In Cell Protein-Protein Contacts: Transient Interactions in the Crowd

Meredith M. Rickard,[†] Yi Zhang,[§] Martin Gruebele,[†,§,‡,*] and Taras V. Pogorelov[†,§,*]

[†]*Department of Chemistry, University of Illinois, Urbana-Champaign, Illinois 61801, USA*

[§,]*Center for Biophysics and Computational Biology, University of Illinois, Urbana-Champaign, Illinois 61801, USA*

[‡,]*Department of Physics, University of Illinois, Urbana-Champaign, Illinois 61801, USA*

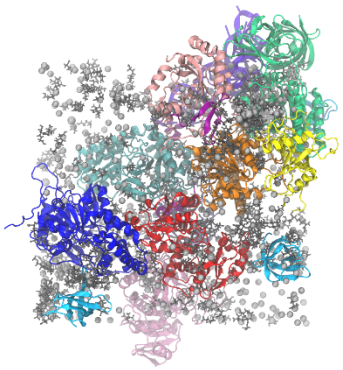
AUTHOR INFORMATION

Corresponding Authors

*****To whom correspondence may be addressed. Email: mgruebel@illinois.edu and pogorelo@illinois.edu

ABSTRACT Proteins *in vivo* are immersed in a crowded environment of water, ions, metabolites, and macromolecules. In-cell experiments highlight how transient weak protein-protein interactions promote (via functional "quinary structure") or hinder (via competitive binding or "sticking") complex formation. Computational models of the cytoplasm are expensive. We tackle this challenge with an all-atom model of a small volume of the *E. coli* cytoplasm to simulate protein-protein contacts up to the 5 microsecond timescale on the special-purpose supercomputer Anton2. We use three CHARMM-derived force fields: C22*, C36m, and C36mCU (with CUFIX corrections). We find that both C36m and C36mCU form smaller contact surfaces than C22*. Although CUFIX was developed to reduce protein-protein sticking, larger contacts are observed with C36mCU than C36m. We show that the lifespan Δt of protein-protein contacts obeys a power-law distribution between 0.03-3 μs, with ~90% of all contacts lasting <1 μs (similar to the timescale for downhill folding).

TOC GRAPHIC



KEYWORDS crowding; hydrophobicity; electrostatics; molecular dynamics

Proteins *in vivo* are surrounded by a complex cellular milieu composed of macromolecules and small metabolite molecules, in addition to the water and inorganic ions that typically surround proteins in experiments *in vitro*. Of particular importance is the high concentration of macromolecules, which can range from 10-40% of the cell by volume.[1–3] In such a crowded environment, proteins have evolved to form transient functional contacts (quinary structure) and to minimize non-functional competitive binding (sticking).[4,5]

High-stability functional protein-protein interactions have been studied for decades, including as potential drug targets.[6,7] More transient interactions are recently receiving more attention.[8] A number of databases have been created to characterize the total space of the interactome.[6] Both computational and experimental studies have been conducted to investigate the physical mechanisms underlying quinary structure (electrostatics, hydrophobic interactions).[4,5,9] Experiments have indicated that crowding and non-specific interactions between neighboring proteins also contribute to the differences in protein dynamics observed *in vitro* and *in vivo*.[4,10]

Recent experiments have highlighted the dependence of a protein's stability[11] and unfolded ensemble[12] on its local environment. Further, several results emphasize the importance of the local electrostatic environment that cannot be replicated with inert crowders.[5,10,13] These results point to a bigger picture of "free for all" of nonspecific protein-protein interactions caused by local charged or hydrophobic patches sticking to complementary patches of neighboring proteins. Attempts to characterize sticking have largely focused on the experimental observation of anomalous translational or rotational diffusion *in vivo*.[4,14,15]

Atomistic computational studies in crowded environments have been limited by the computational cost of simulating such large systems. Early attempts to simulate the crowded cellular environment have involved modelling proteins as hard spheres.[16,17] More recently, protein diffusion *in vivo* has been simulated using Brownian Dynamics or coarse-graining techniques.[18–22] Others have simulated crowded environments by including multiple copies of small proteins.[9,23–25] A $>10^8$ atom model of the *Mycoplasma genitalium* cytosol simulated by all-atom molecular dynamics for 100 ns is the current state-of-the-art.[26] To date, no in-cell simulations using all-atom MD have reached the μs timescale relevant for sampling sticking processes.

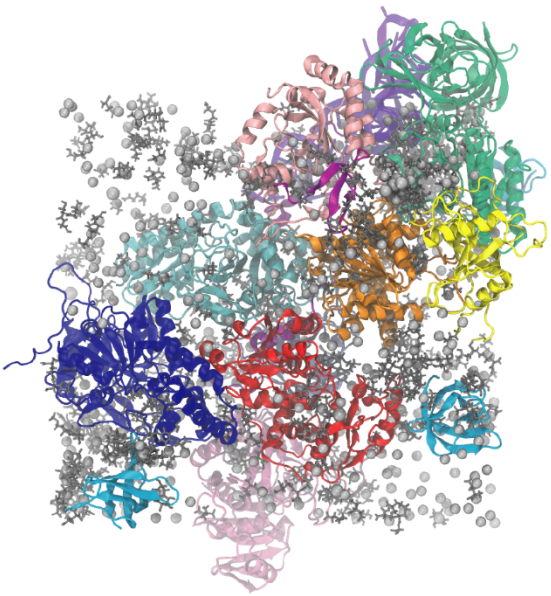
Model design We present a $2{\cdot}10^5$ all-atom model that mimics a small volume of the *E. coli* cytosol to characterize nonspecific protein-protein interactions on a multi-μs timescale. Proteins and RNA

