Submitted to: JPC B

# Crowding, sticking and partial folding of GTT WW domain in a small cytoplasm model

M. M. Rickard,[†] Y. Zhang,[§] T. V. Pogorelov[*,†,§,^,&] and M. Gruebele[*,†,§,‡,^]

[†]Department of Chemistry, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, USA

[§]Center for Biophysics and Quantitative Biology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, USA

[‡]Department of Physics, University of Illinois, at Urbana−Champaign, Urbana, Illinois 61801, USA

[&]National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

[^]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

ABSTRACT: Recent experimental data has shown that protein folding in the cytoplasm can differ from *in vitro* folding with respect to speed, stability, and residual structure. Here we investigate the all-atom molecular dynamics (MD) simulations of 9 copies of the model protein GTT WW domain in a small bacterial cytoplasm model using three force fields. GTT has been well-studied by MD in aqueous solution for comparison. We find that folded copies remain folded for up 25 μs, whereas unfolded copies do not fold for up to 190 μs. Unfolded GTT in our cytoplasm model does populate partly folded intermediates with one of the two hairpins formed. Relative to aqueous solution, GTT gets stuck in metastable states with small RMSD and radius of gyration and extensive burial of surface area against other macromolecules. In particular, GTT is even able to form transient inter-molecular beta sheets with other proteins, resulting in 'chimeric structure' that could be a precursor to oligomeric β-aggregates. We conclude that sticking, enhanced by the non-native mutations of GTT, is largely responsible and we propose, based on our result as well as recent experiments, that co-evolution of protein surfaces with their solvation environment is important for unhindered folding and diffusion of proteins in the cytoplasm.

## INTRODUCTION

*In vitro* experiments,[1] theory,[2] and computational models[3] have shown that protein folding free energy landscapes are shallow. That is, the free energy differences between folded and denatured states and activation free energies are relatively small (0-40 kJ/mole) over the physiological temperature range. Consequently most proteins fold relatively quickly (microseconds to minutes),[4] and their stability is susceptible to changes in solvation conditions including tuning of the temperature,[5] pressure,[6] crowding,[7] ionic strength, acidity,[8] chao-/kosmotopes[9,10] and other solvation parameters.

The native solvation environment for many proteins, the cytoplasm, is no exception to this rule, and experiments have reported cell type,[11] cell cycle,[12] and site-to-site dependence[13] of protein stability and folding kinetics in the cytoplasm, although the general folding mechanism remains unchanged. In cells, proteins are subject to crowding (excluded volume)[14] and sticking interactions[15] in addition to quinary structure (weak functional interactions).[16] For example, a human homolog of a bacterial protein sticks in bacterial cells, while the bacterial protein diffuses more freely, due to differences in electrostatics and surface hydrophobicity.[17]

Here, we examine how a small (~16 proteins, ~262 small solutes, ~430 ions and ~50,000 water molecules) all-atom model of the *E. coli* cytoplasm[18] affects the folding of the 33-residue GTT WW domain compared to similar simulations of GTT in aqueous solution.[19] The now > 200 μs simulation occupies a 'protein-folding niche' between long-time Brownian dynamics simulations[20] and short-time all-atom simulations of larger cytoplasmic models.[21]

GTT is a triple-stranded beta sheet protein derived from the FiP mutant of the signaling module WW domain,[22] with hairpin 1 formed by strands 1+2, and hairpin 2 formed by strands 2+3 (**Fig. 1**). In *in vitro* experiments, GTT folds in ~4.3 μs (at 353 K) and slightly faster using the CHARMM22* force field, via a mechanism that forms hairpin 1 first *ca*. two-thirds of the time, and hairpin 2 *ca.* one-third of the time.[23] Thus GTT is an ideal testbed for simulating the effect of cytoplasmic crowding and sticking on folding dynamics.

We simulated two initially folded and seven initially unfolded copies of GTT in our cytoplasm model using three versions of the CHARMM force field: CHARMM22* (C22*),[24] CHARMM36m (C36m),[25] and CHARMM36m with CUFIX corrections (C36mCU).[26] These force fields were selected due to their propensity to optimize folding (C22*) or reduce 'sticking' interactions between the protein and its surrounding environment (C36m and C36mCU). In all simulations, the

folded copies remained folded. None of the unfolded copies completely folded despite an overall simulation time at least twenty times longer than the expected folding time at the temperature, although native-like structure formed in hairpins 1 and 2 several times during the simulation. Although the optimization of current force fields for folding of isolated small proteins leads them to overestimate macromolecular sticking, a number of general questions can be asked. Here we focus on three: How do crowding and sticking slow down folding dynamics of unfolded GTT and affect fluctuations of the GTT native state? How disruptive can sticking interactions in the cytoplasm become? And what is the role of in-cell water in mediating native-like contact formation?

## METHODS

**Model Construction and Simulations** The cytoplasm model and choice of proteins, ions and metabolites has been described previously.[18] Here we provide a brief summary specific to GTT WW domain and additional analysis information.

All GTT models were constructed by mutating an existing crystal structure (PDB ID: 2F21) at residues N26G, A27T and S28T. Initial unfolded structures were obtained with temperature-jump molecular dynamics simulations using NAMD 2.9.[27] GTT was subjected to equilibrium MD with C22* force field[24] in a 5x5x5 $nm^3$ box with TIP3P water[28] and 0.15 M NaCl, in the NPT ensemble. The Langevin piston Nosé−Hoover procedure were used to maintain temperature and pressure at constant values (500 K and 1 atm).[29,30] The long-range electrostatic forces were calculated without truncation with the particle mesh Ewald (PME) method,[31] using a grid density of 1 $Å^{-3}$. The final structures were equilibrated at 318 K before being placed in the cytoplasm models. The simulation of GTT alone was run in the identical cube of water with a size of 5 nm initially used to obtain the unfolded GTT structures.

The C22* cytoplasm model contained three unfolded copies of GTT, while the C36m and C36mCU models contained 1 folded and 2 unfolded copies of GTT,[25,26] each. Additionally, the C22* model contained a small disordered DNA binding domain (Hns) that was omitted from the C36m and C36mCU models. For the cytoplasm model, initially the macromolecules were simulated with Atomic Resolution Brownian Dynamics to optimally pack into the periodic cell.[32] Biological levels of common metabolites and inorganic ions were added[33] and the models were

solvated with TIP3P water. Models were initially equilibrated locally using NAMD 2.9 using similar parameters as described above before production runs on the special-purpose supercomputer for MD simulations, Anton 2.[34] Altogether, the simulations required ~24 machine-days on Anton 2 (at ~10 μs/day).

