In-cell thermodynamics and a new role for protein surfaces

Austin E. Smitha, Larry Z. Zhoua, Annelise H. Gorensekb, Michael Senskeb, and Gary J. Pie latchac,ded,1

aDepartment of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; bDepartment of Physical Chemistry II, Ruhr-Universität Bochum, 44800 Bochum, Germany; cDepartment of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; and dLineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Edited by Martin Gruebele, University of Illinois at Urbana–Champaign, Urbana, IL, and approved December 14, 2015 (received for review September 21, 2015)

There is abundant, physiologically relevant knowledge about protein cores; they are hydrophobic, exquisitely well packed, and nearly all hydrogen bonds are satisfied. An equivalent understanding of protein surfaces has remained elusive because proteins are almost exclusively studied in vitro in simple aqueous solutions. Here, we establish the essential physiological roles played by protein surfaces by measuring the equilibrium thermodynamics and kinetics of protein folding in the complex environment of living Escherichia coli cells, and under physiologically relevant in vitro conditions. Fluorine NMR data on the 7-kDa globular N-terminal SH3 domain of Drosophila signal transduction protein drk (SH3) show that charge–charge interactions are fundamental to protein stability and folding kinetics in cells. Our results contradict predictions from accepted theories of macromolecular crowding and show that cosolutes commonly used to mimic the cellular interior do not yield physiologically relevant information. As such, we provide the foundation for a complete picture of protein chemistry in cells.

Results and Discussion

Stability in Buffer. In buffer at pH 7.2 and 298 K, \(\Delta G_U^{°} = 0.52 \pm 0.03 \text{ kcal/mol}\), \(\Delta H_U^{°} = 10 \pm 1 \text{ kcal/mol}\), and \(\Delta C_{p,U}^{°} = 0.87 \pm 0.06 \text{ kcal/mol} / K\) (Fig. 1C and SI Appendix, Tables S1 and S2). \(\Delta G_U^{°}\) is concentration independent from 11 \(\mu M\) to 1.1 mM (SI Appendix, Fig. S1) and consistent with reported values (4, 7).

Stability in Cells. Spectra were then acquired in E. coli (Fig. 1B).

At one extreme, no correction was made, such that the equilibrium constant for unfolding equals the area of the composite in-cell (IC) unfolded peak \(\langle U_{IC} \rangle\) over the area of the folded form \(\langle F_{IC} \rangle\):

\[
\Delta G_{U,IC}^{°} = -RT \ln \frac{U_{IC}}{F_{IC}}
\]  

This approach overestimates the population of the unfolded ensemble, and thus gives a minimum value for \(\Delta G_{U,T}\) (Eq. 3, green curve in Fig. 1C).

The second method accounts for the metabolite by using the following equations, which are described below:

\[
U_{basf} = \frac{\langle U_{IC} \rangle - \langle S \rangle \times U_{fac}}{\langle F_{IC} \rangle}
\]

\[
\Delta G_{U,corr}^{°} = -RT \ln \frac{\langle U_{IC} \rangle - \langle S \rangle \times U_{fac}}{\langle F_{IC} \rangle}
\]

The first step removes the contribution from any leaked metabolite. This was accomplished by examining the supernatant spectrum.

Significance

Understanding protein thermodynamics as it occurs inside cells is a fundamental goal of biophysics, and, from a practical point of view, will facilitate the design and improvement of protein-based drugs and catalysts. By measuring the temperature dependence of protein stability inside Escherichia coli cells, we show, contrary to predictions, that proteins are not necessarily stabilized inside cells compared with buffer alone. We also show that crowding-induced charge–charge interactions slow folding because of preferential interactions with the unfolded ensemble, and reducing these interactions increases protein stability.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

See Commentary on page 1684.

1To whom correspondence should be addressed. Email: gary.pielak@unc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1518620113/-/DCSupplemental.
which does not contain any SH3 protein, folded or unfolded. The area of the metabolite resonance in the supernatant spectrum (\(\int S\), Fig. 2B and E) was subtracted from the area of the composite peak in the in-cell spectrum (\(\int U_{IC}\)) to yield \(\int U_{IC} - \int S\).

