translational diffusion coefficients, as a function of solution concentration.¹² However, subsequent simulations reaching longer time scales (between hundreds of nanoseconds and one microsecond) revealed a tendency of OPEP (with the empirical scaling factor set to 1.8; see the Methods section) to excessive slowdown and gelation of the system (see Figure S2 in the Supporting Information). This phenomenon, which was not observed in the previous shorter trajectories, pointed to excessively strong side chain–side chain interactions in OPEP. While those interactions, enhanced by the empirical scaling factor of 1.8, resulted in a good performance of OPEP in capturing folding–unfolding equilibria, they did not correctly describe the weak transient interactions that arise between proteins in crowded environments. The same fate was observed for other coarse-grained models, such as Martini, when dealing with protein–protein interactions being too strong.³¹

To improve the description of crowded systems with the OPEP force field, we benchmarked its predictions of translational diffusion coefficients in homogeneous solutions

of bovine serum albumin (BSA) and binary solutions of chymotrypsin inhibitor 2 (CI2) against neutron-scattering (NS) and nuclear magnetic resonance (NMR) data, respectively, obtained for these solutions.^{32,33} We found that choosing a smaller empirical scaling factor removed the gelation issue and improved the behavior of the protein solutions. Namely, the 1.8/2.1 scaling was used to model the *E. coli* cytoplasm, and the recovered diffusivities of the proteins were found in good agreement with the experimental distribution of the diffusion coefficients, as a function of protein molecular mass in *E. coli*; see ref 34. However, when considering our benchmark systems no single scaling factor ensured the perfect agreement with experimental data for all the model systems (see Figures S3 and S4 in the Supporting Information). For the pure BSA solution, depending on the chosen scaling factor, the diffusion slowdown was either smaller or greater than the experimental estimates. For CI2, for a given scaling factor, depending on the crowders, the slowdown could be greater or smaller than the experimental finding.

A detailed inspection of the interaction histograms in pure BSA and CI2/BSA solutions showed a prominent number of contacts between lysine and glutamate (see Figure S5 in the Supporting Information). On one hand, this reflects a frequent occurrence of these two charged residues on the surfaces of the BSA and CI2 molecules (Figure S5 in the Supporting Information). On the other hand, the lysine–glutamate pair also bears the strongest attractive potential among the original OPEP side chain–side chain interactions. On the surface of the lysozyme, the distribution of charged residues is different from that on the surface of BSA (see Figure S5), with the dominant positively charged residue being arginine rather than lysine. Thus, the ion-pair interactions involving arginine, being weaker than the lysine–glutamate pair, may not have created a sufficient electrostatic attraction between molecules to reproduce the experimentally observed strong slowdown.

OPEPv7 for Crowded Solutions. The above findings suggest that the ion-pair residue interactions are key to correctly describe the diffusion behavior. Furthermore, global down-scaling of all side chain–side chain interactions may excessively weaken other polar and hydrophobic interactions. Therefore, we created OPEPv7, which is a variant with the empirical scaling factor equal to 1.0 and with selectively tweaked attractive interactions between charged residue pairs. In this optimized version of the OPEP force field, we weakened the lysine–glutamate interaction while strengthening all the other ion pairs, in particular, the arginine–glutamate and arginine–aspartate interactions (see the Methods section for details). We were guided by previous potential-of-mean-force calculations,²¹ which revealed stronger interactions for ion pairs involving arginine.

For homogeneous solutions of BSA (see Figure 1A), OPEPv7 yielded a diffusion slowdown in the short-time regime (0.3–5 ns), in agreement with neutron scattering experiments (see Figure 1B). We verified that the reported values of the diffusion slowdown were not impacted strongly by finite-size artifacts (see Figure S10 in the Supporting Information). The agreement between the simulations and available experimental data remained good for a longer time window (see Figure 1C). Regarding the structure of these crowded solutions, we found that, upon moving from a BSA concentration of 107 g/L to 262 g/L, the average number of neighbors being in contact with a given BSA molecule