**Trajectory Analysis** The secondary structure of GTT was calculated every 24 ns using the compute_dssp function in MDTraj.[35,36] RMSD values were calculated using VMD's RMSD trajectory tool.[37] To determine the buried surface area of GTT, the solvent accessible surface area (SASA) of the protein alone was compared to the SASA of the protein in the cytoplasmic environment. SASA calculations were performed in VMD with a radius of the sphere set to the radius of a water molecule set, 1.4 Å.

## RESULTS AND DISCUSSION

**Overall simulation** We ran three simulations of *ca*. 200,000 atoms at 318 K, with 21% of the volume excluded by macromolecules (corresponds to 300 mg/ml macromolecules). The CHARMM22* (C22* for short) simulation of ≈190 μs started with three unfolded states of GTT, the CHARMM36m (C36m) and CHARMM36mCU (C36mCU) simulations of ≈25 μs with one folded and two unfolded GTT WW domains each. When referring to GGT molecules from the same in-cell simulation, we will number them GTT1, GTT2 and GTT3 to distinguish them in the discussion. In none of the simulations did GTT completely fold, although formation of paired strands 1-2 or 2-3 was observed at least 5 times (Figures S1-S3). The two folded WW domains also did not unfold for the 25 μs duration of their simulations. The stability of other proteins in the simulation varied, with $C_\alpha$-RMSD ranging from about 2 to 10 Å, and coldshock proteins (Csp) particularly rigid with small RMSD (Figure S4).

Three typical structures of GTT are shown in a snapshot of the C36mCU simulation in Fig. 1 (WW domains in orange): native (top), highly extended (middle) and non-native helical residual structure (bottom). A clustering analysis in Figure S5 of structures for the three unfolded WW domains in the longest (C22*) simulation shows that unfolded states with transient helical structure near the N-terminus are very prevalent in that force field. They were also observed in aqueous solution simulations where GTT folded (see below). Although this propensity may be exaggerated

by C22*, it is plausible based on the very high helical propensity as evidence by Agadir[38] scores of residues W11 through M15 (numbering as in Figure 5G, scores in Table S1).

**GTT in aqueous solution** A long simulation of GTT in aqueous solution on Anton 2 using C22* has previously shown that GTT folds in 10±2 μs at 395 K, near the melting temperature of GTT with C22*-TIP3P.[19] This is similar to the experimentally measured mean refolding time of ≈6 μs at 363 K.[19] To check that a similar time scale can be obtained at our lower simulation temperature of 318 K, well below $T_m$, we simulated a 25 μs trajectory using C22* (Figure 2). GTT showed large (range 5-20 Å) RMSD fluctuations in the unfolded state, and after populating a short-lived intermediate with a non-native hairpin 1 at 9 μs, it formed hairpin 2 and within 2 μs the native state at 19 μs. Thus GTT can fold on a time scale almost as fast as in the higher temperature simulations, consistent with the prediction of faster pre-factors for WW domain offsetting higher barriers at low temperature.[39]

The formation of transient helical structure itself is not inconsistent with complete refolding. Figure 2 reveals the formation and melting of non-native helix multiple times, generally for < 1 μs. The N-terminal beta-strand is most susceptible to forming non-native helix, as seen in the overall simulation of the cytoplasm model.

**Secondary structure of GTT in the cytoplasm model** We begin by examining the dynamics of GTT in the cytoplasm model via secondary structure. Figure 3 highlights one 5 μs region in the trajectories for each GTT WW domain where significant (if any) changes in secondary structure were observed. The full secondary structure trajectories for the C22*, C36m and C36mCU simulations are shown in SI Figures S1-S3. In four of the seven non-native WW domains non-helical structure persisted throughout the entire simulation, in contrast to Figure 2 and previous simulations of WW domain subject to pressure or temperature perturbations, where transient helical structure lasts for less than a few μs.[40] As expected from the secondary structure propensity using Agadir,[38] all but one of the four GTTs with helical structure throughout show long-lived helix near the N-terminus. Therefore, the cytoplasmic environment enhances helix propensity of WW domain relative to aqueous solution, and this contributes to the reduced folding rate.

The region of turns 1 and 2 is slightly less likely to form non-native helical structure than the β-sheet regions. This is particularly evident in Figure S2 of GTT2 and GTT3 in the C36m

simulation, where turn 1 (residues 16-20) frequently switches from helical to random coil or turn structure, whereas the adjacent sheet regions remain helical. This is consistent with nucleation of the native state occurring at turns, and not in the beta strands, as will be discussed later in conjunction with desolvation during turn formation.

**Compactness of GTT in aqueous solution *vs.* cytoplasm** In the aqueous solution simulation, GTT rapidly explores radii of gyration greater than 20 Å and then swiftly collapses to near-native radius of gyration (Figure 4). This propensity for fast collapse to compact states has been seen in single molecule experiments on unfolded proteins under native conditions.[41] In contrast, one might expect that GTT rarely explores states of high radius of gyration in the crowded cytoplasmic simulations. This is indeed the case, with GTT frequently getting stuck in states with $R_g$<15 Å. (Figure 4).

Two main differences in the dynamics of unfolded GTT in the cytoplasm and aqueous solution are evident. First of all, whereas large fluctuations occur in aqueous solution on a timescale < 100 ns, in the three C22* cytoplasmic copies the protein frequently spends many microseconds in a state with a narrow range of radius of gyration, before switching to another such state. The residence time of GTT3 in Figure 4 for example is greater than 25 μs in three cases, and GTT2 gets stuck in a state with nearly constant $R_g$ for over 70 μs. GTT1 experiences a highly reduced dynamic range of $R_g$ and several long residence times also. The slow time evolution of $R_g$ goes hand in hand with reduced translational diffusion through the cytoplasm, averaging 3.0 $nm^2/\mu s$ in aqueous solution *vs.* 0.03 $nm^2/\mu s$ for the cytoplasmic copies (Figure S6), as determined by fitting the slope of the mean squared displacements *vs.* time of the MD simulation trajectories. Previous experimental measurements in *E. coli* show a slowdown closer to a factor of 10 than a factor of 100,[42,43] indicating that our cytoplasm models overestimate the stickiness of the cytoplasm for GTT.

Secondly, the cytoplasmic GTTs have far less solvent-exposed surface area (Figure 4, bottom) in the C22* simulation, routinely falling below the native state average of 7.5 $nm^2$ in aqueous solution. This observation is consistent with reduced diffusion due to macromolecular sticking and crowding. Both super-resolution displacement mapping of mammalian cells[44] and bleaching measurements in mammalian cells[45] have revealed heterogeneous diffusion consistent with the intermittent sticking and surface area burial seen in the bacterial cytoplasm model in Figure 4.

**Sticking of GTT to nearby macromolecules** The retarded dynamics and lack of exposed surface area of GTT in the cytoplasmic simulations are strong evidence that GTT is not only crowded (excluded volume), but sticks (forms weak surface bonds) to other proteins. We analyzed such sticking previously for a series of proteins.[18] Here we consider two examples of how GTT interacts with the surrounding cytoplasmic matrix.

