QUINARY STRUCTURE ALTERS PROTEIN FOLDING LANDSCAPES

Rachel D. Cohen

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry.

Chapel Hill
2016

Approved by:

Gary J. Pielak

Matthew R. Redinbo

Andrew L. Lee

Eric M. Brustad

John R. Riordan

ABSTRACT

Rachel D. Cohen: Quinary Structure Alters Protein Folding Landscapes
(Under the direction of Gary J. Pielak)

Most knowledge of protein chemistry is derived from experiments performed in dilute, buffered solutions. Although such experiments provide essential information, proteins function in the crowded and complex cellular environment, which imposes an additional level of structure. This quinary structure comprises the transient interactions between macromolecules that provides organization and compartmentalization inside cells. I have used in-cell NMR spectroscopy to characterize quinary structure, and have shown that globular protein stability is affected by quinary interactions involving both the folded state and the unfolded ensemble.

Chapter 1 reviews the history of quinary structure in the context of metabolic pathways, and the technological advances that have yielded recent insight into protein behavior in living cells.

In Chapter 2, I use the K10H variant of the B domain of protein G (GB1, 6.2 kDa) as a pH reporter in *Escherichia coli* cells to show that quinary interactions influence the quality of in-cell ^{15}N–^{1}H HSQC NMR spectra.

In Chapter 3, I quantify the pH-dependence of GB1 stability in cells. At neutral pH, GB1 stability in cells is comparable to that in buffer. As the pH decreases, the increased number of attractive interactions between *E. coli* proteins and GB1

destabilizes GB1 relative to buffer alone. I conclude that electrostatic interactions involving surface residues of the folded state contribute to quinary structure.

Chapter 4 shows that quinary structure can also affect the unfolded state ensemble. It has been known for many years that the unfolded ensemble of GB1 is stabilized by a non-native hydrophobic staple. Exploiting this idea, I made several mutations that do not change the folded state of GB1, but have a large effect on its stability in buffer. These effects are severely attenuated in cells, demonstrating that the cellular environment can remodel the unfolded ensemble.

My work shows that there is more to protein stability than a well-packed hydrophobic core; the key to understanding protein behavior in nature lies in quinary structure.

To Gary, live every week like it’s shark week

ACKNOWLEDGEMENTS

Before graduate school, most of my teachers would describe me as a shy, quiet student. However, all that changed when I started working for my advisor, Dr. Gary J. Pielak. We first met on my visit to UNC in February 2012. During my interview with Dr. Matt Redinbo (who has absolutely no recollection of this), we landed on the topic of swimming, and continued this conversation as he walked me to Gary's office. Even though Gary focused on macromolecular crowding for the rest of our meeting instead of α-synuclein, my initial interest in the Pielak lab, I knew I wanted to work with an advisor who shared my passion for swimming and my sense of humor.

When Gary declined my request to join him and Dr. Will Monteith in Ireland for the summer, I decided to do my summer rotation with Dr. Jack Riordan, where I had the opportunity to study the enigmatic cystic fibrosis transmembrane conductance regulator. This was my first real experience working in a lab; not only was I fascinated by the protein, but the Riordan group taught me good laboratory techniques and helped me gain confidence as a scientist. When I finally rotated through the Pielak lab, I had a difficult time deciding which lab to join. Instead of choosing one, I proposed a collaboration between the two labs, which Gary and Jack accepted.

I want to thank Gary and Jack for taking a chance on me and their continual support and encouragement throughout two years of making minimal progress on two different projects. I also want to thank Gary for deciding to put me on a different project, which worked out quite nicely (for more information, continue reading!) He even let me

attempt another collaboration with Matt but, as it turns out, it is very difficult to get a decent ^{19}F NMR spectrum of a 240 kDa protein.

I doubt I would have made it through graduate school without the support, trust and friendship of such a kind, understanding advisor. Gary, thank you for giving me independence in the lab, letting me learn from my own mistakes, and stepping in before I go too far down a rabbit hole. Thank you for teaching me that communication is the toughest thing we do, that diffusion is not just a good idea, it's the law, and to party like it's 1999. Thank you for knowing when to respond to my emails, and when they should just be ignored. Most importantly, thank you for helping me become a better scientist, scholar and writer. I will be forever grateful for our scientific discussions, the HOS, our trips to the pool, and that we share a similar sense of humor, because that guarantees at least one person will laugh about at my (sometimes inappropriate) jokes during group meeting.

Thank you to all members of the Pielak lab, past and present, for putting up with me! A special thank you to Dr. William Monteith, who developed the quenched lysate method and laid the foundation for my own projects. I am grateful that I was able to use the method to collect the data for my dissertation, and equally grateful that I did not have to sleep in the lab to develop it. I also want to thank soon-to-be Dr. Annelise Gorensek-Benitez, my fellow senior graduate member, for your support and life discussions. You're almost there!

I would also like to acknowledge Dr. Marc ter Horst for maintaining my (seemingly) personal spectrometer, the Varian 600. Not only did Marc quickly address spectrometer problems as they arose, but he was always available to address problems

outside of the laboratory, from life after graduate school to our shared passion for long distance running. Marc, thank you for all our discussions, and your continual encouragement and motivation.

I would also like to thank Dr. Matt Redinbo for finally remembering who I am (what an honor!). I enjoyed our brief collaboration, and am very grateful for your helpful scientific discussions, career advice and complementing Gary's mentorship.

Finally, I would like to thank my family for all your love and support. Mom and Dad, thank you for listening to my practice talks over Skype, and providing encouragement while dad snacks on his chips. Mom, I heard the Protein Society wants you to attend the conference next year! I'm so lucky to parents that are also my best friends. Annie and Maddy, thank you for helping me keep things in perspective, learning to navigate social media, and even though we live far apart, thank you for always being there for me. I love you all!

Thank you to everyone else who has guided me through this journey, I couldn't have done it without you!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

Å	angstrom
°C	degree Celcius
$\Delta G^{\circ'}_{op}$	standard, modified Gibbs free energy of opening
$\Delta G^{\circ'}_{U}$	standard, modified Gibbs free energy of unfolding
$\Delta\Delta G^{\circ'}_{op}$	change in standard, modified Gibbs free energy of opening
$\Delta\Delta G^{\circ'}_{quin}$	standard, modified quinary Gibb free energy of opening
$\Delta\Delta G^{\circ'}_{\text{var}}$	change in standard, modified Gibbs free energy of opening upon mutation
$\Delta H^{\circ'}_{cal}$	standard, modified calorimetric enthalpy of denaturation
δ	chemical shift
μ	micro- (10^{-6})
x *g*	centrifugal force, times gravity
DSC	differential scanning calorimetry
E. coli	*Escherichia coli*
F	folded state
g	gram
GB1	B1 domain of streptococcal protein G
h	hour
HDX	hydrogen deuterium exchange
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HSQC	heteronuclear single quantum coherence
IPTG	isopropyl β-D-1-thiogalactopyranoside
kcal/mol	kilocalories per mole

k_{cl}	closing rate constant
kDa	kilodalton
k_{int}	intrinsic rate constant of hydrogen deuterium exchange
k_{obs}	observed rate of hydrogen deuterium exchange
k_{op}	opening rate constant
K_{op}	equilibrium constant of backbone amide opening
K_U	equilibrium constant of unfolding
L	liter
LB	Luria Bertani medium
M	molar
m	milli- (10^{-3})
min	minute
NmerA	N-terminal metal-binding domain of mercuric ion reductase
NMR	Nuclear magnetic resonance
NOESY	nuclear overhauser spectroscopy
OD_{600}	optical density at 600 nanometers
PBS	phosphate-buffered saline
PGK	phosphoglycerate kinase
pI	isoelectric point
sec	second
SH3	N-terminal SH3 domain of Drosophila signal transduction protein drk
SPHERE	server program for hydrogen exchange rate estimation
SUPREX	stability of unpurified proteins from rates of hydrogen-deuterium exchange