were selected from among the most abundant cytoplasmic components (Fig. 1). We chose smaller proteins to maximize the number of interacting surfaces in our model. We included monomers of several proteins known to oligomerize (e.g. GAPDH) to see whether in absence of their binding partner, their exposed oligomerization interfaces would stick any more than monomeric protein surfaces. Our hypothesis is they would not stick more because they evolved to reduce non-specific binding, just like monomeric proteins. The bacterial cytoplasm models analyzed in this work contained at least sixteen copies of over a dozen different macromolecules, including a cold shock protein, an elongation factor, tRNA and several abundant enzymes (Fig. 1). We also included unfolded copies of the fast-folding protein WW GTT,[27] as a model for sticking of unfolded exposed hydrophobic surface area (SI Fig. S1).

To recreate the cellular environment of *E. coli*, small molecules were then added in accordance with experimentally measured concentrations.[28] Due to the finite size of the box, only the most abundant metabolite species were sampled. Metabolite numbers for each model were either picked from the experimental concentration distribution directly, with a cutoff at <1 molecule/box, or a Monte Carlo code was used to sample the number of copies of each metabolite based on the experimental metabolite concentrations (see SI Table S1/Fig. S2). Across all systems, about 45 species of metabolites were sampled, from the most common (glutamate, 96 mM) to 59th most common (methionine, 0.15 mM). After the addition of inorganic ions and explicit solvent, the final system contains approximately 200,000 atoms. The macromolecular concentration of the systems are ~300 mg/mL with ~220 mM metabolites and ~350 mM inorganic ions.

For production runs, the systems were simulated on the special-purpose supercomputer Anton2,[29] using three different versions of the CHARMM force field: CHARMM 22*,[30] CHARMM 36m,[31] and CHARMM 36m with CUFIX, which was designed to reduce sticking by tuning Lennard-Jones parameters for certain atom pairs.[32] We will use the abbreviations C22*, C36m and C36mCU for modeled systems from here on. The first 15 μs of trajectory were discarded for system equilibration: a drift towards increased contact area occurred initially, before contact area in all simulations settled at ~10-15 μs (see SI Fig. S3). Therefore, the final 5.016 μs of the three trajectories are analyzed here, and the first and second half of this window yield consistent contact area distributions. (Fig. S4) Of course, we cannot exclude that a very long-lived metastable cluster formed after ca. 10 μs. We observe that all proteins associate with this dynamic cluster throughout the rest of the simulation, although proteins leave and rejoin the cluster frequently.

The three models' composition varies slightly. Details of all three model runs (force field used, box size, number of each simulated species) are provided in SI Tables S1-4. The differences (e.g. a folded WW GTT in two simulations substituting a disordered peptide in another; slightly different metabolite distributions) were included for future analyses not addressed in this work. They made a negligible contribution to sticking statistics (see SI Fig. S5-S7), the largest being a contact made by WW GTT 1 (F) (thus included in Fig. 2a).



	Adenylosuccinate Synthetase (ADSS) PDB ID: 1ADE		Mn Superoxide Dismutase (Mn SOD) PDB ID: 1D5N
	Elongation Factor Tu (EFTu) PDB ID: 1OB2		Inorganic Pyrophosphatase (PPA) PDB ID: 2EIP
	Phosphoserine Aminotransferase (PSAT) PDB ID: 1BJO		Cyclophilin A (CYC A) PDB ID: 1LOP
	Aspartate Transcarbamoylase (ATCase) PDB ID: 1ZA1		Cold Shock Protein (3 copies) (CSP) PDB ID: 1MJC
	Malate Dehyrogenase (MD) PDB ID: 2CMD		GTT WW domain (3 copies) (GTT) PDB ID: 2F21 (modified)
	Glyceraldehyde 3-phosphate Dehydrogenase (GAPDH) PDB ID: 1S7C		Phe-tRNA (RNA) PDB ID: 1OB2

Figure 1. A snapshot of the *E. coli* cytoplasm model system. Color-coded macromolecules comprising the system are listed below along with copy number and the PDB ID. Ions and small metabolite molecules are shown in gray. Water molecules are not shown for clarity.

For analysis of protein-protein contacts, snapshots were taken once every 24 ns throughout the trajectories. Although the Phe-tRNA forms multiple contacts with neighboring macromolecules, these contacts are not considered here because we are focusing on protein-protein interactions in this work. To calculate the area of the protein-protein contacts, we measured the difference between the solvent accessible surface area (SASA) of each protein alone and in a protein-protein complex for every potential protein-protein interaction in the system. To account for periodic boundary conditions, the contacts included any extension of a contact area into wrapped regions.

