New Expi293 suite of products for structural biology, inducible expression, and protein labeling

Learn more

THE PROTEIN SOCIETY

Controlling and quantifying protein concentration in *Escherichia coli*

Shannon L. Speer,[1] Alex J. Guseman,[1] Jon B. Patteson,[1] Brandie M. Ehrmann,[1] and Gary J. Pielak [1,2,3,4]*

[1]Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina, 27599

[2]Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina, 27599

[3]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina, 27599

[4]Integrative Program for Biological and Genome Sciences, University of North Carolina, Chapel Hill, North Carolina, 27599

Received 31 March 2019; Accepted 2 May 2019
DOI: 10.1002/pro.3637
Published online 22 May 2019 proteinscience.org

Abstract: The cellular environment is dynamic and complex, involving thousands of different macromolecules with total concentrations of hundreds of grams per liter. However, most biochemistry is conducted in dilute buffer where the concentration of macromolecules is less than 10 g/L. High concentrations of macromolecules affect protein stability, function, and protein complex formation, but to understand these phenomena fully we need to know the concentration of the test protein in cells. Here, we quantify the concentration of an overexpressed recombinant protein, a variant of the B1 domain of protein G, in Tuner (DE3)™ *Escherichia coli* cells as a function of inducer concentration. We find that the protein expression level is controllable, and expression saturates at over 2 m*M* upon induction with 0.4 m*M* isopropyl β-D-thiogalactoside. We discuss the results in terms of what can and cannot be learned from in-cell protein NMR studies in *E. coli*.

Keywords: *Escherichia coli*; flow cytometry; in-cell NMR; mass spectrometry; protein expression

Introduction

Until recently, hard-core excluded volume was thought to be the key to understanding how the crowded and complex cytosol affects protein biophysics compared to dilute solutions.[1,2] Information from in-cell protein NMR studies in *Escherichia coli* cells and lysates show, however, that chemical (aka soft) interactions between macromolecules and the protein being studied (the test protein), can be as important, or even more important, than hard-core repulsions.[3,4] To gain the most information about hard- and soft- interactions it is necessary to know the concentration of test proteins in cells. For instance, to produce the binding isotherms required to quantify the strength of protein–protein interactions using in-cell NMR, it is imperative to know, and be able to control, the concentration of the binding partners. Information about test protein concentration is also required to understand the potential for contributions from test protein–chaperone interactions.[5] However, little is known about the concentration of test proteins in cells.

We set out to control and quantify the concentration of a variant of the B1 domain of streptococcal protein G (GB1)[6] in *E. coli*. The T2Q;L5V;F30V;Y33F;A34F mutant[7,8] was used for protein expression. We call this protein the GB1 variant.

Many expression systems exploit the *lac* operon, whose gratuitous inducer, isopropyl β-D-thiogalactoside (IPTG), acts in a stochastic and binary manner in commonly used *E. coli* strains, including BL21(DE3).[9]

Abbreviations: GB1, B1 domain of protein G; IPTG, isopropyl β-D-thiogalactoside.

Alex J. Guseman's current address is Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15260.

Grant sponsor: NSF Division of Chemistry CHE 1607359 CHE 1726291; Grant sponsor: NSF Division of Molecular and Cellular Biosciences MCB 1410854; Grant sponsor: K.C. Wong Education Foundation; Grant sponsor: National Institutes of Health CA016086 R01GM127291.

**Correspondence to:* Gary J. Pielak, Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599. E-mail: gary_pielak@unc.edu