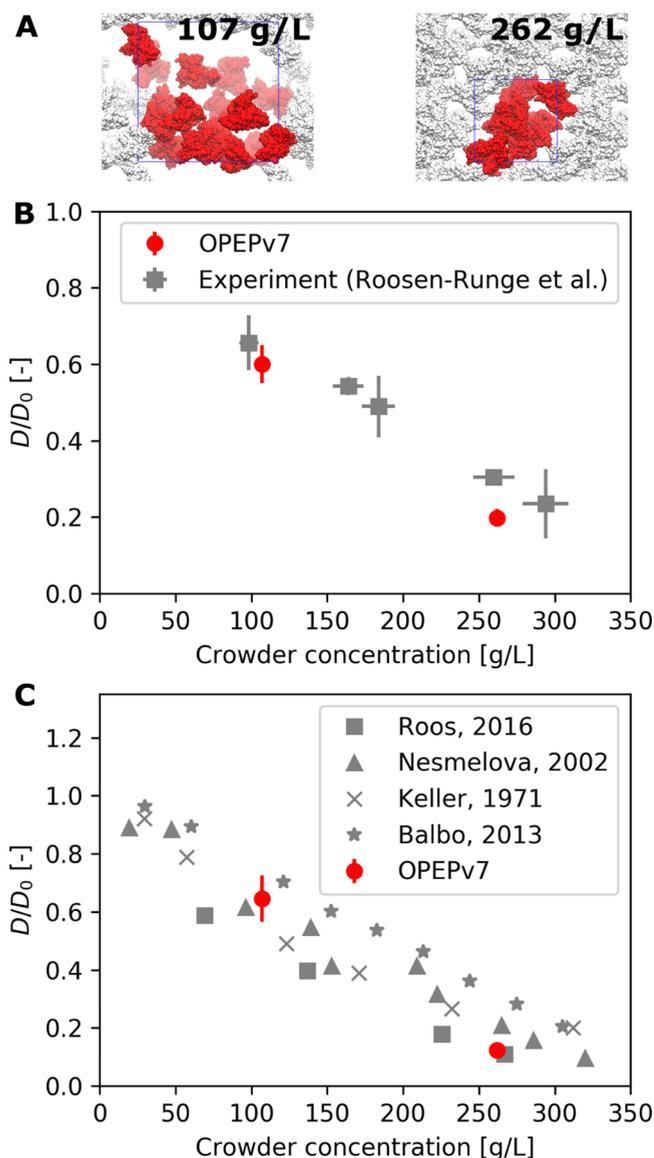


Figure 1. (A) Snapshots of crowded BSA solutions simulated using the OPEPv7 force field. BSA molecules belonging to a single unit cell are shown in red while their periodic images are displayed in white. (B) Decrease in the translational diffusion coefficient (relative to a dilute reference) observed for the 0.3–5 ns time window. The simulation results are compared with experimental data published in ref 32. (C) Diffusion slowdown obtained using the OPEPv7 force field for BSA solutions in the 30–50 ns regime, compared with experimental values^{35–38} characterizing the long-time translational diffusion of BSA.

increased from 1.3 to 3.3. While a BSA molecule spent 23% of time detached from all other proteins in the 107 g/L solution, this fraction dropped to 0.4% in the denser solution. This implies that, in dense protein solutions, the proteins are very rarely found without direct contacts to other proteins molecules but continue to engage in molecular clusters. On average, the less dense solution contained 3.0 molecular clusters with an average size of 4.9 BSA molecules while in the denser solution, virtually all BSA molecules were interconnected in a single cluster, with a single molecule occasionally detaching from the cluster (see Figure 2).

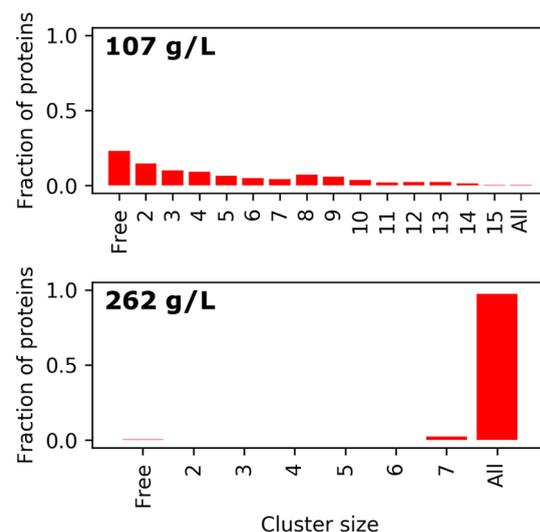


Figure 2. Concentration-dependent partitioning of BSA molecules into clusters of different sizes in crowded BSA solutions. "Free" refers to an isolated BSA molecule floating freely in the solution while "All" denotes a cluster comprising all proteins in the box.

Moving beyond homogeneous systems, we investigated the translational diffusion of chymotrypsin inhibitor 2 (CI2) in crowded solutions of BSA and lysozyme (see Figure 3A). The diffusion of CI2 under crowding was extensively characterized using pulsed-field gradient NMR,³³ revealing a strong drop in the translational diffusion coefficient at higher crowder concentrations, an effect that was particularly marked for lysozyme. In our simulations employing the OPEPv7 force field, we observed a diffusion slowdown that was similar to the experimental results, although somewhat less dramatic in the high-concentration regime for CI2 in lysozyme (see Figure 3B). However, note that the time scales probed by the NMR experiments (milliseconds) exceeded those attained in simulations. Nevertheless, we found that diffusion coefficients calculated with a variable time window started reaching a plateau above 50 ns (see Figure S6 in the Supporting Information).