TCA	tricarboxylic acid
T_m	melting temperature
U	unfolded state ensemble
var	variant
VlsE	variable major protein-like sequence expressed
wt	wildtype

CHAPTER 1: A CELL IS MORE THAN THE SUM OF ITS (DILUTE) PARTS: A BRIEF HISTORY OF QUINARY STRUCTURE[1]

1.1 Introduction

The first four levels of protein structure were proposed by Linderstrøm-Lang in 1952.[1] The amino acid sequence defines the primary structure.[2,3] Secondary structure comprises individual α-helices, β-strands and turns that are stabilized by intramolecular hydrogen bonds.[4] The three-dimensional arrangement of atoms[5] forms the tertiary structure, which is also stabilized by hydrophobic interactions.[6-8] Quaternary structure describes the assembly of individual polypeptide chains in multimeric proteins.[9]

These four levels formed the basis for discovering the relationships between protein folding and function.[10] This hierarchy, however, is only sufficient to describe proteins in isolation. Evidence for cellular compartmentalization of metabolic pathways was established in the 1960s.[11] Quinary structure, the fifth level of protein structure, comprises the interactions between macromolecules that organize the cellular interior. The term was coined three times. In 1973, Vaïnshtein described quinary structure as the level of organization of the "combination of molecules of proteins, nucleic acids and nucleoproteins into aggregates."[12] In 1980, Edelstein extended the four levels of protein structure to include quinary structure to describe the "marked plasticity and inequivalence in the juxtaposition of constituent molecules."[13] In 1982, McConkey

[1]The material for this chapter has been published in Protein Science. The original citation is as follows: Cohen RD, Pielak GJ (2016) A cell is more than the sum of its (dilute) parts: A brief history of quinary structure. Protein Science. RD Cohen and GJ Pielak wrote the paper.

reasoned that the conservation of isoelectric points among homologous proteins was “a consequence of complexities of intracellular organization and the numerous macromolecular interactions in which most polypeptides participate,” which he called “quinary structure”.[14]

The cellular interior is a crowded, dynamic environment; quinary structure is crucial for controlling and maintaining the diverse biochemical processes and signaling cascades. Quinary structure facilitates the assembly of multi-enzyme complexes such the Krebs cycle enzymes,[15,16] the purinosome,[17] and the respirasome.[18,19] Here, we illustrate quinary structure in the context of these systems. We then discuss in-cell NMR and other recent contributions to understanding quinary structure. Finally, we address outstanding questions in the field.

1.2 Quinary structure maintains metabolic pathways

1.2.1 Citric Acid Cycle

Evidence for quinary structure was initially reported in 1948 when Green isolated an aggregated supramolecular complex of the Krebs cycle enzymes, which he called the cyclophorase system.[20] It was shown later that Green had actually isolated mitochondria,[21] but the evidence for intracellular organization of the enzymes continued to accumulate.

The formation of multienzyme complexes facilitates substrate channeling, the direct transfer of metabolites between active centers.[22-26] The bifunctional tryptophan synthase, which catalyzes the last two reactions in L-tryptophan synthesis, is the simplest example of a complex that channels a metabolite between reaction centers.[27-29] A tunnel within the protein transfers the metabolite from the first reaction center to the

second. In 1973, Srere demonstrated the kinetic advantage of a microenvironment and reported higher rates of citrate formation by a group of three immobilized Krebs cycle enzymes than by their soluble counterparts.[30] In 1977, he reported specific interactions between a couple of these enzymes in the presence of polyethylene glycol.[31] Numerous groups have since described specific interactions among these enzymes in solution[32-35] and *in vivo*.[36]

In 1985, Srere coined the term 'metabolon' to describe a "supramolecular complex of sequential metabolic enzymes and cellular structural elements" that was "formed by quinary interactions of complementary surfaces."[37] These complexes facilitate substrate channeling by protecting intermediates[38-40] and result in enhanced rates of catalysis and metabolic flux.[41] Brownian dynamics simulations attribute the transfer efficiency between a fusion protein of two tricarboxylic acid (TCA) cycle enzymes to electrostatic channeling of negatively charged substrates between positively charged active sites.[42,43] The simulation was experimentally validated by demonstrating that high salt concentrations eliminate substrate channeling between the two proteins.[44]

Recently, all eight enzymes in the TCA cycle metabolon were identified by *in vivo* cross-linking and mass spectrometry.[45] Residues important for quinary interactions were identified. In addition, surface electrostatic potentials of the enzymes showed formation of a positively charged channel upon association, in agreement with the studies on the fusion protein.[42-44] Site-directed mutagenesis was used to identify a single arginine residue in citrate synthase crucial for interactions with malate dehydrogenase.[46] While changes to this residue have a negligible effect on enzyme activity, they significantly decrease the overall channeling probability and metabolic flux. These studies provide

important information about quinary structure and its role in governing metabolic pathways.

1.2.2 Purinosome

The purinosome comprises the enzymes responsible for the ten-step biosynthesis of purine.[47] The first evidence for quinary interactions between these enzymes was reported in 1978 when the Benkovic group copurified formyl-methenyl-methylenetetrahydrofolate synthetase and glycineamide ribonucleotide transformylase, raising "the possibility that the two proteins may specifically aggregate or be covalently linked in the native state."[48] Two years later, they reported the co-purification of glycineamide ribonucleotide transformylase with two other enzymes.[49] In both instances, each enzyme exhibited decreased activity upon isolation. The instability of the first intermediate,[50] suggestive of a direct transfer between enzymes, and the identification of multifunctional polypeptide chains[51] provided further evidence for a multi-enzymatic complex.

In 2008, fluorescence microscopy revealed compartmentalization of all six enzymes in the cytoplasm of human cells.[17] An *et al.* found that formation of the six-enzyme complex was dynamic, reversible and depended on purine levels. The Benkovic group coined the term "purinosome" to describe this complex. Its formation is cell-cycle dependent[52] and regulated by G-protein-coupled receptors[53], kinases[54] and specific signaling pathways.[55] Live cell imaging[56-58] was employed to characterize the spatial organization of the purinosome and revealed that its formation was embedded within the microtubule network.[59] Spatial co-localization between the purinosome and mitochondria has since been confirmed.[60]

Application of the Tango assay[61] to characterize protein-protein interactions within the purinosome identified specific interactions between the six enzymes, and suggested that the first three form the "core."[62] The first quantitative measurements of purinosome formation dynamics[63] were made using fluorescence recovery after photobleaching.[64,65] Under purine-rich conditions, each enzyme exhibited its own diffusion coefficient, showing that the enzymes diffuse independently. Under purine-depleted conditions, the diffusion coefficient had two components. These observations were interpreted as dynamic partitioning of the enzymes: the fast component reflected free diffusion, while the second component implied purinosome formation.[63]

1.2.3 Respirasome

The respiratory chain is another example of a pathway governed by quinary structure. The chain comprises four complexes (I through IV) in the inner mitochondrial membrane that drive oxidative phosphorylation. The notion that respiratory enzymes form higher order complexes was proposed in 1947 by Keilin.[66] The organization of the respiratory chain was described initially by the solid state model, which suggested permanent interactions between complexes.[66,67] This model was supported by the isolation of assemblies of multiple complexes.[68-70] Further analyses suggested the complexes assembled stoichiometrically.[71-73]

However, the isolation of independent, functional complexes[74-76] supported the random collision model, in which the individual complexes assemble dynamically.[77] This model was supported by reports of independent diffusion of the complexes,[77-80] and reports of higher order assemblies containing various distributions of complexes.[81,82] Ultimately, the plasticity model[83] prevailed, which synthesized the two previous models

and “envisions the mitochondrial respiratory chain as a combination of free complexes and which are able to adapt to changing conditions.”[84,85] Thus, quinary interactions play a pivotal role in orchestrating the organization and functionality of the respiratory chain.