Figure 5A shows GTT3 from the C36mCU simulation buried in a matrix of four other cytoplasmic proteins. Another copy's (GTT2) surface area also is buried by other proteins (Figure 5B) and varies widely from 20 $nm^2$ to near-zero, showing that the simulation is long enough to sample release and reburial on a time scale of several microseconds. A somewhat shorter time scale for lifetime of protein-protein contacts, following a power law distribution with a broad range from 1 ns to 1 μs has been observed for an average over many fully folded proteins.[18] As discussed above, RMSD is able to range to its highest values only when the protein is less buried, so the buried states tend to be more compact. Figure 5B also decomposes the buried area into hydrophobic, charged and polar interactions. For the CHARMM force fields used here, hydrophobic, polar and charged patches make a similar contribution to GTT-protein contacts. For a larger protein analyzed previously hydrophobic contacts were more dominant.[18]

Figure 6 shows a contact analysis of GTT2 during the portion of the C22* trajectory when the protein is struck for over 70 μs to proteins MDN and EF-TU in Figure 4 (top). During this time its radius of gyration remains frozen at ≈12 Å with fluctuations about 5 times smaller than for the unfolded state in aqueous solution. The reduced $R_g$ is caused by tight contacts between GTT and its host proteins. In particular salt bridges between lysine-glutamate and glutamate-arginine pairs are long lived and often coexist for microseconds 3-4 at a time. Hydrogen bonds are more fleeting, and often act in concert with salt bridges, as can be seen by the strong correlations in Figure 6 (coincidence of red bars). Three hydrophobic interactions between GTT and EF-TU are also particularly prominent, and also correlated with salt bride dynamics. This cooperativity between three sticking modes explains the long-lived nature of this metastable state of GTT. It has been observed previously that MD force fields, including C22*, overstabilize the formation of salt bridges between charged side chains.[46,47] Although this will bias the protein-protein contact formation we observe, other computational studies of proteins in crowded environments also point to the importance of salt bridge formation in protein-protein contacts.[48,49] Additionally,

experimental evidence underscores the critical role protein surface charge plays in protein-protein interactions *in vivo*.[17]

**Chimeric interaction between GTT and other proteins** Cooperativity between interacting proteins in our model cytoplasm can be even more extreme: the intermingling of unfolded GTT with another protein to form intramolecular tertiary structure, which we will refer to as a 'chimeric protein,' in analogy to genetically chimeric proteins where proteins originating from different genes are linked.[50] Such chimeras have been observed experimentally for the WW domain, when monomers are linked together so they can interact frequently, as would be the case with proteins in the cytoplasm.[51]

Figure 7 shows a specific example where GTT1 interacts with superoxide dismutase (SodA) in the C22* simulation. In (b), the proteins make contact and a long helix in SodA partly unfolds. In (c), GTT has diffused away again. In (d), GTT and the unfolded region interact again to form an intermolecular beta sheet composed of two strands (shown in orange and cyan). In (d), they have diffused away again and the beta sheet structure is lost. Finally, in (f), they associate again and the same intermolecular β-sheet strand forms again. Overall, an initially transiently populated helical region in one protein (GTT) interacts with part of a helix in another protein (SodA) to form beta sheet structure. The β-sheet structure is very dynamical and lasted anywhere from several ns to several μs. It was mainly governed by hydrogen bonds between polar residues 30 to 35 in GTT1 (black arrows in Figure 6G), with an Arginine-rich region in SodA (Figure 6H). This happens to be the region of the GTT mutation that distinguishes GTT from wild-type human Pin1 WW domain.[22]

Another such interaction occurred in the C22* simulation between Mdh and GTT2. Thus, chimeric interactions are not rare in this force field. We did not observe them in the albeit shorter C36m and C36mCU simulation (50 μs total), indicating that these force fields may be less sticky. However, such events may present an extreme form of protein-protein interaction in the cytoplasm, and such non-native intermolecular beta sheet structure could eventually resemble pre-amyloid oligomers.

**Drying transition of GTT during partial folding in cytoplasm model** Hairpin formation between either strand 1 and 2, or 2 and 3, was observed several times throughout the three

simulations. Figure 8 illustrates an example of the forward process for loop 2 formation (desolvation and hairpin formation) from the C36mCU simulation in Figure 3c, and Figure S7 illustrates another example of the reverse process (solvation to open the hairpin).

Among the three cases of GTT2's partial hairpin formation and break-up in the C36mCU model, some are more native-like, while others are one residue out of registration. In all cases, we observe a state with a single water molecule intercalating the loop, either just before the 'dry' hairpin forms, or just before additional water molecules solvate the loop while it breaks up. Five samples are not enough for a full transition path analysis, but we propose that the loop intercalated by a single water molecule lies close to the transition state ensemble for loop formation. In all cases the less solvated loop forms before the β-strands interact, and a drying transition that removes the final intercalating water molecule allows the two-stranded beta sheet to form. This 'loop first, then zip up the sheet' is consistent with *in vitro* measurements and models that residues near the turn have the largest effect on folding/unfolding kinetics.[52,53]

## CONCLUSIONS

We find that the C22*, C36m and C36mCU force fields all reduce the in-cell folding rate of GTT WW domain, at least by a factor of *ca.* 10 for C22* relative to aqueous solution. The main reason is sticking of the unfolded protein against other macromolecules in the simulations. In these force fields, the contribution of hydrophobic interactions at protein-protein interfaces is significant. Slow refolding is corroborated by slow diffusion, slow intermittent changes in radius of gyration and surface burial, formation of salt bridges, H-bonds and hydrophobic contacts between proteins, and even formation of chimeric structures. Experimental observations of a few larger proteins have revealed in-cell folding rates comparable to *in vitro*,[54] so the force fields are likely to exacerbate sticking. This should not come as a surprise because they were modified from C22* to simulate folding of a single protein in aqueous solution, where enhanced sticking of a protein (to itself) may enhance collapse and folding. Although sticking in the cell is likely not as severe as in the simulations, sticking will slow down in-cell folding, whereas crowding reduces available volume only by a factor of ca. 2, and has a similarly small effect on folding *in vitro* vs. in-cell or *in vivo*.[11,55]

Another explanation for the sticking of GTT is that the region of the three mutations made to speed up folding (residues 30-32 in Figure 7h) is strongly involved in chimeric structure between

GTT and SodA. Thus, mutating surface-exposed residues away from the wild-type sequence could enhance sticking. Two types of experimental results support the importance of surface residues to avoid in-cell sticking, and hence co-evolution of protein surfaces with their cellular environment.[56] Firstly, in-cell NMR studies have shown that protein surface composition of human *vs.* bacterial proteins strongly influences free *vs.* hindered rotational diffusion in bacterial cells.[17] Secondly, comparison of an intracellular eukaryotic protein (PGK) *vs.* an extracellular prokaryotic protein (VlSE) shows that the latter folds more slowly in mammalian cells.