Published by Wiley © 2019 The Protein Society

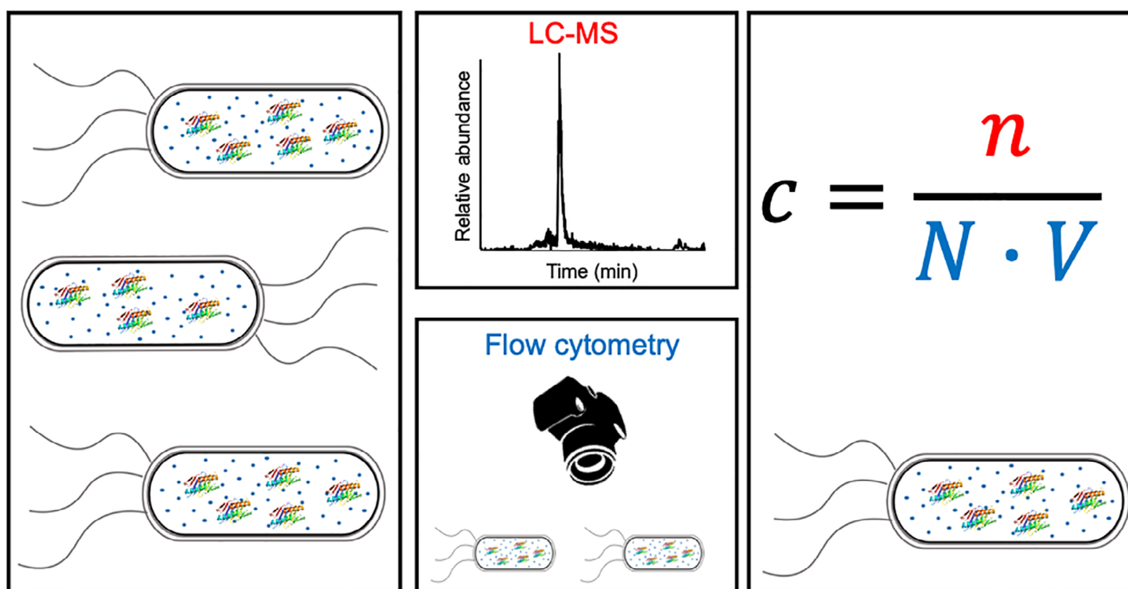

Figure 1. Quantifying protein concentration in cells. *Escherichia coli* cells expressing the GB1 variant (left panel) are subjected to flow cytometry (bottom of middle panel) to obtain the number of cells, *N*, and the cell volume, *V*. The cells are then lysed and subjected to LC–MS (top of middle panel) to determine the moles of protein, *n*, and the average molarity per cell calculated (right panel).

That is, expression is either "on" or "off." Low IPTG concentrations induce synthesis in a small minority of cells. At high concentrations, all cells express protein at a high level. This arrangement is not appropriate for examining protein–protein interactions because a valid binding isotherm requires that every cell contains the same test protein concentration. That is, expression must be homogeneous across all cells. To overcome this problem, we use Tuner (DE3) cells (Novagen), which harbor a deletion (Δ*lacZY*) of lactose permease that makes every cell equally permeable to IPTG.[9,10] The result is homogeneous expression that can be controlled, rheostate-like, by varying the inducer concentration.

To understand our method (Fig. 1), it is useful to consider the definition of molarity: the number of moles of solute per liter of solution. In our case, the solute is the GB1 variant and the "solution" is the cytoplasm. We use liquid chromatography–mass spectrometry (LC–MS) to quantify n, the number of moles, and flow cytometry to quantify both the number of cells, N and their average volume, V. These quantities are combined to give the molarity, C.

$$C = n/N \times V. \tag{1}$$

Results

The IPTG concentration was varied between 0 and 500 μ*M*. Chloramphenicol was used to halt expression. We confirmed the expression of the GB1 variant by comparing LC–MS data from cell lysates to data from the purified GB1 variant (Fig. 2). To quantify the number of moles of the GB1 variant, we analyzed extracted *E. coli* lysates with LC–MS (Fig. 3) using the method of standard addition to overcome the matrix effect.[11]

Cell size was determined from forward and side scatter of cells using flow cytometry. Our data indicate that the average length of an *E. coli* cell is 1.7 ± 0.2 μm and the average width is 1.20 ± 0.02 μm (Fig. 4). The volume, V, was calculated by assuming *E. coli* cells are cylinders. The average cell volume is $2.10 \times 10^{-15} \pm 0.03$ L. This volume is larger than the predicted range of $0.44–1.79 \times 10^{-15}$ L/cell,[12] because it includes the periplasm, which can amount to 20%–40% of the total cell volume under normal growth media.[13] Accounting for the periplasm decreases the volume to $1.7–1.3 \times 10^{-15}$ L, consistent with predictions. The predicted range of volumes[12] arises from different growth conditions and acts as a reminder that analysis of cell size must be performed under the conditions used for in-cell NMR.