In 2000, Schägger and Pfeiffer isolated complexes I, III and IV from mammalian mitochondria and coined the term “respirasome” to describe this 1.7 MDa supercomplex.[81] The respirasome comprises 81 polypeptide chains divided among the three multi-subunit respiratory chain complexes and establishes the proton gradient across the inner mitochondrial membrane that is essential for ATP synthesis and energy conversion. Multiple configurations of this supercomplex have been reported.[86,87] The supercomplex $I_1III_2IV_1$ attracted the most attention.[88-91] Recently, two groups used single particle cryo-electron microscopy to produce high-resolution structures of the mammalian respirasome that reveal important information about the spatial arrangements of the individual complexes and the key motifs necessary for higher order assembly.[18,19] Gu *et al*.[18] reported a 5.4 Å resolution structure purified from porcine heart. Letts *et al*.[19] reported structures of both a ‘tight’ and ‘loose’ conformation (5.8 Å and 6.7 Å, respectively) of the ovine respirasome. In the loose conformation, CIV forms well-defined contacts with CI, whereas in the tight conformation, there are well-defined contacts between all three complexes. Interconversion between the tight and loose conformations may depend on cardiolipin, which stabilizes CIII and CIV interactions.[92] The authors note the absence of scaffold proteins, and demonstrate that the active sites of the complexes are well positioned for substrate channeling.

1.3 Characterizing quinary structure

Most protein biochemistry research is carried out at protein concentrations less than 10 g/L, yet *Escherichia coli* cells contain over 300 g/L of macromolecules.[93] Eukaryotic cells contain a greater variety of macromolecules that participate in more complex cellular processes, and compartmentalization within organelles emphasizes the important of quinary structure.[94-96] In 1984, Clegg asserted “it is no longer reasonable to assume that molecules in cells experience an aqueous cytoplasmic microenvironment comparable to the test tube.”[97] Despite early observations that transient chemical interactions modulate protein behavior in cells, interest in quinary structure remained dormant, partially because technology to study proteins in cells did not yet exist.

The dynamic and reversible assembly of the purinosome, TCA cycle metabolon and the respirasome illustrate the intricacy and essential role of quinary structure. But what do we know about the fundamental nature of these interactions? How do they change protein behavior in cells compared to buffer? Next, we answer some of these questions, and discuss some of the questions that remain.

1.3.1 In-cell NMR spectroscopy

The transient nature of quinary interactions and the challenges associated with studying proteins in living cells has limited fundamental information about the effects of quinary interactions on protein properties. In 2001, Dötsch used heteronuclear nuclear magnetic resonance (NMR) spectroscopy to observe 15N-enriched-enriched proteins overexpressed in living Escherichia coli cells,98 asserting “in-cell NMR spectroscopy will open new avenues of research into protein conformations in their natural

environment, as affected by protein-protein interactions, reversible small molecule binding, and posttranslation modifications."

Investigators, however, quickly realized that in-cell NMR is limited by the fact that many globular proteins cannot be seen in cells.[99] Background resonances from *E. coli* macromolecules are not to blame; the absence or presence of resonances depends on the degree of protein overexpression and the rotational correlation time of the molecule.[100] Disordered proteins, which exhibit fast internal motions, almost always yield high quality in-cell HSQC spectra.[99] Our group constructed a fusion protein comprised of a disordered protein, α-synuclein, and a globular protein, ubiquitin, and compared the in-cell HSQC spectra.[101] The ubiquitin portion failed to produce an in-cell HSQC spectrum and the only observable crosspeaks belonged to α-synuclein. However, both sets of crosspeaks were visible in the cell lysate, which is diluted compared to the in-cell sample. This result demonstrates that the success or failure of in-cell NMR depends on a molecule's rotational correlation time, which for gloublar proteins is increased in the cellular environment, resulting in faster transverse relaxation and line broadening. The increased rotational correlation time of folded, globular proteins in cells is attributed to quinary interactions, which increase the effective molecular weight.

Cytochrome *c* is an important model protein,[102,103] but does not yield an in-cell ^{15}N-^{1}H HSQC spectrum, nor is the spectrum detectable in cell lysates.[104] Crowley *et al*. attributed these observations to interactions between highly basic (pI >10) cytochrome *c* and the excess of primarily negatively charged macromolecules in the cellular environment. They demonstrated this "stickiness" by subjecting whole cell lysates to

size exclusion chromatography and found that cytochrome *c* eluted at a volume corresponding to over ten times its true molecular weight. Interestingly, a charge reversal on the protein's surface altered its elution profile. Replacing three basic residues with glutamates weakened quinary interactions such that the triple variant was detectable by in-cell NMR. This observation suggests that electrostatic interactions are an important component of quinary structure, an idea we discuss later in this review.

Similarly, Gierasch and colleagues investigated the effects of quinary structure on in-cell NMR spectra by characterizing the rotational diffusion of three small globular proteins, GB1, NmerA, and ubiquitin, in *E. coli*.[105] These proteins are roughly the same size (6.2, 6.9 and 8.5 kDa, respectively), but have different surface properties. GB1 and NmerA could be observed in cells, but as discussed above, crosspeaks from ubiquitin were absent. Despite their similar size, GB1 yielded a significantly higher quality in-cell ^{15}N-^{1}H HSQC spectrum than did NmerA. The authors constructed a GB1-L_{15}-NmerA fusion protein and, despite equal levels of protein expression, NmerA suffered severe line broadening. Moreover, the in-cell spectrum of a fusion protein linking two GB1 domains (12.5 kDa) was still of higher quality than that of NmerA. These experiments ruled out protein size as the main contributing factor to the increased rotational correlation time. GB1 is negatively charged (pI 4.2), while both NmerA (pI 7.2) and ubiquitin (pI 7.3) have a net charge of about zero near physiological pH. However, mutational studies revealed that the differences in in-cell NMR spectral quality are due to the differences in distribution of hydrophobic residues on the surfaces of NmerA and ubiquitin. These results demonstrate that protein surfaces determine the ability to

acquire high quality in-cell NMR spectra and that hydrophobic interactions also play a role in quinary structure.