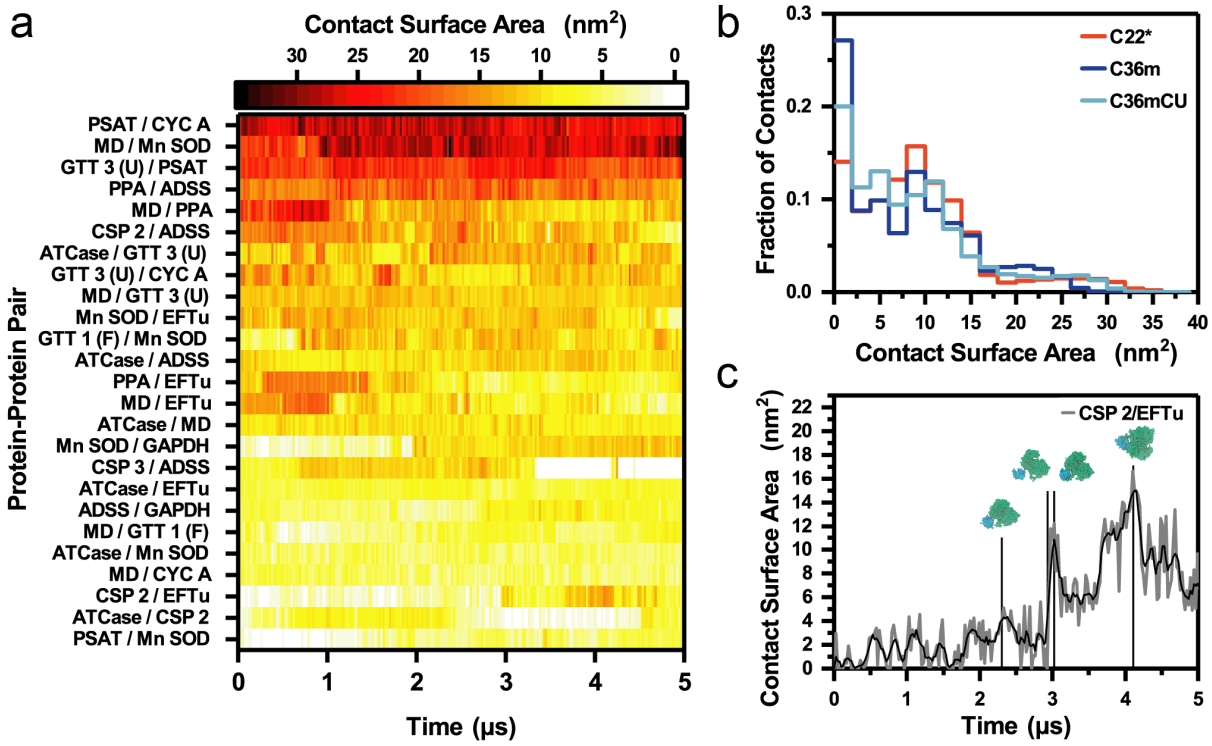


Figure 2. (a) Contact size for largest protein-protein contacts during the C36mCU simulation. Heat maps of the other two models are provided in SI Fig. S8. (b) Distribution of protein-protein contact sizes for CHARMM 22*, 36m, and 36m with CUFIX corrections. (c) Surface area of the CSP2/EFTu contact in the C36mCU model across the trajectory with snapshots at 2.3, 2.9, 3.0, and 4.1 µs.

Protein clustering As has been observed previously even in simulations with less than half the protein concentration,[9,18,23,25] the cytoplasmic proteins tend to cluster, forming protein-rich and water rich regions in the simulation box (see Fig. 1 top). As discussed in reference 20, protein-protein interactions are probably overestimated relative to protein-water interactions in the current generation of force fields, the same effect that is also known to overestimate protein stability in simulations.[33] To quantify such sticking, we calculated the number of unique neighbors each

protein encountered throughout the trajectory. All proteins in all three models formed at least one protein-protein contact; all but one protein in all three models contacts at least two proteins simultaneously at some point. We find that at any given time, proteins on average are in contact with 4.6 unique protein neighbors. These results agree well with previous Brownian Dynamics simulations of crowded cellular environments,[18] which predict 2-7 unique neighbors for similar-sized proteins. Our observation of ~35 ± 3 simultaneous pair contacts out of 105 possible at any given time is also consistent with about 5 neighbors.