Finally, it is interesting to note that formation of chimeric proteins with an intermolecular β-sheet (Figure 8) is a mechanism by which an unfolded protein, here GTT, can initiate formation of β-rich protein oligomers. Such strong sticking could eventually become a nucleus for in-cell formation of amyloid structure, although we did not observe that on the time scale of our simulations.

## ASSOCIATED CONTENT

### Supporting Information

The PDF file describes the full simulations available with text and figures. This material is available free of charge via the Internet at http://pubs.acs.org. Structural files for each simulation available from Anton2 (PSC) provide regular snapshots of the MD simulation.

## AUTHOR INFORMATION

Corresponding Author

E-mail: mgruebel@illinois.edu

**Notes**

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors gladly acknowledge supercomputer time provided on Anton 2 by the Pittsburgh Supercomputing Center (PSC), Allocation Grant MCB100024P. The Anton 2 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 NSF Grant MCB 1803786.

ABBREVIATIONS

VMD,  Visual Molecular Dynamics; CHARMM, Chemistry at Harvard Molecular Mechanics;

REFERENCES


(1) Yang, W. Y.; Gruebele, M. Folding at the Speed Limit. *Nature* **2003**, *423*, 193–197.

(2) Onuchic, J. N.; Wolynes, P. G.; Luthey-Schulten, Z.; Socci, N. D. Toward an Outline of the Topography of a Realistic Protein-Folding Funnel. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 3626–3630.

(3) Freddolino, P. L.; Park, S.; Roux, B.; Schulten, K. Force Field Bias in Protein Folding Simulations. *Biophys. J.* **2009**, *96*, 3772–3780.

(4) Nölting, B.; Golbik, R.; Fersht, A. R. Submillisecond Events in Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 10668–10672.

(5) Griko, Y. V.; Privalov, P. L. Calorimetric Study of the Heat and Cold Denaturation of Beta-Lactoglobulin. *Biochemistry* **1992**, *31*, 8810–8815.

(6) Hawley, S. A. Reversible Pressure–Temperature Denaturation of Chymotrypsinogen. *Biochemistry* **1971**, *10*, 2436–2442.

(7) Zimmerman, S. B.; Minton, A. P. Macromolecular Crowding: Biochemical, Biophysical, and Physiological Consequences. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 27–65.

(8) Brazhnikov, E. V.; Chirgadze YuN, null; Dolgikh, D. A.; Ptitsyn, O. B. Noncooperative Temperature Melting of a Globular Protein without Specific Tertiary Structure: Acid Form of Bovine Carbonic Anhydrase B. *Biopolymers* **1985**, *24*, 1899–1907.

(9) Putnam, F. W.; Erickson, J. O.; Volkin, E.; Neurath, H. Native and Regenerated Bovine Albumin : I. Preparation and Physicochemical Properties. *J. Gen. Physiol.* **1943**, *26*, 513–531.

(10) Washabaugh, M. W.; Collins, K. D. The Systematic Characterization by Aqueous Column Chromatography of Solutes Which Affect Protein Stability. *J. Biol. Chem.* **1986**, *261*, 12477–12485.

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

(12) Wirth, A. J.; Liu, Y.; Prigozhin, M. B.; Schulten, K.; Gruebele, M. Comparing Fast Pressure Jump and Temperature Jump Protein Folding Experiments and Simulations. *J. Am. Chem. Soc.* **2015**, *137*, 7152–7159.

(13) Ebbinghaus, S.; Gruebele, M. Protein Folding Landscapes in the Living Cell. *J. Phys. Chem. Lett.* **2011**, *2*, 314–319.

(14) van den Berg, B.; Ellis, R. J.; Dobson, C. M. Effects of Macromolecular Crowding on Protein Folding and Aggregation. *EMBO J.* **1999**, *18*, 6927–6933.

(15) Chen, T.; Dave, K.; Gruebele, M. Pressure- and Heat-Induced Protein Unfolding in Bacterial Cells : Crowding vs . Sticking. *FEBS Lett.* **2018**, *592*, 1357–1365.

(16) McConkey, E. H. Molecular Evolution, Intracellular Organization, and the Quinary Structure of Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1982**, *79*, 3236–3240.

(17) 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.

(18) Rickard, M. M.; Zhang, Y.; Gruebele, M.; Pogorelov, T. V. In-Cell Protein–Protein Contacts: Transient Interactions in the Crowd. *J. Phys. Chem. Lett.* **2019**, *10*, 5667–5673.

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

(20) 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.

(21) 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.

(22) Jäger, M.; Zhang, Y.; Bieschke, J.; Nguyen, H.; Dendle, M.; Bowman, M. E.; Noel, J. P.; Gruebele, M.; Kelly, J. W. Structure-Function-Folding Relationship in a WW Domain. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 10648–10653.

(23) Davis, C. M.; Zanetti-Polzi, L.; Gruebele, M.; Amadei, A.; Dyer, R. B.; Daidone, I. A Quantitative Connection of Experimental and Simulated Folding Landscapes by Vibrational Spectroscopy. *Chem. Sci.* **2018**, *9*, 9002–9011.

(24) 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.

(25) 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.

(26) 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.

(27) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(28) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(29) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103*, 4613–4621.

(30) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189.

(31) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(32) Yoo, J.; Aksimentiev, A. Competitive Binding of Cations to Duplex DNA Revealed through Molecular Dynamics Simulations. *J. Phys. Chem. B* **2012**, *116*, 12946–12954.

(33) Bennett, B. B. D.; Kimball, E. H. E.; 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.

(34) 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.; Forte, A.; Gagliardo, J.; Gill, G.; Greskamp, B.; Ho, C. R.; Ierardi, D. J.; Iserovich, L.; Kuskin, J. S.; Larson, R. H.; Layman, T.; Lee, L.-S.; Lerer, A. K.; Li, C.; Killebrew, D.; Mackenzie, K. M.; Mok, S. Y.-H.; Moraes, M. A.; Mueller, R.; Nociolo, L. J.; Peticolas, J. L.; Quan, T.; Ramot, D.; Salmon, J. K.; Scarpazza, D. P.; Schafer, U. B.; Siddique, N.; Snyder, C. W.; Spengler, J.; Tang, P. T. P.; Theobald, M.; Toma, H.; Towles, B.; Vitale, B.; Wang, S. C.; Young, C. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. In *SC14: International Conference for High Performance Computing, Networking, Storage and Analysis*; IEEE: New Orleans, LA, USA, **2014**; pp 41–53.

(35) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528–1532.