On average, at the same crowder concentration, a CI2 molecule was found to be in contact with more proteins in the lysozyme solution than in the BSA solution (3.6 versus 2.1 neighbors at 300 g/L). As we showed in our previous work,¹⁴ where a preliminary contact analysis of the 300 g/L CI2/BSA and CI2/lysozyme systems was presented in the Supporting Information, the tighter local packing experienced in lysozyme may result in a stronger excluded-volume effect, stabilizing the native conformation of CI2 at high temperatures. We also noticed a relatively pronounced tendency of CI2 molecules to form dimers and oligomers. In fact, this self-clustering could be an important factor causing the observed diffusion slowdown. We observed that CI2 molecules had numerous hydrophobic contacts among each other (see Figure S7 in the Supporting Information), which could explain the tendency of CI2 to form self-clusters (see Figure 4, as well as Figure S7). Self-association of CI2 into dimers was experimentally observed both in dilute and crowded solutions,³⁹ and the fits of the NMR data suggested that the CI2 homodimer has a lower dissociation constant ($K_d = 15$ mM) than the CI2–BSA heterodimer ($K_d = 35$ mM).³⁹ We observed that the dimerization/oligomerization was less prevalent in highly

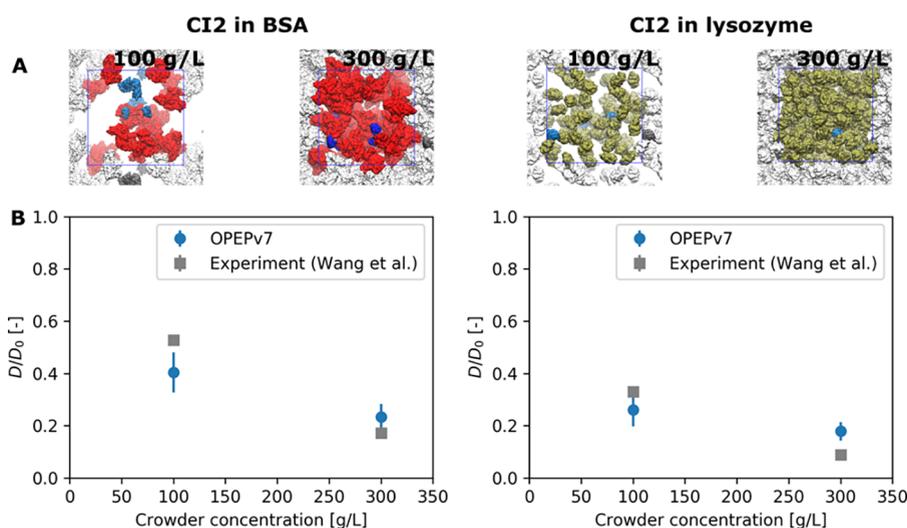


Figure 3. Dynamics of chymotrypsin inhibitor 2 (CI2) in crowded solutions of BSA and lysozyme. (A) Snapshots of the simulation systems. For a single unit cell, CI2 molecules are shown in blue, BSA molecules are shown in red, and lysozyme molecules are shown in green; the periodic images of these molecules are displayed in white. (B) Slownown of the CI2 translational diffusion obtained with the OPEPv7 force field in the 30–50 ns time window. The results are compared with experimental data from pulsed-field gradient NMR.³³