The second step estimates \(U_{\text{frac}}\) (Eq. 4), the fraction of the composite in-cell peak, \(U_{IC}\), that represents the unfolded ensemble. This estimation was accomplished by using the spectrum of the cell lysate (Fig. 2C and F), which contains resolved peaks for the metabolite plus the unfolded ensemble in the lysate. The ratio of the area of the unfolded peak (\(\int U_{\text{lytate}}\)) to the total upfield peak (\(\int (U_{\text{lytate}} + M_{\text{lytate}})\)) provides \(U_{\text{frac}}\) (Eq. 4). The product of the two terms (\(\int U_{IC} - \int S\) * \(U_{\text{frac}}\)) divided by the area from the resonance of the folded form in cells (\(\int F_{IC}\)) gives a metabolite-corrected approximation of the equilibrium constant that was used to estimate \(\Delta G_{U,T}\) in cells (Eq. 5, red curve in Fig. 1C).

Using \(U_{\text{frac}}\) to correct the in-cell data is only an approximation because it assumes the population of the unfolded ensemble does not change on cell lysis. This approach probably overestimates \(\Delta G_{U,T}\) (gives the maximum stability, “cells-corrected” in Fig. 1C) because in vitro studies with protein crowders as well as in-cell studies show that destabilizing weak attractive interactions often dominate stabilizing hard-core excluded volume effect (13-15). We expect the true \(\Delta G_{U,T}\) lies between the two values.

Both curves indicate that SH3 is not stabilized in cells (Fig. 1C). \(T_m\) and \(\Delta G_U\) either decreased or were unchanged compared with buffer (SI Appendix, Tables S1 and S2), consistent with other studies, (13, 15–18) but inconsistent with previously accepted crowding theory (1, 2). \(\Delta C_{H^0,U}\) is the same in buffer and in cells (SI Appendix, Table S1). The stability decrease from the uncorrected data (0.53 ± 0.07 kcal/mol at 298 K) arises from a decrease in \(\Delta H_{U}\) (SI Appendix, Table S3), which is also inconsistent with theory. The corrected data indicate no change in \(\Delta H_{U,T}\). Further, diluted cell lysates and 100 g/L reconstituted lysate (14, 19) had little effect on SH3 stability (SI Appendix, Fig. S2). Because stabilizing hard-core repulsions are always present, our data show that these repulsions can be completely offset by attractive interactions in cells.

**Charge–Charge Interactions.** To probe electrostatic interactions between SH3 (\(p_l=6\), Fig. 1D–F) and protein crowders, we then performed in vitro experiments at several pH values in 100-g/L solutions of BSA (66 kDa, \(p_l=6\)) or lysozyme (14 kDa, \(p_l=9\)).

At pH 7.2, lysozyme destabilized SH3 relative to buffer (\(\Delta \Delta G_{U,T,298 K} = -0.92 \pm 0.03\) kcal/mol, Fig. 1D) and broadened its resonances (Fig. 3A and SI Appendix, Table S2). We attribute both effects primarily to attractive, nonspecific, charge-charge interactions between the protein surfaces. Consistent with this idea, adding
0.15 M NaCl to screen the charge–charge interaction decreased both the destabilization (ΔΔG°′U,298K = −0.70 ± 0.03 kcal/mol) and line width (SI Appendix, Table S2). BSA, which has a net charge similar to SH3 at this pH, was slightly stabilizing (ΔΔG°′U,298K = 0.09 ± 0.06 kcal/mol) and caused weak broadening, as expected for proteins having the same net charge (3).

We then decreased the pH to 5.4. SH3 is a polyanion at both pH 7.2 and 5.4, lysozyme remains a polycation but with a lower net positive charge, and BSA changes from a polyanion to polycation. Lysozyme is less destabilizing (ΔΔG°′U,298K = −0.76 ± 0.09 kcal/mol, Fig. 1E) at the lower pH, whereas BSA changes from slightly stabilizing to having no effect on stability. In addition, lysozyme causes less broadening at the lower pH, whereas BSA causes more broadening. All these observations reinforce the idea that charge–charge interactions play a key role in modulating stability. However, hydrogen bonds, weakly polar interactions, and hydrophobicity (20) must also contribute to attractive intermolecular interactions because at pH 3.0 (Fig. 1F), lysozyme and SH3 are both positively charged, yet SH3 remains destabilized (ΔΔG°′U,298K = −0.39 ± 0.05 kcal/mol).