(36) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577–2637.

(37) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(38) Lacroix, E.; Viguera, A. R.; Serrano, L. Elucidating the Folding Problem of Alpha-Helices: Local Motifs, Long-Range Electrostatics, Ionic-Strength Dependence and Prediction of NMR Parameters. *J. Mol. Biol.* **1998**, *284*, 173–191.

(39) Liu, F.; Nakaema, M.; Gruebele, M. The Transition State Transit Time of WW Domain Folding Is Controlled by Energy Landscape Roughness. *J. Chem. Phys.* **2009**, *131*,

(40) Wirth, A. J.; Platkov, M.; Gruebele, M. Temporal Variation of a Protein Folding Energy Landscape in the Cell. *J. Am. Chem. Soc.* **2013**, *135*, 19215–19221.

(41) Lipman, E. A. Single-Molecule Measurement of Protein Folding Kinetics. *Science* **2003**, *301*, 1233–1235.

(42) Elowitz, M. B.; Surette, M. G.; Wolf, P. E.; Stock, J. B.; Leibler, S. Protein Mobility in the Cytoplasm of Escherichia Coli. *J. Bacteriol.* **1999**, *181*, 197–203.

(43) Mullineaux, C. W.; Nenninger, A.; Ray, N.; Robinson, C. Diffusion of Green Fluorescent Protein in Three Cell Environments in Escherichia Coli. *J. Bacteriol.* **2006**, *188*, 3442–3448.

(44) Xiang, L.; Chen, K.; Yan, R.; Li, W.; Xu, K. Super-Resolution Displacement Mapping of Unbound Single Molecules Reveals Nanoscale Heterogeneities in Intracellular Diffusivity. *Nat. Methods* **2020**, *17*, 524-530.

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

(46) Debiec, K. T.; Gronenborn, A. M.; Chong, L. T. Evaluating the Strength of Salt Bridges: A Comparison of Current Biomolecular Force Fields. *J. Phys. Chem. B* **2014**, *118*, 6561–6569.

(47) Debiec, K. T.; Cerutti, D. S.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T. Further along the Road Less Traveled: AMBER Ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J. Chem. Theory Comput.* **2016**, *12* , 3926–3947.

(48) 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.

(49) Bülow, S. von; 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.

(50) Campbell, R. K.; Bergert, E. R.; Wang, Y.; Morris, J. C.; Moyle, W. R. Chimeric Proteins Can Exceed the Sum of Their Parts: Implications for Evolution and Protein Design. *Nat. Biotechnol.* **1997**, *15*, 439–443.

(51) Dave, K.; Gasic, A. G.; Cheung, M. S.; Gruebele, M. Competition of Individual Domain Folding with Inter-Domain Interaction in WW Domain Engineered Repeat Proteins. *Phys. Chem. Chem. Phys.* **2019**, *21*, 24393–24405.

(52) Muñoz, V.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. A Statistical Mechanical Model for Beta-Hairpin Kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 5872–5879.

(53) Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. Folding Dynamics and Mechanism of Beta-Hairpin Formation. *Nature* **1997**, *390*, 196–199.

(54) Gelman, H.; Wirth, A. J.; Gruebele, M. ReAsH as a Quantitative Probe of In-Cell Protein Dynamics. *Biochemistry* **2016**, *55*, 1968–1976.

(55) Ebbinghaus, S.; Dhar, A.; Douglas Mcdonald, J.; Gruebele, M. Protein Folding Stability and Dynamics Imaged in a Living Cell. *Nat. Publ. Group* **2010**, *7*, 319–319.

(56) Guin, D.; Gruebele, M. Weak Chemical Interactions That Drive Protein Evolution: Crowding, Sticking, and Quinary Structure in Folding and Function. *Chem. Rev.* **2019**, *119*, 10691–10717.