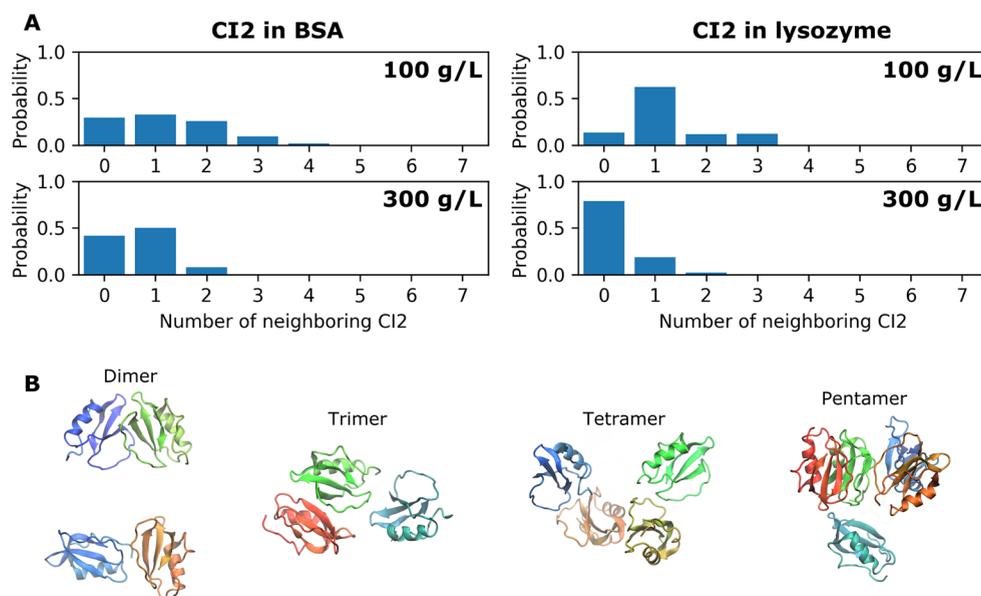


Figure 4. Contacts of chymotrypsin inhibitor 2 (CI2) in crowded solutions of BSA and lysozyme. (A) Histograms mapping the number of other CI2 molecules being in contact with a CI2 molecule. The data for the 300 g/L trajectories were presented in the Supporting Information of ref 14. (B) Representative geometries of oligomeric structures formed by CI2 in the 100 g/L CI2/BSA solution. Each CI2 molecule in the given oligomer is depicted using a different color.

crowded solutions (see Figure 4A), an effect that may stem from competing interactions with crowders.

CI2 vs α -Synuclein under Crowding. Finally, to compare the diffusion of a folded globular protein (CI2) with that of an intrinsically disordered one, we examined the diffusion of α -synuclein (α Syn) in a 300 g/L BSA solution. The structure of α Syn was modeled using a flexible model based on OPEPv7, with a single bead per amino acid (see the Methods section) and with bond potentials taken from ref 23. Our results qualitatively agreed with an experimental observation⁴⁰ demonstrating that the intrinsically disordered protein experiences a less drastic slowdown in crowded conditions, compared to a folded globular protein, represented by CI2 (see Figure 5). However, when compared to experiments, the

difference between the diffusivities of the two type of proteins is less marked. This is probably caused by the fact that our simple model produces a compact structure of α Syn with an average gyration radius of 14 Å. A very similar value was obtained by simulations including constraints from cross-linking experiments; however, note that different experimental techniques lead to values ranging from 22.6 Å to 50 Å.⁴¹

Compared to CI2, α Syn had fewer hydrophobic contacts with BSA while polar and charged contacts were enhanced (see Figure S8 in the Supporting Information). Given its glycine- and alanine-rich amino-acid sequence, α Syn frequently interacted with BSA via these two residues (see Figure S8). Similar to CI2, α Syn exhibited a tendency to form oligomers, which were promoted by hydrophobic contacts; in particular,

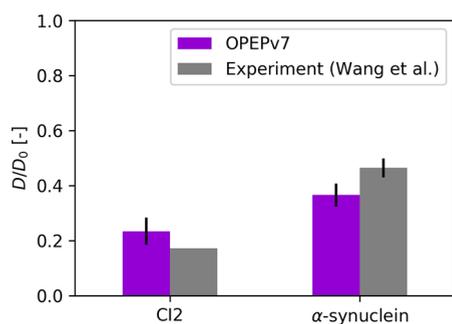


Figure 5. Slowdown of the translational diffusion coefficient in a 300 g/L BSA solution: a comparison between the globular CI2 protein and the disordered α Syn. Simulation results obtained with the OPEPv7 force field in the 30–50 ns regime are set side by side with experimental data from pulsed-field-gradient NMR.⁴⁰

those formed by VAL–VAL pairs (see Figure S8). This clustering of α Syn decreased its average diffusion coefficient.

Heterogeneous Crowding of the Cytoplasm. In this final section, we test OPEPv7 against a heterogeneous crowded solution modeling the *E. coli* cytoplasm. The model is composed of 197 proteins belonging to 35 different protein families, and simulated in a cubic box of 400 Å at a solution concentration of 279 g/L. Details on the system preparation are given in ref 34. The system was evolved for 1 μ s using LBMD. In Figure 6, we report the distribution of the translational diffusion coefficients calculated for each protein in the system as a function of the protein molecular mass. The diffusion coefficient was estimated by fitting the mean square displacement in the 30–50 ns time window. In Figure 6A, we

report the data from the previous simulation of the same model but based on the earlier version of OPEP and using the 1.8/2.1 scaling factor.³⁴ The computed diffusion coefficients already show a good agreement with the model⁴² derived from experimental data; see the solid line in Figure 6A. In Figure 6B, we report the results obtained with OPEPv7, which also works very well, reproducing the experimental trend. With respect to the ad hoc scaling, OPEPv7 somewhat reduces the slowdown caused by the high protein concentration. In order to quantify the agreement, we attempted the fit of our data using the same functional form as used in ref 42 but leaving a single free parameter. For OPEPv7, we retrieved the same value of the free parameter as that obtained for the fit of the experimental data; see Figure S11 in the Supporting Information.