The number of cells in a sample, N, was determined using a counting-flow cytometer and confirmed using OD_{600}.[12,14–18] We then quantified the average molarity per cell as a function of IPTG concentration

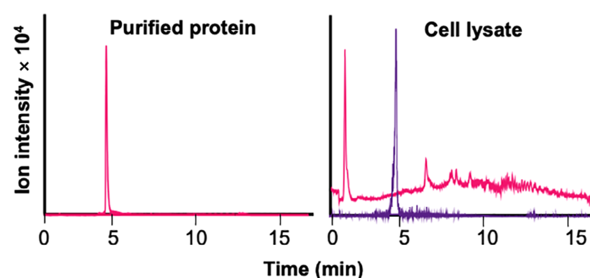

Figure 2. Quantifying GB1 variant concentration. Total ion chromatogram of purified GB1 (left panel), and a cell lysate (right panel, pink) with its extracted ion chromatogram (purple *m/z* 1040.6487 [M + H]$^+$) in *E. coli* lysate (right panel).

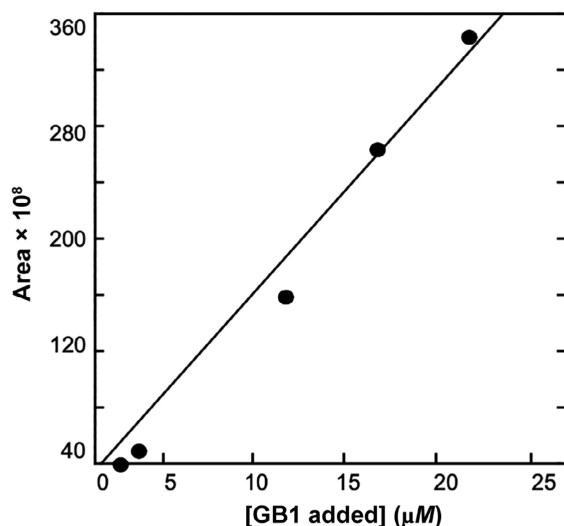

Figure 3. Correcting for matrix effects. *Escherichia coli* cells expressing the GB1 variant are lysed, and purified GB1 variant is added to the cell lysate. The cells are then subjected to LC–MS to determine the moles of GB1.

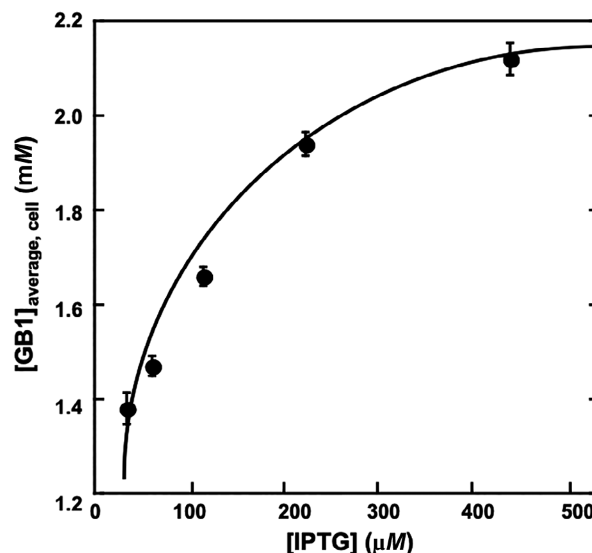

Figure 5. Average concentration of the GB1 variant in Tuner (DE3) cells as a function of inducer concentration. Uncertainties are propagated from triplicate analysis. The smooth curve is of no theoretical significance.

using the number of moles of the variant, the average cell volume, and the number of cells (Fig. 5).