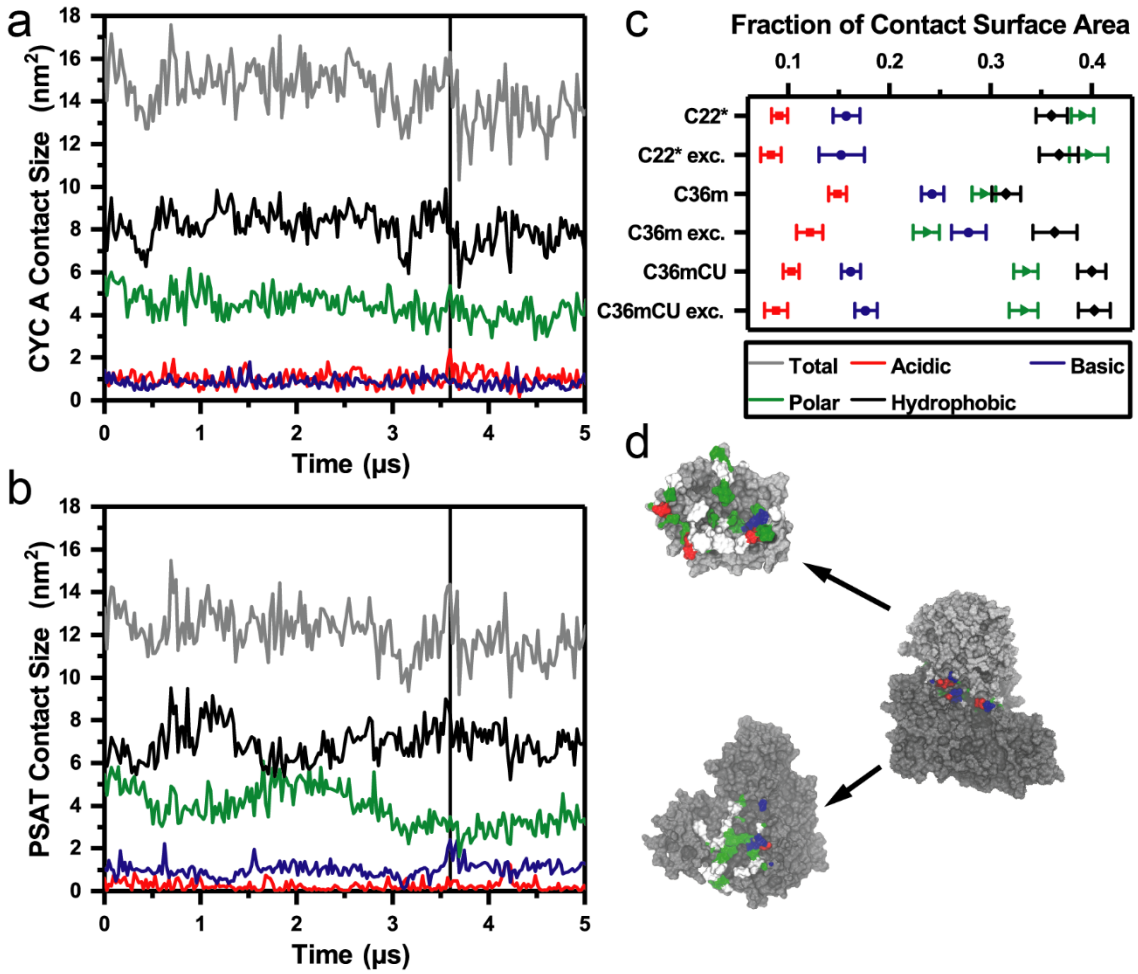


Figure 3. (a), (b) Contact surface areas of the PSAT/CYC A contact in the C36mCU simulation. For each contributing protein, hydrophobic, polar, acidic, and basic contributions are shown. (c) Relative contributions of hydrophobic, polar, acidic, and basic residues in protein-protein contacts across all systems. Error bars are based on standard deviations, which in turn are calculated over time. "Exc." refers to the exclusion of oligomer-oligomer contacts from the dataset. (d) Snapshot of the protein-protein contact at 3.6 μs. Residues at the protein-protein interface are colored according to residue type.

The difference of protein-protein contact size distributions between force fields (Figure 2b) indicates that the C36m system is markedly less sticky than C22* and C36mCU. C22* also has more very large (>30 nm^2) protein-protein contacts. CHARMM 22* was originally designed to fold single proteins in water, and thus may over-stabilize contacts between amino acids that are not close together in protein sequence, such as tertiary structure within a single protein, or surface amino acid contacts between different proteins.

Previous experimental studies of nonspecific protein-protein interactions in crystal structures have indicated that on average, these have a contact area of ~5.7 nm^2.[34] In our three systems, the mean contact area was 9.5 nm^2 (C22*), 8.1 nm^2 (C36m) and 8.5 nm^2 (C36mCU). In addition, a plateau in the distribution was observed at ~10 nm^2 for all three models. The plateaus cannot be wholly attributed to unfolded WW domain model proteins included in the simulation (see below), as they persist when these proteins are removed from the average (see SI Fig. S7).

Unfolded protein contacts The unfolded copies of the GTT model protein (a 33 residue WW domain) are more prone to forming large, long-lived contacts than folded protein surfaces in the simulation. Of the 10 largest contacts on average, a single unfolded copy of GTT participates in 4 (Figure 2a) even though it represents only 1 of up to 17 macromolecules in the simulation. We attribute this to exposure of hydrophobic core residues, in agreement with previous experimental reports[15,35,36] and computational modeling of in-cell systems[18] that point to hydrophobic interactions as an important source of nonspecific protein-protein sticking.

Additionally, other computational[9] and experimental[4,5] studies have pointed to electrostatics driving protein-protein contacts. Figure 3 shows, for contacts in the >10 nm^2 range, what fraction of the contact surface is covered by hydrophobic, hydrophilic, and charged amino acid side chains, relative to their abundances on the complete set of protein surfaces (averaged for the 3 models) simulated here. Hydrophobic and polar residues tend to dominate over the formation of salt bridges between charged residues in all three models.