Simplified Version of OPEPv7. To facilitate the implementation of OPEPv7 in other simulation packages, we developed a simplified version of the force field (OPEPv7 LJ; see the Methods section). In this version, the interaction between attractive side-chain beads was described using the standard 12–6 LJ potential. On the basis of our preliminary verifications, the OPEPv7 LJ version appears to exhibit a behavior consistent with that of OPEPv7. More specifically, a 400 ns OPEPv7 LJ simulation of the 262 g/L BSA system yielded a short-time diffusion coefficient of $1.0 \pm 0.2 \text{ \AA}^2/\text{ns}$ (0.3–5 ns regime), which is not very different from the value of $1.5 \pm 0.2 \text{ \AA}^2/\text{ns}$ obtained with OPEPv7. We noticed that OPEPv7 LJ slightly increased the interprotein contacts with, on average, a BSA molecule having 3.9 neighbors, compared to 3.3 neighbors for OPEPv7. In addition, the histograms of inter-residue contacts gave a consistent overall picture, with OPEPv7 LJ somewhat enhancing the contacts involving

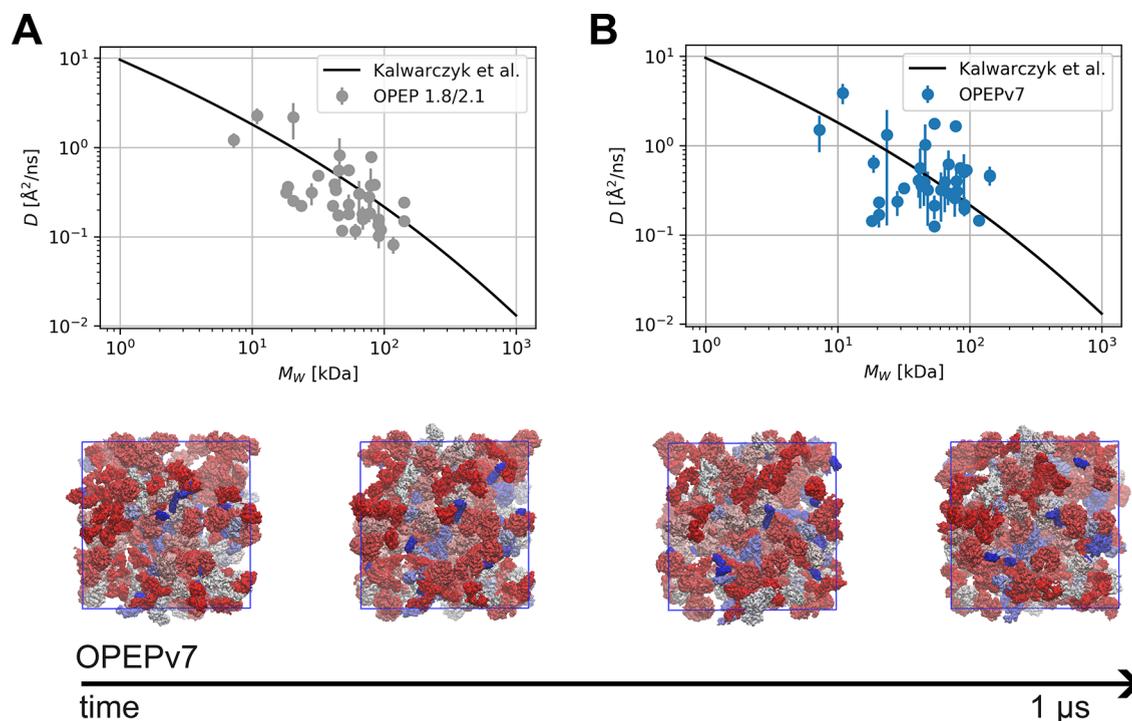


Figure 6. Distribution of the protein diffusion coefficients extracted from the simulation of a model of the *E. coli* cytoplasm (279 g/L concentration) and calculated in the 30–50 ns time interval as a function of protein molecular weight. (A) Data from a simulation based on the earlier version of OPEP but using the ad hoc scaling factor of 1.8/2.1. (B) Results from a simulation based on OPEPv7, i.e., the final optimized version. The distributions are compared with the analytical fit of experimental data reported in ref 42. In the bottom panel of the figure, we show a representation of the cytoplasm system simulated with OPEPv7 at different time frames. Each protein species is represented using a different color.

hydrophobic residues (see Figure S9 in the Supporting Information). This behavior is due to the attractive part of the potential function, which is more long-range than that of the original OPEPv7 formulation (see the Supporting Information) and which can be controlled by tuning the interaction cutoff. Note that this impact should be less strong under intermediate crowding conditions.