**Synthetic Polymer and Their Monomers are Not Biologically Relevant.** The sucrose-, glucose-, and ethylene-glycol–based polymers, Ficoll, dextran, and PEG, respectively, have been used for decades to...
mimic the cellular interior (2). Contrary to what is observed in cells (Fig. 1C), the polymers stabilize SH3 relative to buffer (Fig. 4 and SI Appendix, Tables S1–S3). Furthermore, the monomers are more stabilizing than the polymers, the opposite of what is usually implied by the term “macromolecular crowding” (21). In addition, theory predicts that stabilization arises from entropic effects, yet ethylene glycol and PEG stabilize SH3 enthalpically, whereas the sugar-induced stabilization is entropic. Clearly, synthetic polymers are poor mimics of the cellular interior, and existing theories need to be modified. The modifications must account for nonspecific chemical interactions that act enthalpically to destabilize the protein and nonspecific repulsive chemical interactions that act enthalpically to stabilize the protein (22). The picture is even more complicated because the accounting must also consider solvent (23), including its size relative to the crowding molecules (21).

Biologically Relevant Crowders Interact More Strongly with the Unfolded Ensemble. Resonance broadening (Fig. 3A) is only an approximate measure of attractive intermolecular interactions. To obtain more detailed knowledge, we used $^{35}$F relaxation data to estimate tumbling times ($\tau_m$) (10) of SH3 under crowded conditions. In buffer, $\tau_m$ is 4 ns/rad for the folded state and 3 ns/rad for the unfolded ensemble (Fig. 3B and SI Appendix, Table S4), similar to published values (24). Ficoll (300 g/L) increases $\tau_m$ fourfold and fivefold for the folded state and unfolded ensemble, respectively. These increases likely arise from an increase in microscopic viscosity (19). The viscosity of the 100 g/L lysozyme solution is only 1.3 times that of water (19), yet $\tau_m$ increases 5-fold (folded state) and 13-fold (unfolded ensemble). BSA at 100 g/L has a similar viscosity, yet increases $\tau_m$ 3-fold (folded state) and 25-fold (unfolded ensemble). The large effect on the unfolded ensemble suggests that interactions with the unfolded ensemble cause the stability changes shown in Fig. 1.

Folding and Unfolding Rates Confirm Preferential Interactions of Biologically Relevant Crowders with the Unfolded Ensemble. We also quantified folding and unfolding rates (Fig. 3C and SI Appendix, Table S5). Rate data were acquired in BSA, lysozyme, urea (all at 100 g/L), and Ficoll (300 g/L). Ficoll decreased the folding and unfolding rates threefold. Slower folding is consistent with the viscosity increase. Slower unfolding in Ficoll is consistent with both viscosity and an entropic pressure for protein compaction (25, 26); however, limiting the explanation to viscosity and compaction effects is probably too simple. In contrast, BSA had only small effects, whereas lysozyme slowed folding fivefold but had no effect on unfolding. Like lysozyme, urea slowed folding fivefold, but increased unfolding threefold.

We speculate that urea’s small size allows it to penetrate the folded state to speed unfolding, while its interaction with the backbone in the unfolded ensemble slows folding, whereas lysozyme is too large to penetrate and affect unfolding, but slows folding via the aforementioned interactions with the unfolded ensemble. These observations reinforce the idea that biologically relevant conditions can destabilize proteins by facilitating favorable interactions with the unfolded ensemble.

Conclusions

Our data show that physiologically relevant information about protein exteriors has been hidden because proteins are studied in buffer instead of in cells. This limitation does not matter for protein cores; they yield relevant information in buffer because interior atoms experience the same environment in cells as they do in buffer—they are surrounded by other protein atoms. Exteriors are fundamentally different. In buffer, the surface is exposed to mostly water, but the cytoplasm is dramatically complex and crowded (27). The data indicate that crowding proteins interact with test protein surfaces, and these interactions affect both the equilibrium and kinetics of folding. Although synthetic polymers are important in industrial applications, they are poor models of the cellular interior. Our data also show that theories to explain protein behavior in cells must consider the folded state, the unfolded ensemble, and include terms for hard-core repulsions, solvent, hydrogen bonds, charge–charge–, hydrophobic–, and weakly polar interactions, all of which contribute to the enthalpic and entropic components of crowding effects. Recent modifications to theory (22), as well as simulations of intracellular dynamics, are pointing the way (27). Most importantly, our data, and those of others (13, 15–18), show that studying protein folding in living cells is key to gaining information needed to understand the many roles of proteins in biology.