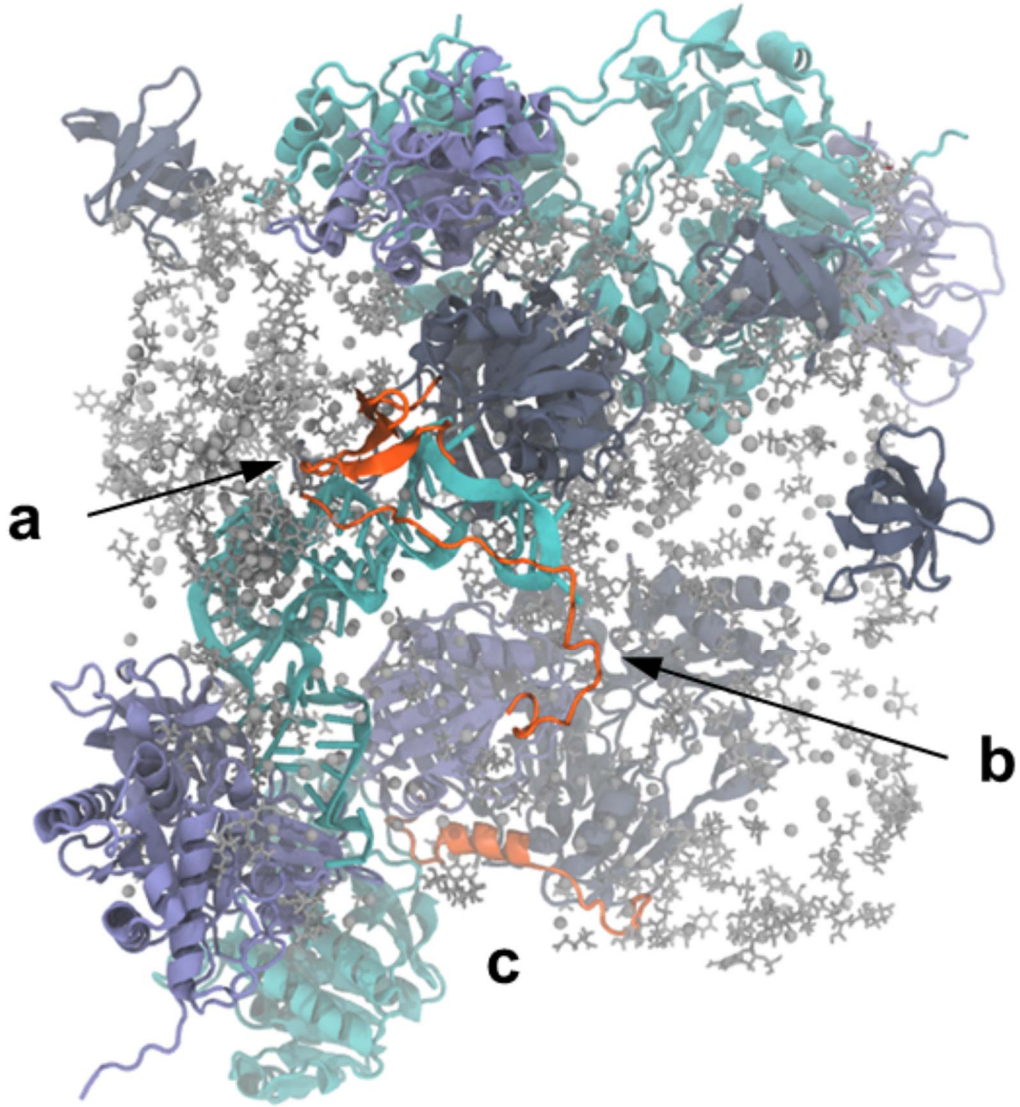


**Figure 1.** Overall view of the Charmm36mCU (C36mCU) simulation after ~4.99 μs of all-atom simulation, showing three states of GTT in shades of orange. (a) native structure with *Q=0.93*. (b) highly extended unfolded structure with *Q*=0.01. (c) unfolded with non-native helical structure, *Q*=0.03. Other macromolecules in the model are illustrated in blue/green and metabolites and inorganic ions are shown in gray.

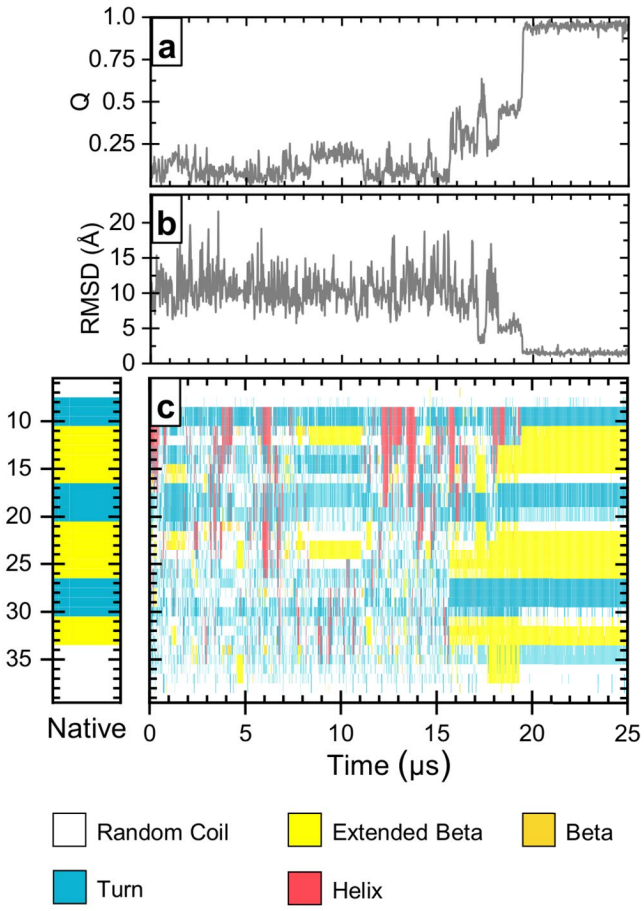


**Figure 2.** CHARMM22* (C22*) trajectory of GTT in aqueous solution at 318 K, showing a complete refolding event in less than 20 μs, consistent with ref. [19]. (A) Q > 0.85 in the folded state. (B) RMSD undergoes fluctuations between 5 and 20 Å before settling to <3 Å upon folding. (C) Secondary structure involves fluctuating helix especially between residues 3-10 (strand 1). Turn 1 forms temporarily at 8 μs, until folding is initiated by formation of turn 2 at >15 μs and subsequent formation of a native-like turn 1 at 20 μs.

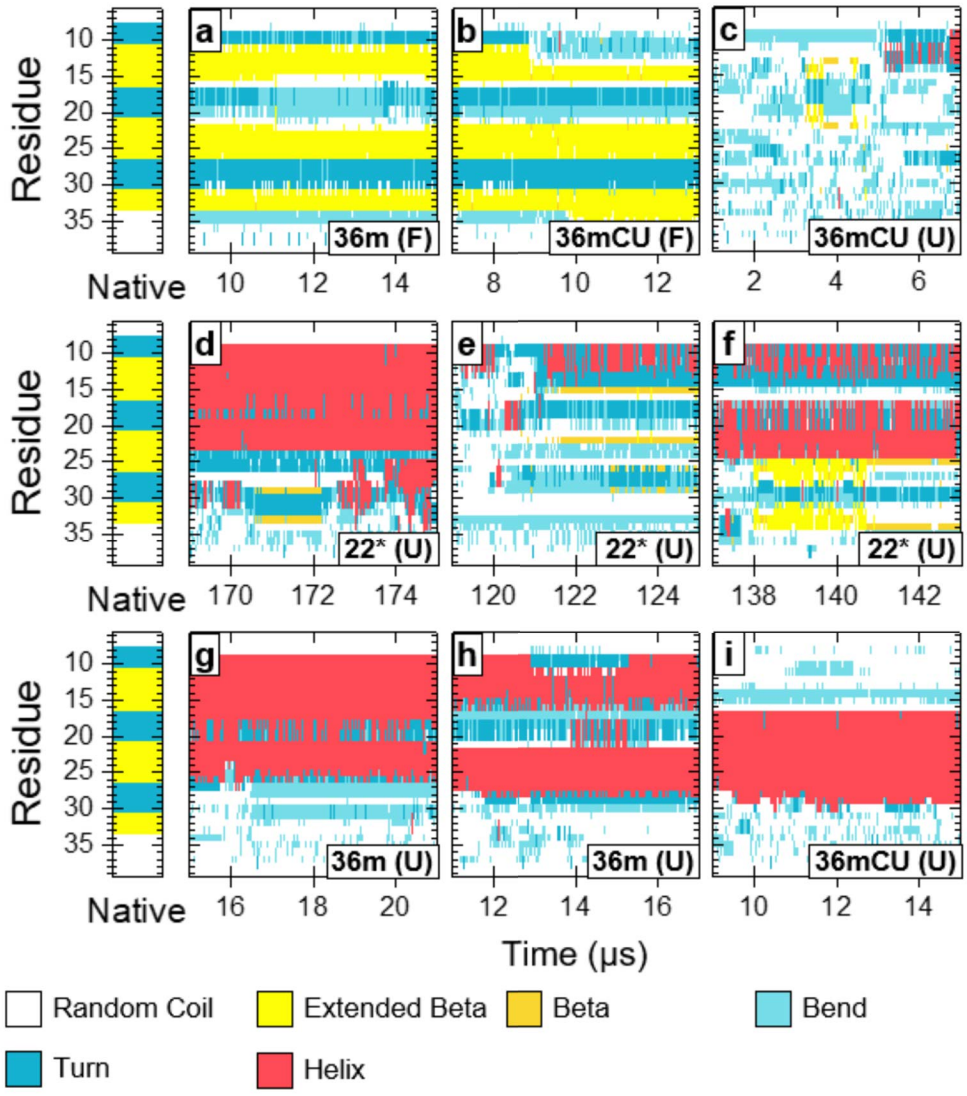


**Figure 3.** Secondary structure of GTT throughout trajectories as calculated by DSSP (reference DSSP, mdtraj). The secondary structure at each residue of the native structure is given at left. Each panel (a-i) represents a 6 μs slice of each GTT copy's trajectory. The lower right of each panel indicates the force field used to simulate and whether the protein started in the folded (F) or unfolded (U) state. Partial β hairpin formation was observed in several of the unfolded trajectories (c-f) but was frequently register shifted. Secondary structure calculations of the full trajectories can be found in Figures S1-S3.