CONCLUSION

By reporting a high sensitivity of the translational diffusion coefficient to changes in the OPEPv7 potential, our work corroborates previous findings describing direct protein interactions as an important factor determining the diffusion slowdown in crowded protein solutions.^{8,9,11,35} We stress that, at high crowding, the effect of direct protein interactions can also be clearly seen in the short-time regime (as follows from the comparison of the BSA slowdown in Figures 1B, as well as Figures S3 and S4 in the Supporting Information), previously thought to be dominated by hydrodynamic interactions.³²

With its focus on side chain–side chain interactions, the OPEPv7 force field is suitable for the simulation of folded globular proteins. While our simulations of α Syn diffusion suggest a wider applicability of the force field to unstructured proteins, further modifications of the model are necessary in order to accurately describe the behavior of IDPs. Namely, a fine-tuning of some hydrophobic interactions should be considered, along with the implementation of intramolecular terms for ensuring chain extension.¹⁹

In conclusion, we present OPEPv7, which is a coarse-grained force field at amino-acid resolution, suited for rigid-body simulations of the structure and dynamics of heterogeneous crowded solutions formed by globular proteins. States of local packing sampled in these coarse-grained simulations can be converted into the all-atom resolution for a more-detailed exploration of protein interactions and conformations in the crowded environment.¹⁴ OPEPv7 is currently used to investigate the reorganization of functional or pathological protein assemblies, such as metabolons or amyloid systems. With this model in hand and with the flexibility of the lattice Boltzmann technique, the *in silico* rheology of crowded protein solutions is becoming possible. Finally, we offer a simplified version of the force field where the majority of the interactions are modeled with a standard LJ 12–6 potential. This formulation will allow testing the model in other molecular dynamics simulation engines and including alternative coupling schemes with hydrodynamics (e.g., Brownian dynamics with hydrodynamic interactions).

ASSOCIATED CONTENT

Data Availability Statement

The initial structures and the trajectories of the crowded systems simulated using OPEPv7 are available at [10.5281/zenodo.7778703](https://doi.org/10.5281/zenodo.7778703). Text files with force field parameters (OPEPv7 and the simplified version) can be downloaded at [10.5281/zenodo.7302991](https://doi.org/10.5281/zenodo.7302991).

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.3c00253>.

Details on the functional form of the potentials; evaluation of PBC artifacts; figures showing the results of additional trajectory analysis; values of the coupling parameter (PDF)

AUTHOR INFORMATION

Corresponding Authors

Stepan Timr – Laboratoire de Biochimie Théorique (UPR 9080), CNRS, Université de Paris, Paris 75005, France; Institut de Biologie Physico-Chimique, Fondation Edmond de Rothschild, PSL Research University, Paris 75005, France; J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences, Prague 8 18223, Czech Republic; orcid.org/0000-0002-5824-4476; Phone: +420 266 052 077; Email: stepan.timr@jh-inst.cas.cz

Fabio Sterpone – Laboratoire de Biochimie Théorique (UPR 9080), CNRS, Université de Paris, Paris 75005, France; Institut de Biologie Physico-Chimique, Fondation Edmond de Rothschild, PSL Research University, Paris 75005, France; orcid.org/0000-0003-0894-8069; Phone: +33-1-58415169; Email: fabio.sterpone@ibpc.fr

Authors

Simone Melchionna – IAC–CNR, 00185 Rome, Italy; Lexma Technology, Arlington, Massachusetts 02476, United States

Philippe Derreumaux – Laboratoire de Biochimie Théorique (UPR 9080), CNRS, Université de Paris, Paris 75005, France; Institut de Biologie Physico-Chimique, Fondation Edmond de Rothschild, PSL Research University, Paris 75005, France; Institut Universitaire de France, 75005 Paris, France; orcid.org/0000-0001-9110-5585

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jpcb.3c00253>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme, under the Marie Skłodowska-Curie Grant Agreement No. 840395. The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (No. FP7/2007-2013) under REA Grant Agreement No. PCOFUND-GA-2013-609102, through the PRESTIGE programme coordinated by Campus France. We acknowledge the financial support by the "Initiative d'Excellence" program from the French State (Grant "DYNAMO", ANR-11-LABX-0011-01, and "CACISICE", ANR-11-EQPX-0008). Part of this work was performed using HPC resources from GENCI [CINES, TGCC, IDRIS] (Grant No. x20196818) and LBT. We thank Geoffrey Letessier for technical support at the LBT.