Nevertheless, in-cell NMR has provided the first atomic level detail of protein structure in cells. In-cell NMR was used to determine the three dimensional structure of several proteins in their physiological state.[106,107] Quinary structure is unlikely to affect the tertiary structure of globular proteins because Anfinsen showed the structure of a globular protein represents the lowest free energy state,[10] and Richards showed that the inside of a globular protein is almost as efficiently packed as perfectly packed spheres.[108]

On the other hand, FlgM is a protein that is disordered in solution, but gains structure in cells.[109] There has also been much debate about the oligomeric state of α-synuclein,[110] but careful in-cell NMR and EPR data reveal its physiological state is a disordered monomer.[111,112]

Although prokaryotic cells are established systems that enable rapid overexpression of proteins, they lack the breadth of biological activities and post-translational modifications of eukaryotic cells. In 2006, Selenko and colleagues injected the B1 domain of protein G (GB1), a model protein into *Xenopus laevis* oocytes.[113,114] The utility of in-cell NMR has also been extended to proteins overexpressed in human cells.[115,116] Majumder *et al.* used in-cell NMR to identify key residues involved in quinary interactions for four different proteins in both prokaryotic and eukaryotic cells.[117] The key residues comprise hydrophobic and hydrophilic residues on the protein surface, and were highly conserved across species. RNA was also identified as a component of quinary structure. Similarly, Barbieri *et al.* showed that replacing positively charged

residues on the surface of the human protein profilin 1 with neutral or negatively charged residues improved the quality of in-cell HSQC spectra in both bacterial and human cells.[118] They were able to differentiate between residues involved in specific interactions and non-specific interactions, contributing to the mounting evidence that electrostatic interactions play an important role in forming quinary structure.

^{19}F NMR spectroscopy is another approach for studying protein structure and stability,[119,120] and fluorine was one the first nuclei exploited for in cell NMR.[36,121-124] Fluorine is rarely found in biology,[125,126] and the ^{19}F nucleus is NMR-active, extremely sensitive and does not perturb protein structure. Fluorine-containing amino acids can be incorporated into proteins at a specific site,[123] or by feeding *E. coli* fluorine-containing precursors, such a fluoroindole.[127] Since these strategies incorporate fluorine at one or more specific residues, depending on the protein, spectral analysis is simplified and ^{19}F chemical shifts and line widths provide information on site-specific structure and dynamics. One dimensional ^{19}F NMR experiments are less time-consuming than traditional two-dimensional methods; spectra can be acquired in minutes with protein concentrations as low as 100 μM.[124] Although ^{19}F NMR does not offer the breadth of information afforded by heteronuclear multidimensional in-cell NMR experiments, it is a simple approach for obtaining general information about proteins in cells, including larger proteins that may not afford two-dimensional in-cell spectra. Furthermore, ^{19}F NMR can be used to study protein-ligand interactions[128,129] with biomedical applications.[130-133]

1.3.2 Quantitative assessments of quinary structure

In 2000, the Fitzgerald and Oas laboratories developed a mass spectrometry based method to measure the stability of unpurified proteins from rates of hydrogen-deuterium exchange (SUPREX).[134] Briefly, they resuspended unpurified protein extracts in deuterium-containing buffer and used mass spectrometry to measure protein mass as a function of time. The data were fit to an exponential function to obtain the rate of exchange and to estimate protein stability. A year later, Ghaemmaghami and Oas adapted SUPREX to provide the first quantitative comparison of protein stability in cells and in buffered solution.[135] They used the SUPREX method to show that the stability of monomeric λ repressor in *E. coli* is comparable to the stability *in vitro*, which was a radical idea at that time, because crowding was thought to only stabilize proteins.

In 2014, our lab presented the "quenched lysate method" to quantify protein stability in living *E. coli* cells.[136] Adapted from SUPREX,[135] the quenched lysate method pairs amide proton exchange and NMR spectroscopy[137,138] to yield residue level information about protein stability in cells. Poor signal-to-noise ratio prevented quantification of exchange directly from in-cell NMR. Instead, the protein of interest is overexpressed in ^{15}N-containing media, and the cells are transferred to D_2O-containing buffer after chloramphenicol is added to halt protein expression. Aliquots are removed at discrete time points and hydrogen exchange is quenched while cells are lysed simultaneously. The lysate is transferred to an NMR tube and an ^{15}N-^{1}H HSQC spectrum is acquired. Samples are collected until the signal decays two complete half-lives.

The quenched lysate enabled quantification of the stability of several GB1 variants in cells. In buffer, mutations to residue D40, a solvent exposed residue not involved in secondary structure formation, had a negligible effect on GB1 stability. However, as the mutations became increasingly positively charged, GB1 stability decreased.[139] These results support observations that surface charge plays an important role in quinary structure.[104,105]

To explore the role of charge-charge interactions in quinary structure, we quantified GB1 stability in *E. coli* cells at different pH values. Using the K10H variant of GB1 to report on intracellular pH and a buffer to control the intracellular pH, we showed that the quality of the in-cell HSQC spectrum degrades at lower pH values.[140] We used the quenched lysate method to quantify GB1 stability at different pH values to show that lowering the pH destabilizes the protein to a greater extent in cells than in buffer.[141] At pH 5.0, the protein is destabilized by over 1 kcal/mol in cells compared to buffer at 37 °C, providing further evidence of an electrostatic contribution to quinary structure.

The fact that GB1 is not natural to *E. coli* is also important. The lack of homologs in *E. coli* means there are no specific interactions to interfere with the proposed studies of how the intracellular environment affects the protein. Simply put, the proposed work defines a path to knowledge about the potential of quinary interactions. The situation is analogous to the early days of site-directed mutagenesis with its focus on model proteins such as T4 lysozyme,[142] GB1,[143] eglin c[144] and the barnase/barstar complex.[145] This classic work delivered vital information about key protein properties, including the propensity of amino acids to form α-helices and β-sheets, the importance of packing, and the energetics of protein complex formation. Studies of heterologous systems like

GB1 have the same potential to advance both fundamental and practical knowledge about how proteins behave in cells.

Additional methods have been developed to quantify protein thermodynamics in cells. In 2015, Danielsson *et al.* used in-cell NMR to quantify the stability of a marginally stable β-barrel protein (I135A SOD1) in mammalian cells.[146] Since crosspeaks are observable for both the folded and unfolded state of this protein, the ratio of crosspeak volumes was used to quantify stability. Their work shows that the protein is destabilized in cells relative to buffer. The temperature-dependence of the stability shows that the destabilization is enthalpic.

Smith *et al.*[147] used a similar approach to quantify the stability of the N-terminal SH3 domain of *Drosophila* signal transduction protein drk (SH3). SH3 has a $\Delta G^{\circ'}_{U}$ of ~0 kcal/mol at 37 °C, meaning that the folded and the unfolded states are equally populated at equilibrium. By incorporating 5-fluoroindole at a single residue, they used ^{19}F NMR to show that the stability of SH3 in cells is comparable to, or less than, its stability in buffer. These results show that attractive quinary interactions offset, and perhaps dominate, the stabilizing effect of hard-core repulsions.[148]

Studies of protein folding in living cells using Förster resonance energy transfer experiments have also yielded important results. Gruebele's group showed that phosphoglycerate kinase (PGK) has a more compact structure and experiences accelerated folding rates in the nucleus of U2OS bone tissue cancer cells, compared to buffered solutions.[149] In contrast, the cytoplasm slows PGK folding relative to buffer. These results emphasize the differential effects resulting from cellular compartmentalization. These authors also showed that PGK (pI 7.8) is stabilized in

cells, while the temperature-dependence of the folding kinetics remains similar to those in buffer.[150] In contrast, a truncation of the Variable major protein-Like sequence Expressed (VlsE) is destabilized and its folding kinetics are slowed in U2OS cells compared to buffer.[151] While PGK and VlsE are similar in size and isoelectric point, the authors suggest "hydrophobic contacts of unfolded VlsE… could stabilize the unfolded state when it sticks to cellular crowders" to explain the different effects of the cellular environment on the stability of the two proteins.