Discussion

The intracellular concentration of GB1 varies between 1.4 m*M* using 31 μ*M* IPTG, and 2.1 m*M* at 500 μ*M* IPTG (Fig. 5). These concentrations are larger than that of any natural protein in *E. coli*,[19] including chaperones.[5] The protein with the highest concentration in *E. coli* (~100 μ*M*) is the chain elongation protein, EF-Tu.[19] Gro EL, the most highly expressed chaperone in *E. coli* has a concentration of less than 50 μ*M*.[5]

What do these findings mean for using in-cell protein NMR in *E. coli* to gain physiologically relevant data? The fact that the level of the test protein is often more than 10 times the concentration of the most abundant chaperone[5] and the fact that experiments are

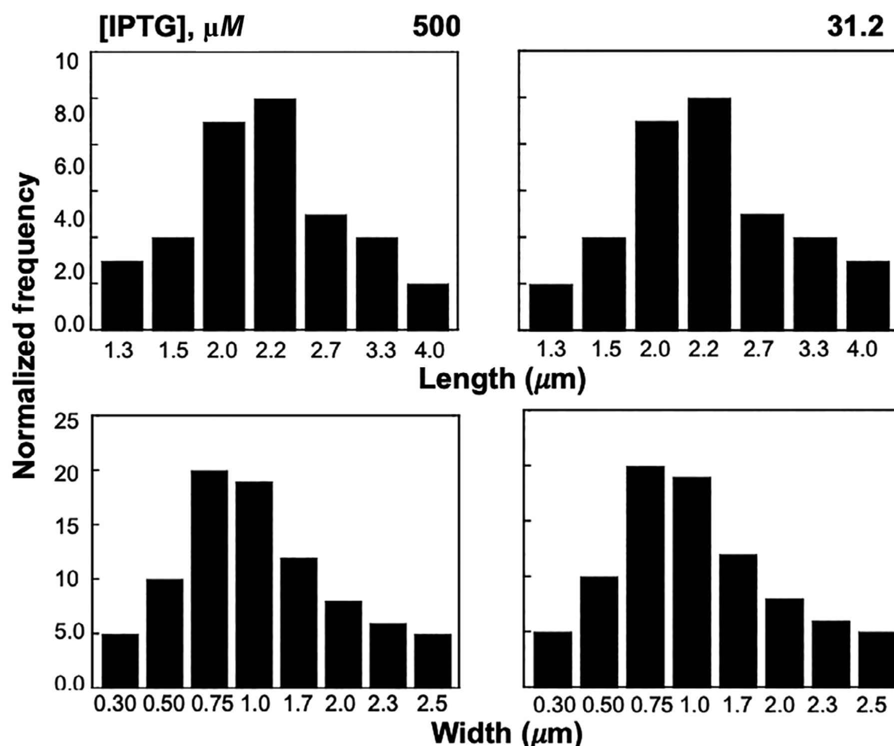

Figure 4. Inducer does not change the average size of *E. coli* cells. Cells are subjected to flow cytometry, and histograms of lengths and widths are obtained from forward and side scatter.

limited to no more than a few hours (because of viability and leakage problems)[20] means that chaperone–test protein interactions are probably not important.

The data also provide insight into what can be learned about protein–protein interactions in cells. The current detection limit for prokaryotic in-cell protein NMR is ~10 μ*M*, but 100 μ*M* is required for acquiring high-quality data, even with a cold probe.[21–24] Therefore, given the data in Figure 5, dissociation constants of ~50 μ*M* up to ~1 m*M* should be quantifiable.

The intracellular concentration of macromolecules under common growth conditions is 200–300 g/L[25,26] and does not increase when upon overexpression of a test protein.[27] Nevertheless, test-protein overexpression limits the physiological relevance of in-cell NMR data. For instance, one of the next frontiers in biophysics will be understanding quinary interactions in cells, the transient interactions between macromolecules that provide organization and compartmentalization inside cells.[28] Unfortunately, overexpression of even native *E. coli* proteins eliminates the potential for providing detailed information about physiologically relevant quinary interactions because overexpression necessarily spoils their stoichiometry. Such studies nonetheless provide essential information about protein biophysics in cells because the data reveal the types (charge–charge, hydrophobic, and hydrogen bonding) and strengths of interactions that comprise physiologically relevant quinary structure.[29]

Materials and Methods

A pET11a plasmid harboring the GB1 T2Q;L5V;F30V; Y33F;A34F mutant was used for protein expression. The plasmid was transformed into Tuner (DE3) *E. coli* cells (Novagen) by heat shock. A single colony was used to inoculate 5 mL of Lenox Broth (10 g/L tryptone, 5 g/L yeast, and 5 g/L NaCl) supplemented with 100 μg/mL ampicillin. This culture was incubated for 6–8 h at 37°C with shaking (New Brunswick Scientific Innova I26, 225 rpm), after which 500 μL was used to inoculate 50 mL of Lenox Broth and the culture was shaken overnight at 37°C.