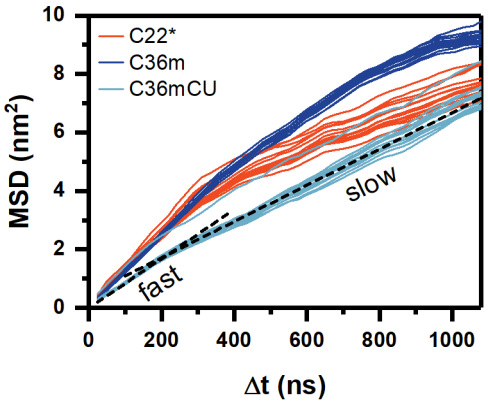


Figure 4: Mean squared displacement of proteins in C22* (orange), C36m (dark blue), and C36mCU (light blue) models versus elapsed time. The straight dotted lines superimposed on the C36m CU simulations illustrate the 'fast' and 'slow' linear regimes. Cold Shock Protein 1 in the C22* model and Cold Shock Proteins 1 & 3 in the C36mCU model were outliers that diffused more quickly than other proteins and were omitted from this figure.

Diffusion is the transport mechanism preceding macromolecular sticking. The effective diffusion could be slowed down and made anomalous by sticking, particularly if measured over long times. As has been discussed,[37] the bacterial cytoplasm has glassy properties at long scales, modulated by metabolic activity. Protein clustering that contributes to slow dynamics can be seen in Fig. 1, and may be exacerbated by the small volume and periodic boundary condition of our simulation. Figure 4 shows the squared displacement of individual proteins versus elapsed time in all three models. All models show evidence of anomalous diffusion over long time periods: There is a 'fast' linear regime from ~30 to 200 ns, a 'slow' linear regime from ~200 to 800 ns, and an overall roll-over at long times. C36m in particular is least well separated into 'fast' and 'slow' linear regimes, with just a gradual roll-over.

Previous all-atom simulations of proteins in crowded environments[9,24,26] have indicated larger diffusion coefficients than we find here. Our 'fast' linear regime produces diffusion coefficients (~1.4-2 $nm^2/\mu s$, see SI Table S5) slower than experiment[38] by about a factor of 3-5 when left uncorrected for finite size effects.[39] The subsequent 'slow' linear regime goes hand-in-hand with the protein sticking noted above—as proteins crowd or stick together and form clusters, they become less free to diffuse about the box. It is worth noting that previous simulations measure diffusion coefficients using much shorter timescales than the work presented here, highlighting the potential importance protein sticking for long-time transport dynamics.

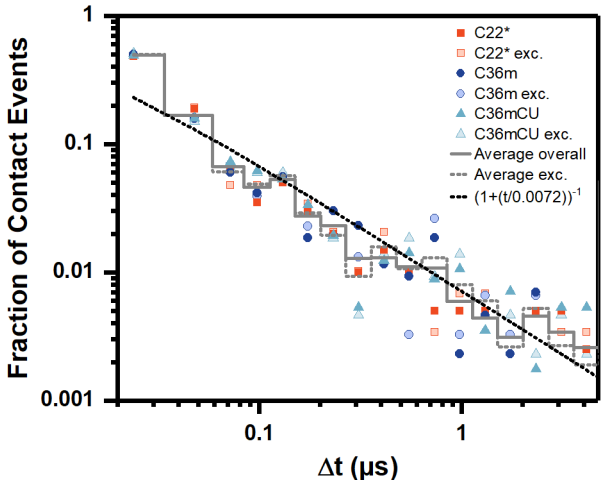


Figure 5. Distribution of protein-protein contact lifetimes. Only ≈7% of transient contacts last longer than 1 μs in all three models; about half of these lasted throughout the simulation window of 5 μs, so their ultimate potential duration is unknown and not shown. "Exc." refers to the exclusion of oligomer-oligomer contacts from the dataset. A minimum 0.15 nm^2 contact area threshold was used; larger thresholds produce similar distributions (see SI Fig. S9).

Protein-protein contact lifetimes Our results for all protein-protein pairs indicate that contacts are highly dynamic (Figure 2a.). For example, the contact between the cold shock protein number 2 (CSP 2) and the elongation factor Tu (EFTu) is broken and reformed repeatedly; in one case, a contact forms and reaches >10 nm^2 in under 50 ns (Figure 2c). To determine the timescales of the contact events, we compare the distribution of protein-protein contact lifespans Δt in all three simulations (Figure 5). Despite differences observed in contact area distribution between the three force fields, the timescales of protein-protein contacts do not differ appreciably between the three force fields simulated: with a 15 $Å^2$ threshold defining a contact, 90% the probability $P(\Delta t)$ decays on the order of ~1 μs.