1.4 Conclusions and perspectives

Evidence of quinary structure emerged over half a century ago, but efforts to characterize and understand the fundamental physical chemistry of these transient interactions were undertaken only recently. In-cell NMR spectroscopy is a valuable tool for obtaining both molecular-level and residue-level information about proteins in cells. In recent years, we have learned that quinary interactions can impede the acquisition of in-cell NMR spectra of globular proteins, largely due to electrostatic and hydrophobic interactions between the test protein and the intracellular environment. In-cell NMR has also been used to identify key surface residues that form quinary interactions. The cellular environment also changes the folding landscape of proteins, and hydrophobic interactions in the unfolded state must also be considered when trying to understand quinary structure. Most importantly, we have learned that the protein surfaces, both in the folded or unfolded state, mediate interactions with the intracellular environment, and that the effects of quinary structure depend on the inherent properties of the protein of interest.

These same ideas concerning the importance of protein surfaces are also applicable to the formulation of biologics. The ability to successfully formulate these life-saving, protein-based drugs depends on the lack of what the pharmaceutical industry calls higher order structure[152] of highly concentrated solutions. Such structure can result highly viscous solutions that are impossible to inject. It is known that the surface properties of biologics contribute to viscosity.[153]

Although textbooks discuss the four levels of protein structure, they now must include five. Quinary structure is responsible for the elegant organization of the dynamic, complex cellular environment and governs the billions of signaling pathways and cascades vital for cell survival. Experiments using purified proteins have formed the basis for our knowledge about protein structure, function and dynamics; however, this information must be put in the context of the physiological environment to fully understand the implications of these characteristics. We believe the key to understanding protein behavior in nature lies at the surface, in the residues that mediate interactions with the rest of the cellular environment to form quinary structure.

CHAPTER 2: INTRACELLULAR PH MODULATES QUINARY STRUCTURE[1]

2.1 Introduction

The pH is a fundamental parameter in protein science because it influences many protein properties, including their stability and solubility.[154] Traditionally, proteins were studied in simple solutions, containing salt to control the ionic strength and buffer to control the pH. However, most proteins function inside cells, where the concentration of macromolecules can exceed 300 g/L.[93] In-cell NMR spectroscopy[98,155,156] is a powerful, non-perturbing tool for studying proteins in their native environment. NMR and other techniques have been used to show that crowding modulates the stability of proteins, both enthalpically and entropically.[157-160] The cellular interior introduces an additional element: quinary structure[14,139,161-164] supports intracellular organization and arises from transient interactions between macromolecules that are extremely weak in buffer.

From a practical point of view, the sum of these individually weak interactions between the test protein and intracellular components can hinder the acquisition of in-cell NMR spectra of globular proteins, particularly in *Escherichia coli.*[156,164,165] The B domain of protein G (GB1, 6.2 kDa)[166,167] is one of the few globular proteins that yields high quality ^{15}N-^{1}H HSQC spectra in cells,[104-106] making it a quintessential test protein

[1]The material for this chapter has been published in Protein Science. The original citation is as follows: Cohen RD, Guseman AJ and Pielak GJ. (2015) Intracellular pH modulates quinary structure. Protein Science 24 1748-1755. RD Cohen and GJ Pielak wrote the paper.

for in-cell NMR. Here, we show that pH modulates both its stability and quinary interactions.

For proteins that undergo two-state equilibrium folding, stability is defined as the Gibb's free energy of the ensemble of unfolded forms minus that of the folded state.[10] The effects of crowding can be broken down into two types of interactions: hard-core repulsions and soft (chemical) interactions.[148] Hard-core effects arise from the physical volume occupied by macromolecules and can only stabilize globular proteins because the crowded environment favors the more compact, native conformation.[168] Soft interactions, such as quinary interactions, can be stabilizing or destabilizing.[148] For instance, charge-charge repulsions add to the effect of hard-core repulsions, shifting the equilibrium even further toward the native state. Attractive non-specific interactions between the macromolecules and the protein favor the unfolded ensemble, because the ensemble displays more attractive surface. The net effect determines the impact on stability.

Proteins contain several ionizable side chains with pK_a values between 4 and 12. Therefore, the net charge changes with pH. The pH where the net charge is zero is the isoelectric point (pI), which can be estimated from the amino acid sequence or measured by isoelectric focusing. Proteins are positively charged at pH values below the pI and negatively charged above the pI. Therefore, it is not surprising that globular protein stability is pH-dependent.[169-171] This dependence is exemplified by acid-denaturation, which results from the accumulation of localized positive charge on the compact native state that disperses upon unfolding.[172-176] Here, we use NMR-detected amide proton exchange to quantify GB1 stability in buffer as a function of pH.

Changes in protein charge also affect quinary structure. The average pI of *E. coli* proteins is ~5.[177] Therefore, most *E. coli* proteins are negatively charged at neutral pH. Since GB1 is also negatively charged at neutral pH (pI 4.5), charge-charge repulsions facilitate acquisition of high quality in-cell spectra of ^{15}N-enriched GB1.[104,178] Here, we use a GB1 variant to report the intracellular pH and illustrate the pH dependence of quinary interactions.

2.2 Results

2.2.1 GB1 stability decreases with decreasing pH

We used NMR-detected amide proton exchange[137,138] to quantify GB1 stability at the residue-level. For globular proteins, each backbone amide participates in an equilibrium between an open (op) state and a closed (cl) state. Exchange of the amide proton for a deuteron occurs only in the open state (equation 1):[179,180]

$$closed\ (N-H) \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} open\ (N-H) \xrightarrow{k_{\mathrm{int}}} open\ (N-D)$$

The rates of opening and closing are k_{op} and k_{cl}, respectively. The intrinsic rate of exchange, k_{int}, is the rate in an unstructured peptide and depends on factors such as primary structure, pH and temperature,[181,182] but k_{int} does not change in cells[183] or under crowded conditions.[158,184]. Values of k_{int} can be predicted using the computer program SPHERE.[185] The equilibrium constant for opening, $K_{op,}$ equals $k_{op}/k_{cl,}$ and the free energy of opening is:

$$\Delta G^{\circ'}_{op} = -RT\ln(K_{op})$$

Provided that k_{int} is rate determining, which we know to be true for GB1,[136,186] the equilibrium constant for opening is:

$$K_{op} = \frac{k_{obs}}{k_{\text{int}}}$$

The volume of each backbone amide nitrogen-amide proton crosspeak in ^{15}N-^{1}H HSQC spectra is proportional to its concentration. Fitting the volume-time data to an exponential decay yields k_{obs}. In this way, we obtain information about the equilibrium thermodynamic stability of individual amide bonds.[187] These individual values, however, are difficult to interpret,[188] so we focus on averages, which, for our data, approximate the free energy for global unfolding.[136,189]

To quantify the pH-dependence of GB1 stability in buffer, we performed NMR-detected amide proton exchange experiments in a buffer comprising 20 mM citrate, 150 mM NaCl (99.9% D_2O) at pH values (corrected for deuterium) of 4.4, 5.7, 6.0, 6.7 and 7.6. Spectra were serially acquired over 24 h. Data for twelve residues could be fitted at all pH values (Table 2.1). Figure 2.1 shows the opening free energies for each quantifiable residue at each pH. Figure 2.2 shows the average free energy of opening as a function of pH.