The next day, 10 mL of the culture was used to inoculate 200 mL of supplemented M9 minimal media [50 m*M* Na_2HPO_4, 20 m*M* KH_2PO_4, 9 m*M* NaCl, 4 g/L glucose, 1 g/L NH_4Cl, 0.1 m*M* $CaCl_2$, 2 m*M* $MgSO_4$, 10 mg/L thiamine, 10 mg/L biotin, and 150 mg/L ampicillin (pH 7.4)]. The culture was incubated at 37°C and its optical density at 600 nm (OD_{600}) was monitored (Bio-Rad Spectra Plus). Once the OD_{600} reached 0.6, protein expression was induced by adding varying concentrations of IPTG (30–500 μ*M* final concentration). After 1 h, the OD_{600} was measured and chloramphenicol (50 μg/mL final concentration) added to halt protein expression.

Aliquots of 1 mL were collected and analyzed to determine the dimensions (Amnis ImageStreamX Mark II). Calibration beads (1 μm diameter) were included to monitor instrument performance (Apogee Flow). Small-angle, forward-scattered light was used to measure cell size. The samples were then diluted 1:1 with trypan blue to determine the total number of cells (Thermo Fisher Attune NxT).

Aliquots (1 mL) were collected and centrifuged for 10 min at 8000*g* (Eppendorf model 5430). These pellets were resuspended in 1 mL of autoclaved, deionized H_2O. Cells were lysed by sonication (Fischer Scientific Sonic Dismembrator model 500, 15% amplitude, 0.50 s on, and 0.5 s off for 1 min). The lysates were loaded onto solid phase exchange columns (Micro Bio-Spin 6) and centrifuged for 4 min at 1000*g*. Aliquots of the extracted lysates (75 μL aliquots) were flash frozen and lyophilized for 12 h (Labconco Freezone). GB1 standards were expressed and purified as described.[30]

The lyophilized samples were resuspended in 500 μL of LC–MS grade H_2O containing 0.01% formic acid (Optima LC-MS grade, Fisher Chemical). The resuspended samples were split into two samples, one of which was analyzed alone. Varying amounts (1–10 μ*M* final concentration) of a GB1 standard were spiked into the other half of the sample. Aliquots (5 μL) were analyzed by LC–MS. Separation was achieved with a Waters CSH C18 column (1.7 μm, 150 mm × 2.1 mm) in a gradient of 0%–100% mobile phase B over 11.25 min (initial mobile phase: H_2O, 0.01% formic acid; mobile phase B: acetonitrile, 0.1% formic acid). A Thermo Q Exactive™ HF-X Orbitrap™ mass spectrometer with an electrospray ionization source operating in positive ion mode was used. The GB1 variant was detected and quantified in the cell lysate extract ([M + H] *m/z* 1040.6487, Fig. 3).

Acknowledgments

Our research is supported by the National Science Foundation (MCB 1410854 and CHE 1607359 to GJP) and the National Institutes of Health (R01GM127291). We thank Elizabeth Pielak for comments on the manuscript. GJP thanks the K.C. Wong Education Foundation for travel support. The University of North Carolina Department of Chemistry Mass Spectrometry Core Laboratory is supported by the National Science Foundation (CHE 1726291). The UNC Flow Cytometry Core Facility is supported in part by the Cancer Center Core Support Grant to the UNC Lineberger Comprehensive Cancer Center (CA016086).

References

1. Minton AP (1981) Excluded volume as a determinant of macromolecular structure and reactivity. Biopolymers 20:2093–2120.
2. Ellis RJ (2001) Macromolecular crowding: obvious but underappreciated. Trends Biochem Sci 26:597–604.
3. Kyne C, Crowley PB (2017) Short arginine motifs drive protein stickiness in the *Escherichia coli* cytoplasm. Biochemistry 56:5026–5032.
4. Davis CM, Gruebele M, Sukenik S (2018) How does solvation in the cell affect protein folding and binding? Curr Opin Struct Biol 48:23–29.