REFERENCES

- (1) Ellis, R. J.; Minton, A. P. Join the crowd. *Nature* **2003**, *425*, 27–28.
- (2) Mittal, S.; Chowhan, R. K.; Singh, L. R. Macromolecular crowding: Macromolecules friend or foe. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 1822–1831.
- (3) Kuznetsova, I. M.; Turoverov, K. K.; Uversky, V. N. What macromolecular crowding can do to a protein. *Int. J. Mol. Sci.* **2014**, *15*, 23090–23140.
- (4) Muramatsu, N.; Minton, A. P. Tracer diffusion of globular proteins in concentrated protein solutions. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 2984–2988.

- (5) Kekenus-Huskey, P. M.; Eun, C.; McCammon, J. A. Enzyme localization, crowding, and buffers collectively modulate diffusion-influenced signal transduction: Insights from continuum diffusion modeling. *J. Chem. Phys.* **2015**, *143*, 094103.
- (6) Gruebele, M.; Thirumalai, D. Perspective: Reaches of chemical physics in biology. *J. Chem. Phys.* **2013**, *139*, 121701.
- (7) Yu, I.; Mori, T.; Ando, T.; Harada, R.; Jung, J.; Sugita, Y.; Feig, M. Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *eLife* **2016**, *5*, No. e19274.
- (8) von Bülow, S.; Siggel, M.; Linke, M.; Hummer, G. Dynamic cluster formation determines viscosity and diffusion in dense protein solutions. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 9843–9852.
- (9) McGuffee, S. R.; Elcock, A. H. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput. Biol.* **2010**, *6*, No. e1000694.
- (10) Ando, T.; Skolnick, J. Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 18457–18462.
- (11) Nawrocki, G.; Wang, P.-h.; Yu, I.; Sugita, Y.; Feig, M. Slow-down in diffusion in crowded protein solutions correlates with transient cluster formation. *J. Phys. Chem. B* **2017**, *121*, 11072–11084.
- (12) Sterpone, F.; Melchionna, S.; Tuffery, P.; Pasquali, S.; Mousseau, N.; Cragolini, T.; Chebaro, Y.; St-Pierre, J.-F.; Kalimeri, M.; Barducci, A.; et al. The OPEP protein model: From single molecules, amyloid formation, crowding and hydrodynamics to DNA/RNA systems. *Chem. Soc. Rev.* **2014**, *43*, 4871–4893.
- (13) Sterpone, F.; Derreumaux, P.; Melchionna, S. Protein simulations in fluids: Coupling the OPEP coarse-grained force field with hydrodynamics. *J. Chem. Theory Comput.* **2015**, *11*, 1843–1853.
- (14) Timr, S.; Sterpone, F. Stabilizing or destabilizing: Simulations of chymotrypsin inhibitor 2 under crowding reveal existence of a crossover temperature. *J. Phys. Chem. Lett.* **2021**, *12*, 1741–1746.
- (15) Chiricotto, M.; Melchionna, S.; Derreumaux, P.; Sterpone, F. Hydrodynamic effects on β -amyloid (16–22) peptide aggregation. *J. Chem. Phys.* **2016**, *145*, 035102.
- (16) Chiricotto, M.; Melchionna, S.; Derreumaux, P.; Sterpone, F. Multiscale aggregation of the amyloid A β 16–22 peptide: From disordered coagulation and lateral branching to amorphous prefibrils. *J. Phys. Chem. Lett.* **2019**, *10*, 1594–1599.
- (17) Sterpone, F.; Derreumaux, P.; Melchionna, S. Molecular mechanism of protein unfolding under shear: A lattice Boltzmann molecular dynamics study. *J. Phys. Chem. B* **2018**, *122*, 1573–1579.
- (18) Brandner, A. F.; Timr, S.; Melchionna, S.; Derreumaux, P.; Baaden, M.; Sterpone, F. Modelling lipid systems in fluid with Lattice Boltzmann Molecular Dynamics simulations and hydrodynamics. *Sci. Rep.* **2019**, *9*, 16450.
- (19) Languin-Cattoën, O.; Laborie, E.; Yurkova, D. O.; Melchionna, S.; Derreumaux, P.; Belyaev, A. V.; Sterpone, F. Exposure of von Willebrand factor cleavage site in A1A2A3-fragment under extreme hydrodynamic shear. *Polymers* **2021**, *13*, 3912.
- (20) Timr, S.; Gnutt, D.; Ebbinghaus, S.; Sterpone, F. The unfolding journey of superoxide dismutase 1 barrels under crowding: Atomistic simulations shed light on intermediate states and their interactions with crowders. *J. Phys. Chem. Lett.* **2020**, *11*, 4206–4212.
- (21) Sterpone, F.; Nguyen, P. H.; Kalimeri, M.; Derreumaux, P. Importance of the ion-pair interactions in the OPEP coarse-grained force field: Parametrization and validation. *J. Chem. Theory Comput.* **2013**, *9*, 4574–4584.