2.2.2 K10H as a pH reporter

We introduced a histidine residue at the protein surface as an in-cell pH probe.[190] HSQC spectra of the original protein and the variant in buffer are shown in Figure 2.6. The titration curve of the variant is shown in Figure 2.3. Shift changes for both the amide proton and the backbone amide nitrogen of histidine 10 were observed. We used the pH-insensitive backbone ^{1}H and ^{15}N chemical shifts of W43 as an internal reference.

The shifts of H10 nuclei minus those of W43 nuclei, $\Delta^{1}H$ and $\Delta^{15}N$, were weighted by their gyromagnetic ratios to yield a composite chemical shift[191] change:

$$\Delta\delta_{comp} = [(\Delta^{1}H\,ppm) + (\Delta^{15}N\,ppm \bullet 0.154)^2]^{\frac{1}{2}}$$

The data were fitted to the modified Henderson-Hasselbalch equation[192,193] to give an apparent pK_a of 6.7 (R^2 >0.99).

2.2.3 Intracellular pH can be controlled with buffer

The K10H variant can be used to monitor the pH inside *E. coli* cells. We compared the composite shift change of the histidine 10 resonance in cells to the titration curve (Figure 2.3) to estimate the intracellular pH (Figure 2.4). After the cells were washed and resuspended in a weak buffer (1x M9 salts pH 7.4), serial acquisition of ^{15}N-^{1}H HSQC spectra demonstrate that the pH in *E. coli* cells decreased by 0.7 units, from pH 7.4 to pH 6.7, in 6 h. However, when the cells were washed and resuspended in a stronger buffer (75 mM bis-tris propane, 75 mM HEPES, 75 mM citrate, pH 7.5), the intracellular pH dropped by less than 0.3 units. We used this buffer to monitor in cells spectra acquired at several internal pH values. Higher buffer concentrations degraded the quality of the spectra, consistent with the observation that high ionic strength decreases the quality of spectra, especially when a cold probe is used.[194]

2.2.4 pH-dependence of in-cell HSQC spectrum quality

Using the three-buffer formulation described above, we acquired in-cell ^{15}N-^{1}H HSQC spectra of the K10H variant at pH values of 7.5, 6.0, 5.0, and in diluted lysate. At pH 7.5, relatively sharp crosspeaks were observed (Figure 2.5A), comparable to spectra of GB1 in diluted lysate (Figure 2.5D). At pH 6.0, the crosspeaks broaden, and some disappear (Figure 2.5B). At pH 5.0, only a few crosspeaks are apparent (Figure 2.5C),

and these probably represent metabolites.[195] When the cells from the pH 5.0 experiments are lysed, the crosspeaks reappear (Figure 2.5D). The disappearance of crosspeaks strongly suggests that the protein has not leaked from the cells, and control experiments on supernatants indicate that the HSQC spectra represent GB1 in cells.

2.3 Discussion

We quantified the backbone amide proton exchange rates of twelve residues that exchange *via* global unfolding such that $\Delta G^{\circ'}_{op} \approx \Delta G^{\circ'}_{U}$ and interpret the average $\Delta G^{\circ'}_{op}$ value as the global free energy of unfolding for that condition.[136,189] GB1 stability increases with increasing pH over the range studied here (Figures 2.1 and 2.2).

Previously, the backbone NH of lysine 10 in wild-type GB1 has been used to sense the intracellular pH.[139] However, both the backbone ^{1}H and ^{15}N resonances of K10 follow the ionization of the glutamic acid 56 side chain and exhibit an apparent pK_a of 4.0.[196] To create a pH probe within the physiological range, we mutated lysine 10 to histidine. The mutation should be innocuous because the side chain of K10 is solvent exposed.[166] Indeed, its HSQC spectrum is nearly identical to that of wild-type GB1, except at the site of the mutation (Figure 2.6).

We used the K10H variant to monitor the pH inside cells over time. In the absence of a strong buffer the intracellular pH decreased from ~7.5 to ~6.7 in 6 h. A similar acidification of the *E. coli* cytoplasm during in-cell NMR experiments was been observed at low buffer concentration in protein probe/protein systems comprising H50 in α-synuclein[183,197] and K10 in GB1.[139]

Low buffer concentrations are usually used in NMR experiments because high ionic strength degrades spectral quality.[194] However, as shown here and previously,

[139,183,197] low ionic strength buffers do not maintain a constant pH, which can affect data interpretation. We formulated a buffer to help maintain the intracellular pH without excessively degrading the quality of HSQC ^{15}N-^{1}H spectra. The combination of 75 mM bis-tris propane (pK_a 6.8), 75 mM HEPES (pK_a 7.5) and 75 mM citrate (pK_1 3.1, pK_2 4.8, pK_3 6.4) gives a buffering range from pH 2 to nearly pH 9.

Having observed that this buffer moderates the change in intracellular pH (Figure 2.4), we used it to compare in-cell HSQC spectra of the K10H variant at different pH values (Figure 2.5). The crosspeaks broadened as the pH was decreased, and all but disappeared at pH 5.0. However, the crosspeaks reappeared in the diluted lysate, indicating that the protein is still folded.

Although we did not observe crosspeaks from inside cells at pH 5.0, there is ample evidence that GB1 remained folded in the cells at this pH. First, *in vitro* studies,[189,198] including the data shown in Figures 2.1 and 2.2, show that GB1 is stable and folded well below pH 5.0. Second, amide proton exchange data for GB1 in cells under these conditions[136] are similar to those observed in dilute solution.[166,189] Third, *E. coli* in the large intestine survive both acid pH and high packing densities,[199] the cells in the NMR tube remain viable under the conditions used here.[136] Fourth, if the protein was unfolded we would expect to observe the crosspeaks because disordered proteins yield high quality ^{15}N-^{1}H HSQC spectra due to their increased internal motion compared to globular proteins.[200,201]

Since GB1 appears to be folded in cells at the pH values studied, we conclude the broadening and disappearance of crosspeaks from the in-cell spectra at lower pH is the result of faster transverse relaxation caused by slower tumbling.[200] The width of a

crosspeak is inversely proportional to the transverse relaxation time, T2, which measures the lifetime of the coherent transverse magnetization. Slow tumbling yields shorter T2 values, giving rise to broad crosspeaks.[156,202,203] This rotational motion decreases with increasing molecular weight; however, the only parameter changed in our experiments is the pH. The increase in apparent molecular weight arises from increased protein-protein interactions. That is, the accumulation of positive charge on GB1 as the intracellular pH is decreased strengthens interactions with negatively charged *E. coli* proteins.[177] These interactions increase the effective molecular weight of GB1, thereby increasing its correlation time and degrading the quality of the HSQC spectra.

We first noted this temporal decrease in pH in cells expressing GB1 while preparing a manuscript on stability and quinary structure.[139] The observation has led us to correct our data from on stability in cells,[136] where we assumed the intracellular pH was 7.6. Specifically, in that study, k_{int} values were calculated for pH 7.6 when the internal pH was 5.8.[139] We have recalculated the $\Delta G^{\circ'}_{op}$ values in cells based on the correct internal pH (Figure 2.7).

This cytoplasmic acidification may also explain anecdotal observations about the variable quality of in-cell ^{15}N-^{1}H HSQC spectra. That is, in the absence of a strong buffer, the intracellular pH depends not only on the time between preparing the sample and acquiring the spectrum, but also the protein under investigation[183,197] and even the *E. coli* strain used to express the protein (Figure S2). An important conclusion from this work is that pH and its control are as important for in-cell NMR spectroscopy as they are for *in vitro* studies.