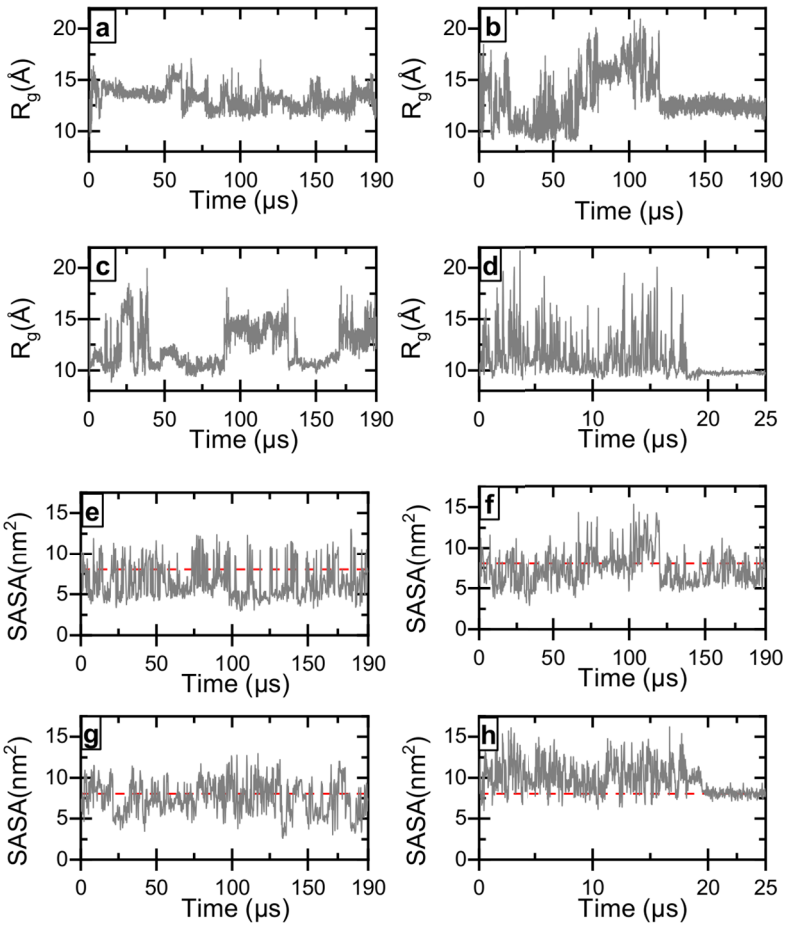


**Figure 4.** Top 4 panels: Radius of gyration $R_g$ of GTT in the CHARM22* (C22*) cytoplasm model (a-c**)** and in water (d). In water, the protein makes excursions between native-like $R_g$ and > 20 Å while in the unfolded state. In the cytoplasm model, $R_g$ fluctuates much less in the unfolded state (GTT1 in panel a), switches between long-lived states of similar compactness (GTT2 in panel b), and can end up in very long-lived metastable states due to sticking to the surface of other macromolecules (see Figures 5 and 6). Bottom 4 panels: in the aqueous solution C22* simulation, the exposed hydrophobic surface area of GTT (h) is almost always larger than in the native state (red dotted line). In the C22* cytoplasm model, unfolded GTT1-3 (e-g) most often exposes even less hydrophobic area than aqueous folded GTT due to burial against other macromolecules (see Figures 5 and 6 for examples).

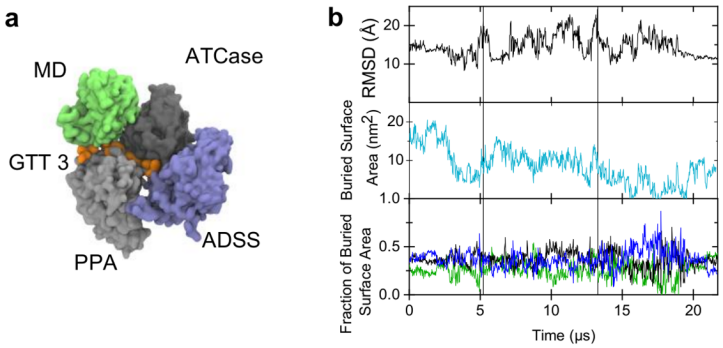


**Figure 5**. (a) Snapshot of GTT3 buried by surrounding macromolecules in the CHARMM36mCU (C36mCU) trajectory. (b) RMSD of GTT2 in the C36mCU trajectory, surface area of GTT that is buried by protein-protein contacts throughout the C36mCU trajectory. (c) contribution of different residue types: polar (green), hydrophobic (black), and charged (blue) to the buried surface area.