- (22) Marino, S. M.; Gladyshev, V. N. Cysteine function governs its conservation and degeneration and restricts its utilization on protein surfaces. *J. Mol. Biol.* **2010**, *404*, 902–916.
- (23) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Best, R. B.; Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Comput. Biol.* **2018**, *14*, No. e1005941.
- (24) Tan, C.; Jung, J.; Kobayashi, C.; Torre, D. U. L.; Takada, S.; Sugita, Y. Implementation of residue-level coarse-grained models in GENESIS for large-scale molecular dynamics simulations. *PLoS Comput. Biol.* **2022**, *18*, No. e1009578.
- (25) Ahlrichs, P.; Dünweg, B. Simulation of a single polymer chain in solution by combining lattice Boltzmann and molecular dynamics. *J. Chem. Phys.* **1999**, *111*, 8225–8239.
- (26) Bernaschi, M.; Melchionna, S.; Succi, S.; Fyta, M.; Kaxiras, E.; Sircar, J. K. MUPHY: A parallel MULti PHYSics/scale code for high performance bio-fluidic simulations. *Comput. Phys. Commun.* **2009**, *180*, 1495–1502.
- (27) Succi, S. *The Lattice Boltzmann Equation: For Fluid Dynamics and Beyond*; Oxford University Press, 2001.
- (28) Benzi, R.; Succi, S.; Vergassola, M. The lattice Boltzmann equation: Theory and applications. *Phys. Rep.* **1992**, *222*, 145–197.
- (29) Ortega, A.; Amorós, D.; De La Torre, J. G. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* **2011**, *101*, 892–898.
- (30) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.
- (31) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thalmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; et al. Martini 3: A general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, *18*, 382–388.
- (32) Roosen-Runge, F.; Hennig, M.; Zhang, F.; Jacobs, R. M. J.; Sztucki, M.; Schober, H.; Seydel, T.; Schreiber, F. Protein self-diffusion in crowded solutions. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 11815–11820.
- (33) Wang, Y.; Li, C.; Pielak, G. J. Effects of proteins on protein diffusion. *J. Am. Chem. Soc.* **2010**, *132*, 9392–9397.
- (34) Di Bari, D.; Timr, S.; Guiral, M.; Giudici-Orticoni, M.-T.; Seydel, T.; Beck, C.; Petrillo, C.; Derreumaux, P.; Melchionna, S.; Sterpone, F.; et al. Diffusive dynamics of bacterial proteome as a proxy of cell death. *ACS Central Sci.* **2023**, *9*, 93–102.
- (35) Roos, M.; Ott, M.; Hofmann, M.; Link, S.; Rössler, E.; Balbach, J.; Krushelnitsky, A.; Saalwächter, K. Coupling and decoupling of rotational and translational diffusion of proteins under crowding conditions. *J. Am. Chem. Soc.* **2016**, *138*, 10365–10372.
- (36) Nesmelova, I. V.; Skirda, V. D.; Fedotov, V. D. Generalized concentration dependence of globular protein self-diffusion coefficients in aqueous solutions. *Biopolymers* **2002**, *63*, 132–140.
- (37) Keller, K. H.; Canales, E. R.; Yum, S. Tracer and mutual diffusion coefficients of proteins. *J. Phys. Chem.* **1971**, *75*, 379–387.
- (38) Balbo, J.; Mereghetti, P.; Herten, D. P.; Wade, R. C. The shape of protein crowders is a major determinant of protein diffusion. *Biophys. J.* **2013**, *104*, 1576–1584.
- (39) Li, C.; Pielak, G. J. Using NMR to distinguish viscosity effects from nonspecific protein binding under crowded conditions. *J. Am. Chem. Soc.* **2009**, *131*, 1368–1369.
- (40) Wang, Y.; Benton, L. A.; Singh, V.; Pielak, G. J. Disordered protein diffusion under crowded conditions. *J. Phys. Chem. Lett.* **2012**, *3*, 2703–2706.
- (41) Nguyen, P. H.; Ramamoorthy, A.; Sahoo, B. R.; Zheng, J.; Faller, P.; Straub, J. E.; Dominguez, L.; Shea, J.-E.; Dokholyan, N. V.; De Simone, A.; et al. Amyloid oligomers: A joint experimental/computational perspective on Alzheimer's disease, Parkinson's disease, type II diabetes, and amyotrophic lateral sclerosis. *Chem. Rev.* **2021**, *121*, 2545–2647.
- (42) Kalwarczyk, T.; Tabaka, M.; Holyst, R. Biologistics—Diffusion coefficients for complete proteome of *Escherichia coli*. *Bioinformatics* **2012**, *28*, 2971–2978.