2.4 Summary

In summary, this work highlights the importance of quinary interactions in furthering our understanding of proteins inside living cells. Studies in dilute solution demonstrate that fluctuations in pH modulate protein properties; this effect grows even more complex in the dense cellular environment. The ability to maintain cells at a desired pH extends beyond NMR spectral quality and is essential to unlock the secrets of protein behavior.[199,204,205]

2.5 Materials and Methods

2.5.1 Expression, purification and NMR-detected H/D exchange

The pET-11a plasmid with the gene encoding GB1 T2Q has been described.[136] The Agilent Quikchange mutagenesis kit was used with the following primers to produce the K10H variant: forward 5' C CTG AAC GGT CAT ACC CTG AAA GGT GAA ACC ACC 3', reverse 5' GGT GGT TTC ACC TTT CAG GGT ATG ACC GTT CAG G 3'. The mutation was confirmed by DNA sequence analysis (GeneWiz, Research Triangle Park, NC).

Expression, purification and NMR-detected amide proton exchange experiments were performed as described.[136] A timer was initiated when lyophilized GB1 was resuspended to a final concentration of 1 mM in 500 µL of buffer prepared in 99.9% D_2O. pH was corrected for the isotope effect. For glass electrodes in D_2O, pH = pH_{read} + 0.4.[206] ^{15}N-^{1}H HSQC spectra were acquired serially over 24 h on a Varian Inova 600 MHz spectrometer at 37 ^{o}C. Each 20-min experiment comprised 64 increments in the ^{15}N dimension with eight scans per increment. Data were processed *via* nmrPipe. Peak volumes were obtained by using NMRViewJ.[207,208] Volumes were plotted as a function

of time and fit to an exponential decay by using SigmaPlot to yield k_{obs} values (SigmaPlot).

2.5.2 Titration curve

The lyophilized K10H variant was resuspended to a final concentration of 500 μM in 50 mM bis-tris propane, 50 mM citrate, 50 mM borate, 50 mM HEPES, 5% D_2O at twelve pH values (5.1, 5.5, 6.0, 6.2, 6.4, 6.7, 6.9, 7.3, 7.7, 8.2, 8.6, 8.9). HSQC spectra were acquired at 298 K on a 700 MHz Bruker Avance III HD spectrometer equipped with a Bruker TCI cryoprobe. Sweep widths were 9600 Hz and 1950 Hz in the ^{1}H and ^{15}N dimensions, respectively. Each spectrum comprised 64 increments in the ^{15}N dimension with eight scans per increment. Total acquisition time was ~ 20 min.

2.5.3 In-cell NMR

E. coli BL21(DE3) cells containing the plasmid encoding GB1 K10H were grown in 200 mL of M9 minimal media.[136] Chloramphenicol was added to a final concentration of 50 μg/mL to halt expression, and the cells were harvested by centrifugation (1,000*g*, 20 min, 4 °C). Buffers comprised either 1x M9 salts (48 mM Na_2HPO_4, 22 mM KH_2PO_4, 9 mM NaCl, pH 7.4) or 75 mM bis-tris propane/75 mM HEPES/75 mM citrate at various pH values. The cells were resuspended and washed with 30 mL of buffer, followed by centrifugation. After three wash/spin cycles, the pellet was resuspended in 50 μL of buffer and transferred to an NMR tube. Spectra were acquired on the 600 MHz spectrometer using sweep widths of 12,000 Hz and 2,500 Hz in the ^{1}H and ^{15}N dimensions, respectively. ^{15}N-^{1}H HSQC spectra were acquired serially for six h. Each spectrum comprised 64 increments in the ^{15}N dimension with eight scans per increment.

After the experiment, the sample was centrifuged, and a spectrum of the supernatant was obtained to ensure that the signal came from protein inside cells.

2.6 Figures

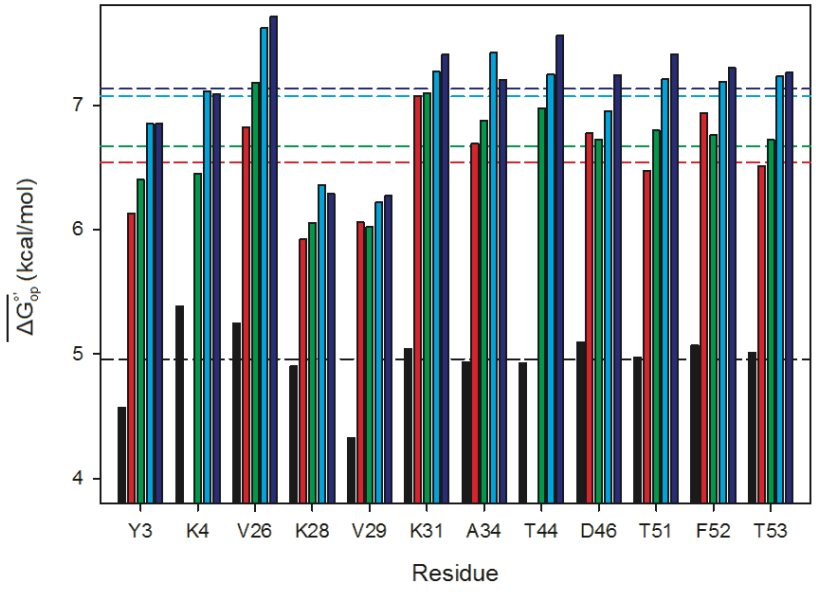

Figure 2.1 GB1 stability increases with increasing pH. $\Delta G^{\circ'}_{op}$ values at 37 °C for wild-type GB1 at pH 4.4 (black), 5.7 (red), 6.0 (green), 6.7 (magenta) and 7.6 (blue). The horizontal lines represent the average $\Delta G^{\circ'}_{op}$ at each pH value.

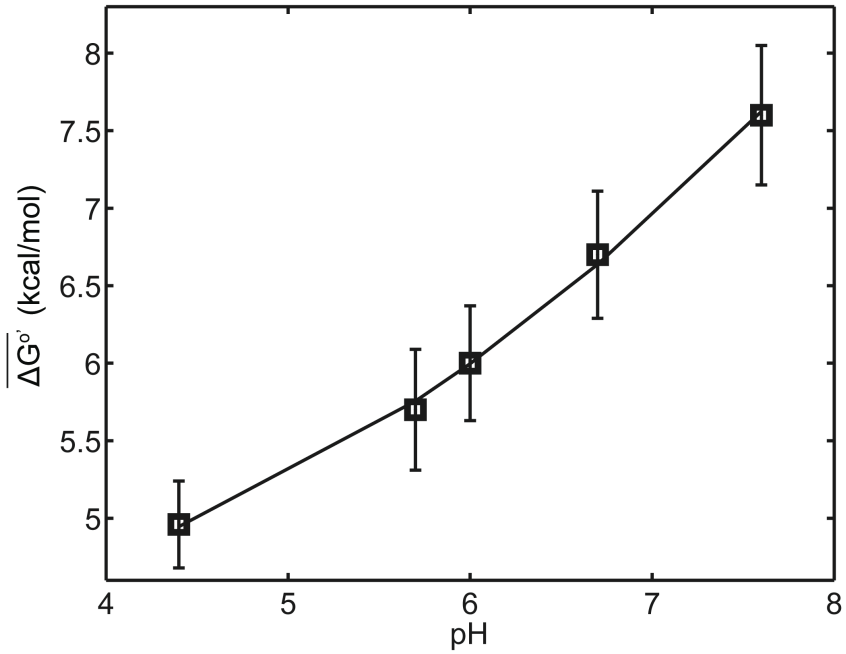

Figure 2.2. Plot of average $\Delta G^{\circ'}_{op}$ versus pH from the data in Figure 1.