Whether interactions between oligomeric proteins are excluded ("exc." in Fig. 5) or not has little effect on the probability distribution in Fig. 5. This supports our hypothesis that while oligomeric proteins are of course prone to interacting with their partners, as monomers they are not any more prone to interact with other proteins in the cytoplasmic matrix than are monomeric proteins. The average distribution (gray stepped curves in Fig. 5) is essentially the same either way. Thus, it appears that binding interfaces of oligomeric proteins, just like surfaces of monomeric proteins, have co-evolved with the cytosol to avoid non-specific interactions, as also indicated by previous studies of protein-protein interaction networks.[40–43]

P(Δt) was fitted between 50 ns and 4 μs to a stretched exponential and to power laws (see SI Fig. S10). The best fit was obtained with a power law probability distribution finite as t approaches 0,[44] which yielded the equation $P(\Delta t) \sim (1+(\Delta t/0.0072))^{-1}$. As discussed in ref. 45, for 'birth-death' processes with constant rates (here: sticking-unsticking processes with constant rate coefficients for making or breaking protein-protein contact pairs), both stretched exponentials and power laws with exponents close to -1 are possible.[45] The latter happens when the number of contacts at any given time is proportional to the lifespan Δt of contacts and the distribution of contacts is sampled rapidly. In terms of the power spectrum $P(\Delta\omega)$, where $\Delta\omega$ is the contact frequency Fourier-conjugate to contact lifespan Δt, a $1/\Delta t$ probability distribution simply implies a constant power spectrum $P(\Delta\omega)$.[46]

Origins of sticking The decreased diffusion coefficients, observed protein clustering, and larger than expected protein-protein contacts indicate that all three force fields are "too sticky"— they over-stabilize the frequency of formation and persistence of protein-protein contacts. The large contribution of hydrophobic residues to the largest contacts we observe here indicate that the source of this stickiness may lie in hydrophobic interactions. The 36m model, which has the smallest overall protein-protein contact areas, shows a larger contribution of charged residues relative to hydrophobic residues. We hypothesize that the CUFIX corrections, which have focused largely on charged and polar interactions,[32] destabilize these charge-charge interactions but not hydrophobic interactions that drive the protein-protein sticking we observe.

Hydrophobicity results from a water-water interactions at the expense of water-protein interactions.[47] Previous simulations have shown that the TIP3P water model we use leads to overly compact disordered states of unfolded proteins.[48] If water-water interactions are stabilized in current CHARMM force fields, excessive protein-protein sticking and lowered macromolecular diffusion would result. This would also explain current force fields' tendencies to overestimate protein melting temperatures,[33] as proteins would be more likely to favor protein-protein over protein-water interactions and remain folded at higher temperatures. Recent results indicate that simulations using the TIP4P-D water model[48] can reproduce experimental viscosity measurements of crowded protein solutions.[24] Other groups have attempted to correct this sticking problem by increasing Lennard-Jones protein-water interactions in crowded systems.[9,49]

We note that the timescale of protein-protein contacts decays on the order of ~1 μs, with ~7% of protein-protein contacts surviving longer than 1 μs without being broken (Fig. 5). This ~1 μs

timescales is on the order of the speed limit for downhill protein folding.[50–52] In the crowded cytoplasm, the speed limits of intra- and inter-molecular protein contacts are not appreciably different. Just as tertiary structure is formed by hydrophobically-driven docking of two secondary structural elements within a single protein, protein-protein contacts in Fig. 3 are driven by hydrophobic and polar interactions in close vicinity.

We report the first μs-timescale all-atom MD simulation of protein dynamics in a bacterial cytoplasm model. We find that three versions of the CHARMM forcefield over-stabilize non-specific sticking between proteins and their surrounding environment, with C36m the least sticky. In the screened high-ion-concentration milieu of our model cytoplasm, protein-protein interaction is driven mainly by hydrophobic and polar contacts and leads to translational diffusion 4-8 times slower than experimental values. Sticking occurs on the μs timescale or faster, similar to the time scale of the intra-molecular tertiary structure formation. Competition with sticking may have driven evolution of fast tertiary contact (re)formation in already-folded proteins that become partly unfolded. Our study demonstrates the potential of crowded all-atom models to offer new insight into protein dynamics *in vivo*. As molecular dynamics to continues to grow as a technique to study biomolecules *in vivo*, care must be taken to develop force fields that accurately replicate sticking and non-competitive binding in crowded environments.

ASSOCIATED CONTENT

Supporting Information. Supporting information contains supplementary methods, supplementary figures, supplementary tables, parameter and coordinate files of metabolites that were parametrized for this study, and a molecular dynamics movie.

AUTHOR INFORMATION

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

Anton2 computer time was provided by the Pittsburgh Supercomputing Center (PSC) through Grant R01GM116961 from the National Institutes of Health. The Anton2 machine at PSC was

generously made available by D.E. Shaw Research. T. V. P. acknowledges support from the Department of Chemistry, University of Illinois at Urbana-Champaign. M.R., and M.G. were supported by the NSF grant MCB 1803786. We thank Prof. James Imlay for helpful discussions during model construction and Christopher Maffeo for assistance with Brownian Dynamics.

References

(1) Theillet, F.-X.; Binolfi, A.; Frembgen-Kesner, T.; Hingorani, K.; Sarkar, M.; Kyne, C.; Li, C.; Crowley, P. B.; Gierasch, L.; Pielak, G. J.; et al. Physicochemical Properties of Cells and Their Effects on Intrinsically Disordered Proteins (IDPs). *Chem. Rev.* **2014**, *114*, 6661–6714.

(2) Minton, A. P. The Influence of Macromolecular Crowding and Macromolecular Confinement on Biochemical Reactions in Physiological Media. *J. Biol. Chem.* **2001**, *276*, 10577–105780.