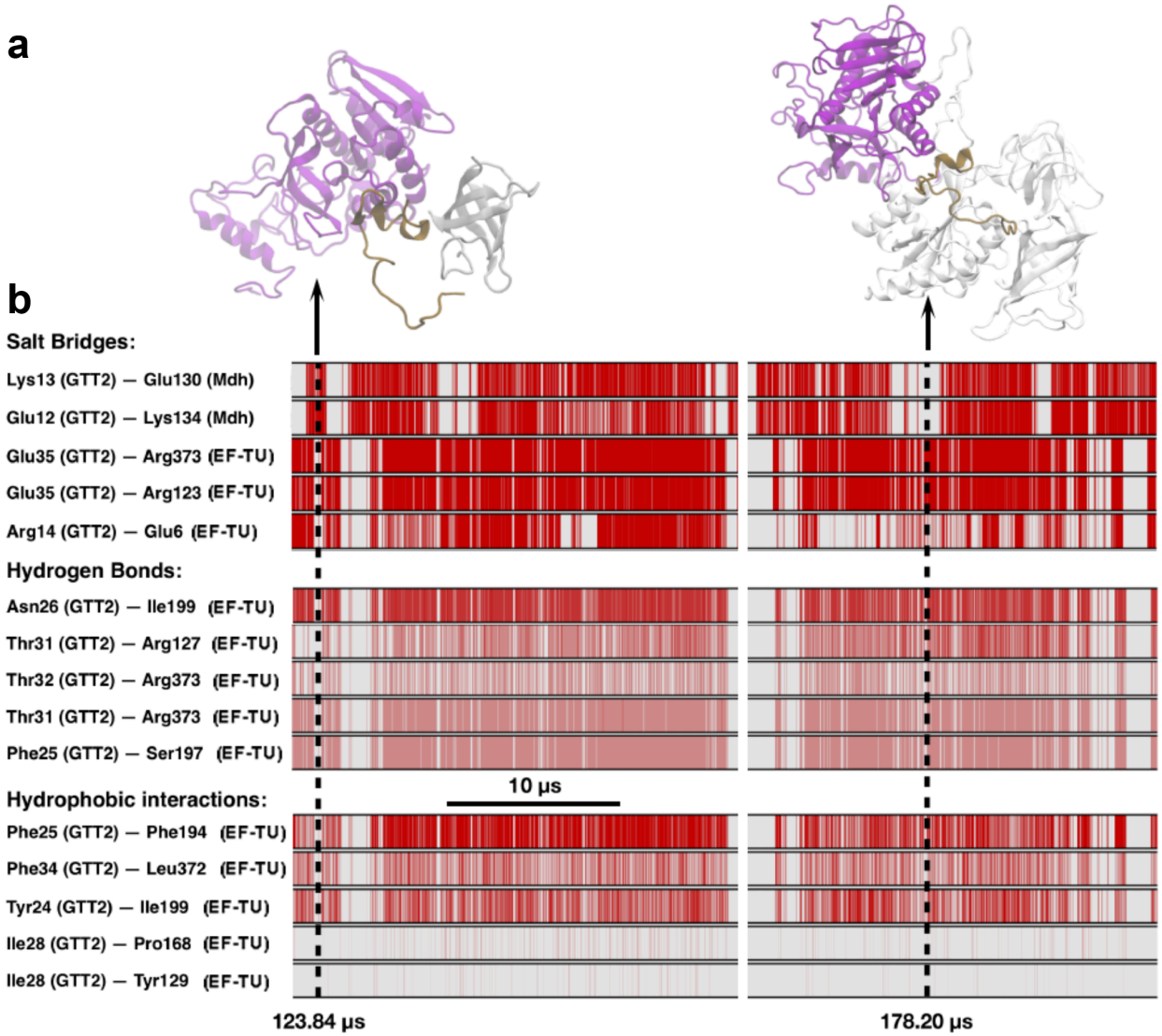


**Figure 6.** Another example of cytoplasmic sticking from the C22* simulation, corresponding to the flat region in the GTT2 radius of gyration after 120 µs in Figure 4b. GTT is interacting mainly with (a) Mdh and EF-TU via (b) salt bridges, hydrogen bonds, and hydrophobic interactions. At the very bottom are shown two residues on EF-TU that make only sporadic contact for comparison.

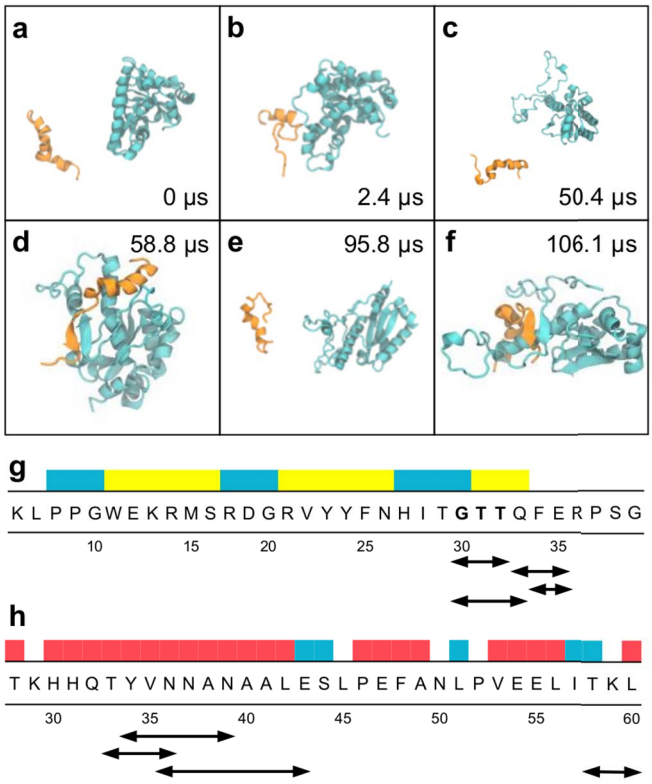


**Figure 7.** Chimeric interactions between GTT1 (orange) and Mn SOD (blue) in the C22* trajectory. (a) The proteins were not in contact at the beginning of the simulation. (b) GTT forms a contact with Mn SOD and unfolds Mn SOD's α-helix (residues 30-42) (c-f) GTT and Mn SOD continue to form and break protein-protein contacts including 1 antiparallel chimeric interaction (d) and multiple parallel chimeric interactions (f). These chimeric interactions can be remarkably long-lived. The interaction shown in (d) lasted for ~4.44 μs. The complete sequence of GTT (g) and a partial sequence of Mn SOD (h) are shown with their secondary structures indicated. The residues that form chimeric interactions during each event are indicated with black arrows below the respective sequences.

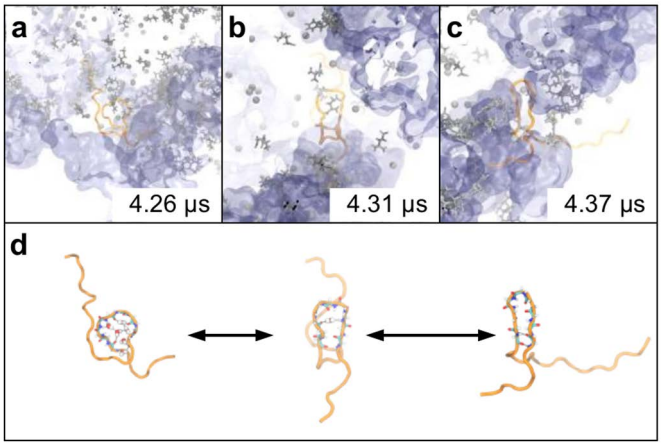


**Figure 8.** Representative snapshots of turn hydration during the β hairpin formation of GTT2 (U) in the C36mCU simulation. We observe a potential transition state that involves a single water bridging the backbone of the nascent turn (b). This state in a few observed cases either immediately precedes or immediately follows hairpin formation (c), or further solvation of the turn (a) (also see Figure S4). Panel (d) shows the same three states for clarity without the surrounding proteins, ions and metabolites, and highlights the bridging water molecules.

Table of Contents Image (TOC)