5. Powers E, Powers D, Gierasch L (2012) FoldEco: a model for proteostasis in *E. coli*. Cell Rep 1:265–276.
6. Gronenborn AM, Filpula DR, Essig NZ, Achari A, Whitlow M, Wingfield PT, Clore GM (1991) A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. Science 253:657–661.
7. Byeon I-JL, Louis JM, Gronenborn AM (2003) A protein contortionist: Core mutations of GB1 that induce dimerization and domain swapping. J Mol Biol 333:141–152.
8. Byeon I-JL, Louis JM, Gronenborn AM (2004) A captured folding intermediate involved in dimerization and domain-swapping of GB1. J Mol Biol 340:615–625.
9. Nadeau JL. Introduction to experimental biophysics: biological methods for physical scientists. Boca Raton: CRC Press, 2016.
10. Abramson J, Smirnova I, Kasho V, Verner G, Kaback HR, Iwata S (2003) Structure and mechanism of the lactose permease of *Escherichia coli*. Science 301:610–615.
11. Taylor PJ (2005) Matrix effects: the achilles heel of quantitative high-performance liquid chromatography-electrospray-tandem mass spectrometry. Clin Biochem 38:328–334.
12. Koch AR, Button D (1996) Deduction of the cell volume and mass from forward scatter intensity and of bacteria analyzed by flow cytometry. J Microbiol Methods 27: 49–61.
13. Stock JB, Rauch B, Roseman S (1977) Periplasmic space in *Salmonella typhimurium* and *Escherichia coli*. J Biol Chem 252:7850–7861.
14. Bouvier T, Troussellier M, Anzil A, Courties C, Servais P (2001) Using light scatter signal to estimate bacterial biovolume by flow cytometry. Cytometry 44:188–194.
15. Bainer R, Park H, Cluzel P (2003) A high-throughput capillary assay for bacterial chemotaxis. J Microbiol Methods 55:315–319.
16. Slade KM, Baker R, Chua M, Thompson NL, Pielak GJ (2009) Effects of recombinant protein expression on green fluorescent protein diffusion in *Escherichia coli*. Biochemistry 48:5083–5089.
17. Tzur A, Moore JK, Jorgensen P, Shapiro HM, Kirschner MW (2011) Optimizing optical flow cytometry for cell volume-based sorting and analysis. PLoS One 6:e16053.
18. Volkmer B, Heinemann M (2011) Condition-dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling. PLoS One 6:e23126.
19. Link AJ, Robison K, Church GM (1997) Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. Electrophoresis 18: 1259–1313.
20. Barnes CO, Pielak GJ (2011) In-cell protein NMR and protein leakage. Proteins 79:347–351.
21. Burz DS, Dutta K, Cowburn D, Shekhtman A (2006) Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). Nat Methods 3:91–93.
22. Burz DS, Shekhtman A (2008) In-cell biochemistry using NMR spectroscopy. PloS One 3:e2571.
23. Pielak GJ, Li C, Miklos AC, Schlesinger AP, Slade KM, Wang GF, Zigoneanu IG (2009) Protein nuclear magnetic resonance under physiological conditions. Biochemistry 48:226–234.
24. Majumder S, DeMott CM, Burz DS, Shekhtman A (2014) Using singular value decomposition to characterize protein–protein interactions by in-cell NMR spectroscopy. ChemBioChem 15:929–933.
25. Cayley S, Lewis BA, Guttman HJ, Record MTJ (1991) Characterization of the cytoplasm of *Escherichia coli* K-12 as a function of external osmolarity: Implications for protein-DNA interactions *in vivo*. J Mol Biol 222:281–300.
26. Zimmerman SB, Trach SO (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. J Mol Biol 222: 599–620.
27. Dedmon MM, Patel CN, Young GB, Pielak GJ (2002) FlgM gains structure in living cells. Proc Natl Acad Sci USA 99:12681–12684.
28. Cohen RD, Pielak GJ (2017) A cell is more than the sum of its (dilute) parts: a brief history of quinary structure. Protein Sci 26:403–413.
29. Sarkar M, Li C, Pielak GJ (2013) Soft interactions and crowding. Biophys Rev 5:187–194.
30. Guseman AJ, Perez Goncalves GM, Speer SL, Young GB, Pielak GJ (2018) Protein shape modulates crowding effects. Proc Natl Acad Sci USA 115:10965–10970.