(3) Mittal, S.; Chowhan, R. K.; Singh, L. R. Macromolecular Crowding: Macromolecules Friend or Foe. *Biochim. Biophys. Acta - Gen. Subj.* **2015**, *1850*, 1822–1831.

(4) Mu, X.; Choi, S.; Lang, L.; Mowray, D.; Dokholyan, N. V; Danielsson, J.; Oliveberg, M. Physicochemical Code for Quinary Protein Interactions in Escherichia Coli. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E4556–E4563.

(5) Cohen, R. D.; Pielak, G. J. Electrostatic Contributions to Protein Quinary Structure. *J. Am. Chem. Soc.* **2016**, *138*, 13139–13142.

(6) De Las Rivas, J.; Fontanillo, C. Protein–Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks. *PLoS Comput. Biol.* **2010**, *6*, e1000807.

(7) Modell, A. E.; Blosser, S. L.; Arora, P. S. Systematic Targeting of Protein–Protein Interactions. *Trends Pharmacol. Sci.* **2016**, *37*, 702–713.

(8) Sukenik, S.; Ren, P.; Gruebele, M. Weak Protein–Protein Interactions in Live Cells Are Quantified by Cell-Volume Modulation. *Proc. Natl. Acad. Sci.* **2017**, *114*, 6776–6781.

(9) 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.

(10) Davis, C. M.; Gruebele, M. Non-Steric Interactions Predict the Trend and Steric Interactions the Offset of Protein Stability in Cells. *ChemPhysChem* **2018**, *19*, 2290–2294.

(11) Feng, R.; Gruebele, M.; Davis, C. M. Quantifying Protein Dynamics and Stability in a Living Organism. *Nat. Commun.* **2019**, *10*, 1179.

(12) Wang, Y.; Sukenik, S.; Davis, C. M.; Gruebele, M. Cell Volume Controls Protein Stability and Compactness of the Unfolded State. *J. Phys. Chem. B* **2018**, *122*, 11762–11770.

(13) Monteith, W. B.; Cohen, R. D.; Smith, A. E.; Guzman-Cisneros, E.; Pielak, G. J. Quinary

Structure Modulates Protein Stability in Cells. *Proc. Natl. Acad. Sci.* **2015**, *112*, 1739–1742.

(14) Wang, Q.; Zhuravleva, A.; Gierasch, L. M. Exploring Weak, Transient Protein--Protein Interactions in Crowded in Vivo Environments by in-Cell Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **2011**, *50*, 9225–9236.

(15) Guo, M.; Gelman, H.; Gruebele, M. Coupled Protein Diffusion and Folding in the Cell. *PLoS One* **2014**, *9*, e113040.

(16) Bicout, D. J.; Field, M. J. Stochastic Dynamics Simulations of Macromolecular Diffusion in a Model of the Cytoplasm of Escherichia Coli. *J. Phys. Chem.* **1996**, *100*, 2489–2497.

(17) Ridgway, D.; Broderick, G.; Lopez-Campistrous, A.; Ru'aini, M.; Winter, P.; Hamilton, M.; Boulanger, P.; Kovalenko, A.; Ellison, M. J. Coarse-Grained Molecular Simulation of Diffusion and Reaction Kinetics in a Crowded Virtual Cytoplasm. *Biophys. J.* **2008**, *94*, 3748–3759.

(18) McGuffee, S. R.; Elcock, A. H. Diffusion, Crowding & Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm. *PLoS Comput. Biol.* **2010**, *6*, e1000694.

(19) 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.

(20) Mereghetti, P.; Gabdoulline, R. R.; Wade, R. C. Brownian Dynamics Simulation of Protein Solutions: Structural and Dynamical Properties. *Biophys. J.* **2010**, *99*, 3782–3791.

(21) Mereghetti, P.; Wade, R. C. Atomic Detail Brownian Dynamics Simulations of Concentrated Protein Solutions with a Mean Field Treatment of Hydrodynamic Interactions. *J. Phys. Chem. B* **2012**, *116*, 8523–8533.

(22) Wang, Q.; Cheung, M. S. A Physics-Based Approach of Coarse-Graining the Cytoplasm of Escherichia Coli (CGCYTO). *Biophys. J.* **2012**, *102*, 2353–2361.

(23) Cossins, B. P.; Jacobson, M. P.; Guallar, V. A New View of the Bacterial Cytosol Environment. *PLoS Comput. Biol.* **2011**, *7*, e1002066.

(24) von Bülow, S.; Siggel, M.; Linke, M.; Hummer, G. Dynamic Cluster Formation Determines Viscosity and Diffusion in Dense Protein Solutions. *Proc. Natl. Acad. Sci.* **2019**, *116*, 9843–9852.

(25) Petrov, D.; Zagrovic, B. Are Current Atomistic Force Fields Accurate Enough to Study Proteins in Crowded Environments? *PLoS Comput. Biol.* **2014**, *10*, e1003638.

(26) 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*, e19274.

(27) Piana, S.; Sarkar, K.; Lindorff-Larsen, K.; Guo, M.; Gruebele, M.; Shaw, D. E. Computational Design and Experimental Testing of the Fastest-Folding Beta-Sheet Protein. *J. Mol. Biol.* **2011**, *405*, 43–48.