Materials and Methods

Protein Expression for In-Cell NMR. Plasmids harboring the gene encoding SH3 were transformed into Agilent BL21 DE3 Gold cells by heat-shock. A single colony was used to inoculate a 5-mL culture of Lennox broth (10 g/tryptone, 5 g yeast extract, 5 g NaCl) supplemented with 100 µg/mL ampicillin. The culture was incubated with shaking at 37 °C (New Brunswick Scientific Innova I26, 225 rpm). After 8 h, 50 µL of the saturated culture was used to inoculate 50 mL of supplemented M9 media (50 mM Na$_2$HPO$_4$, 20 mM KH$_2$PO$_4$, 9 mM NaCl, 4 g/L glucose, 1 g/L NH$_4$Cl, 0.1 mM CaCl$_2$, 2 mM MgSO$_4$, 10 mg/L thiamine, 10 mg/L biotin, and 150 mg/L ampicillin, pH 7.4).

The 50-mL culture was shaken at 37 °C overnight. The culture was then diluted to 100 mL with supplemented M9 media. Five-fluoroorioleic (in dimethyl sulfoxide) was added to a final concentration of 0.1 g/L, and the culture was shaken for 30 min. Isopropyl β-D-thiogalactopyranoside (IPTG, 1 mM final

![Fig. 4. Synthetic polymers and their monomers. (A) Glucose and dextran, (B) sucrose and Ficoll (all at 300 g/L), (C) ethylene glycol, 8 kDa PEG, and 35 kDa PEG (all at 200 g/L) stabilize the SH3 domain. Buffer (black) curve is reproduced from Fig. 1C. Error bars for the 298-K data are the SD of three trials.](https://www.pnas.org/content/pnas/1518620113/fig/fig4)
In-cell samples were prepared as described above. For in vitro experiments, purified fluorine-labeled protein was added to NMR buffer (50 mM acetic acid/dimethyl acetamide, Hepes, bis-Tris propane, pH 7.2) containing the stated concentration of cosolute. The concentration of BSA/lysozyme was maintained at 0.10, 0.25, 0.50, 0.80, 1.00, and 1.50 mM. NMR experiments were performed on a Bruker 800 MHz spectrometer equipped with a Bruker QCI cryoprobe. Resonances were referenced to HDO at 9.4 ppm. Typical cell slurries were 50% wet cells by volume. In the presence of cell crowders, the concentration of BSA/lysozyme was maintained at 70 ppm in both dimensions; 1,024 complex points were collected with 64 or 72 complex points in each mixing time. The difference in the mixing time was used to drive the analysis. For the in-cell data, a similar method was used. Triplicate data were obtained for the 10°C, 25°C, and 40°C datasets. For the other datasets, uncertainties from the nearest-neighbor triplicate dataset were used to drive the analysis. For example, the 5°C and 15°C used the uncertainty associated with the 10°C dataset.

For in vitro experiments in the presence of crowders, triplicate 25°C data were acquired on the same sample. The SD was used to scale the uncertainties at other temperatures based on the SDs of the buffer dataset. These scaled values were used to drive the Monte Carlo Carlo analysis.

For the in-cell data, the value acquired from inversion recovery experiments for in vitro crowded conditions.
ACKNOWLEDGMENTS. We thank Linda L. Spremulli for inspiring us to study protein in cells, Paul J. Sapiezna and Andrew L. Lee for assistance with relaxation analysis, Alex J. Gusman for help with mass spectrometry, Matthew R. Redinbo and Elizabeth Pielak for helpful comments, and Gregory B. Young for spectrometer maintenance. This work was supported by the National Science Foundation (MCB 1410854) and the Cluster of Excellence RESOLV (EXC1069) funded by the Deutsche Forschungsgemeinschaft. M.S. acknowledges financial support from the Verband der Chemischen Industrie.