(28) Bennett, B. D.; Kimball, E. H.; Gao, M.; Osterhout, R.; Van Dien, S. J.; Rabinowitz, J. D. Absolute Metabolite Concentrations and Implied Enzyme Active Site Occupancy in Escherichia Coli. *Nat. Chem. Biol.* **2009**, *5*, 593–599.

(29) Shaw, D. E.; Grossman, J. P.; Bank, J. A.; Batson, B.; Butts, J. A.; Chao, J. C.; Deneroff, M. M.; Dror, R. O.; Even, A.; Fenton, C. H.; et al. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. In *International Conference for High Performance Computing, Networking, Storage and Analysis, SC*; 2014; 41–53.

(30) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100*, L47–L49.

(31) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14*, 71–73.

(32) Yoo, J.; Aksimentiev, A. New Tricks for Old Dogs: Improving the Accuracy of Biomolecular Force Fields by Pair-Specific Corrections to Non-Bonded Interactions. *Phys. Chem. Chem. Phys.* **2018**, *20*, 8432–8449.

(33) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E4758–E4766.

(34) Janin, J.; Rodier, F. Protein-Protein Interaction at Crystal Contacts. *Proteins Struct. Funct. Genet.* **1995**, *23*, 580–587.

(35) Smith, A. E.; Zhou, L. Z.; Gorensek, A. H.; Senske, M.; Pielak, G. J. In-Cell Thermodynamics and a New Role for Protein Surfaces. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 1725–1730.

(36) Dhar, A.; Girdhar, K.; Singh, D.; Gelman, H.; Ebbinghaus, S.; Gruebele, M. Protein

Stability and Folding Kinetics in the Nucleus and Endoplasmic Reticulum of Eucaryotic Cells. *Biophys. J.* **2011**, *101*, 421–430.

(37) Parry, B. R.; Surovtsev, I. V.; Cabeen, M. T.; O'Hern, C. S.; Dufresne, E. R.; Jacobs-Wagner, C. The Bacterial Cytoplasm Has Glass-like Properties and Is Fluidized by Metabolic Activity. *Cell* **2014**, *156*, 183–194.

(38) Nenninger, A.; Mastroianni, G.; Mullineaux, C. W. Size Dependence of Protein Diffusion in the Cytoplasm of Escherichia Coli. *J. Bacteriol.* **2010**, *192*, 4535–4540.

(39) Yeh, I.-C.; Hummer, G. System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. *J. Phys. Chem. B* **2004**, *108*, 15873–15879.

(40) Heo, M.; Maslov, S.; Shakhnovich, E. Topology of Protein Interaction Network Shapes Protein Abundances and Strengths of Their Functional and Nonspecific Interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (10), 4258–4263.

(41) Johnson, M. E.; Hummer, G. Nonspecific Binding Limits the Number of Proteins in a Cell and Shapes Their Interaction Networks. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (2), 603–608.

(42) Levy, E. D.; De, S.; Teichmann, S. A. Cellular Crowding Imposes Global Constraints on the Chemistry and Evolution of Proteomes. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (50), 20461–20466.

(43) Peleg, O.; Choi, J.-M.; Shakhnovich, E. I. Evolution of Specificity in Protein-Protein Interactions. *Biophys. J.* **2014**, *107* (7), 1686–1696.

(44) Ervin, J.; Sabelko, J.; Gruebele, M. Submicrosecond Real-Time Fluorescence Sampling: Application to Protein Folding. *J. Photochem. Photobiol. B Biol.* **2000**, *54*, 1–15.

(45) Reed, W. J.; Hughes, B. D. From Gene Families and Genera to Incomes and Internet File Sizes: Why Power Laws Are so Common in Nature. *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* **2002**, *66*, 067103.

(46) Bracewell, R. N. Fourier and Other Mathematical Transforms. In *Digital Encyclopedia of Applied Physics*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2009; pp 1149–1175.

(47) Southall, N. T.; Dill, K. A.; Haymet, A. D. J. A View of the Hydrophobic Effect. *J. Phys. Chem. B* **2001**, *106*, 521–533.

(48) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113–5123.

(49) Nawrocki, G.; Karaboga, A.; Sugita, Y.; Feig, M. Effect of Protein–Protein Interactions and Solvent Viscosity on the Rotational Diffusion of Proteins in Crowded Environments. *Phys. Chem. Chem. Phys.* **2019**, *21*, 876–883.

(50) Yang, W. Y.; Gruebele, M. Folding λ-Repressor at Its Speed Limit. *Biophys. J.* **2004**, *87*, 596–608.

(51) Hagen, S. J.; Hofrichter, J.; Szabo, A.; Eaton, W. A.; Wipf, P.; Kelly, J. W.; Gruebele, M. Diffusion-Limited Contact Formation in Unfolded Cytochrome c: Estimating the Maximum Rate of Protein Folding. *Proc. Natl. Acad. Sci.* **1996**, *93*, 11615–11617.

(52) Liu, F.; Du, D.; Fuller, A. A.; Davoren, J. E.; Wipf, P.; Kelly, J. W.; Gruebele, M. An Experimental Survey of the Transition between Two-State and Downhill Protein Folding Scenarios. *Proc. Natl. Acad. Sci.* **2008**, *105*, 2369–2374.