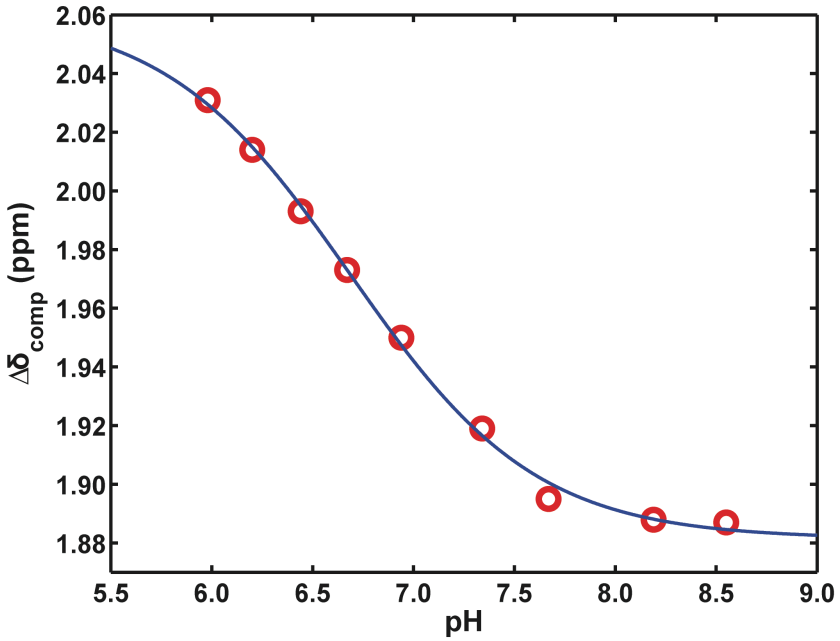

Figure 2.3. Titration curve from the composite chemical shift[191] change ($\Delta\delta_{comp}$) of the H10 resonances at 37 °C. The solid curves are fits to the modified Henderson-Hasselbalch equation[192,193] to yield a pK_a of 6.7.

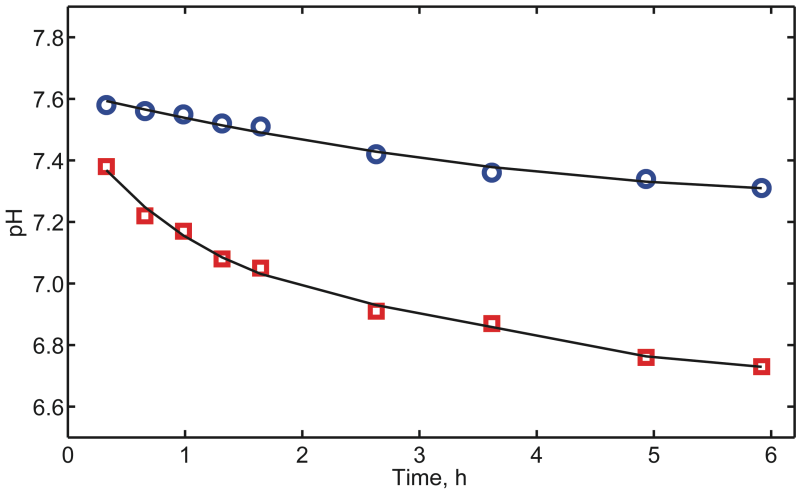

Figure 2.4. Change in intracellular pH over time measured using the K10H variant in 1x M9 salts (initial pH pH 7.4, red) and in 75 mM bis-tris propane/75 mM HEPES/75 mM citrate (initial pH 7.5, blue). Values represent the composite corrected chemical shift. pH values were extrapolated from the data in Figure 3.

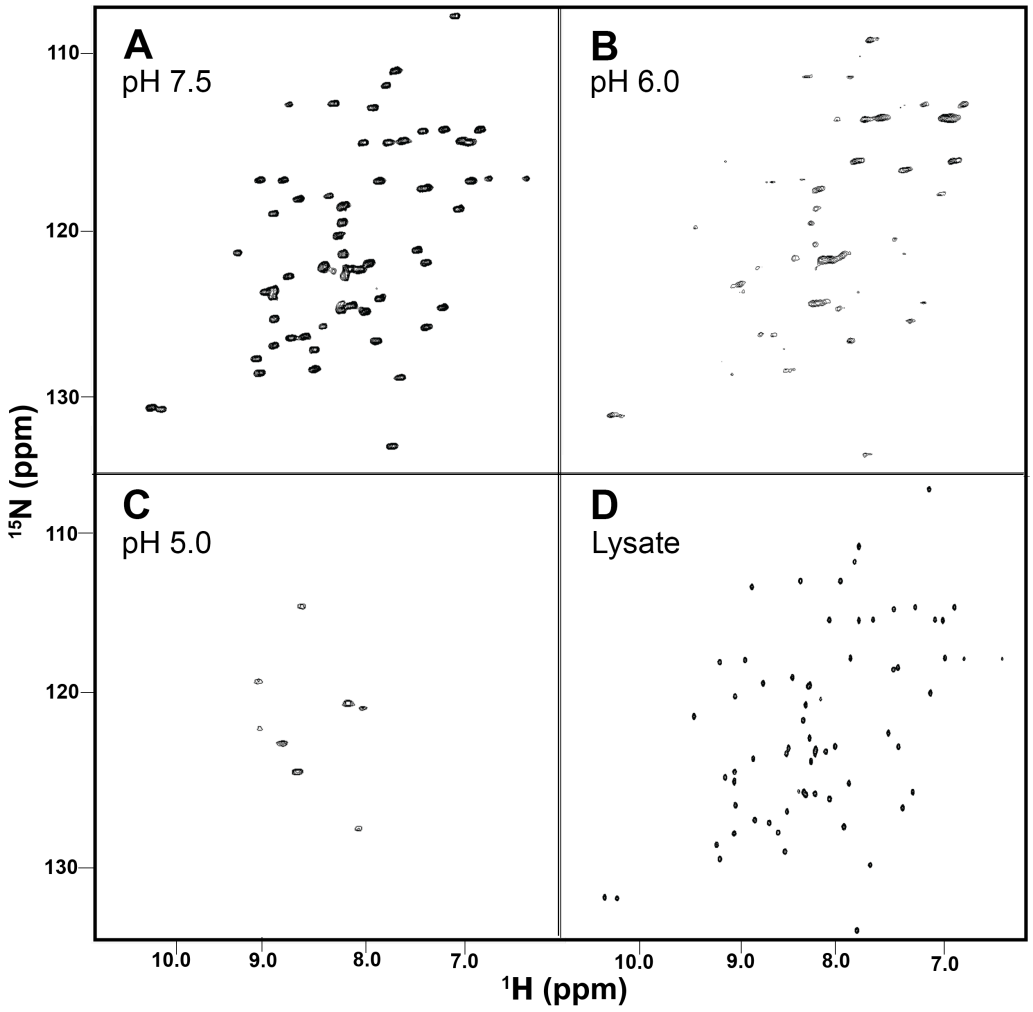

Figure 2.5. In-cell ^{15}N $-^{1}H$ HSQC spectra of K10H GB1. Cells were washed and resuspended in 75 mM bis-tris propane/HEPES/citrate at 37 °C at pH 7.5 (A), 6.0 (B) and 5.0 (C). The cells from panel C were lysed and the resultant HSQC spectrum (pH 5.0) is shown (D).

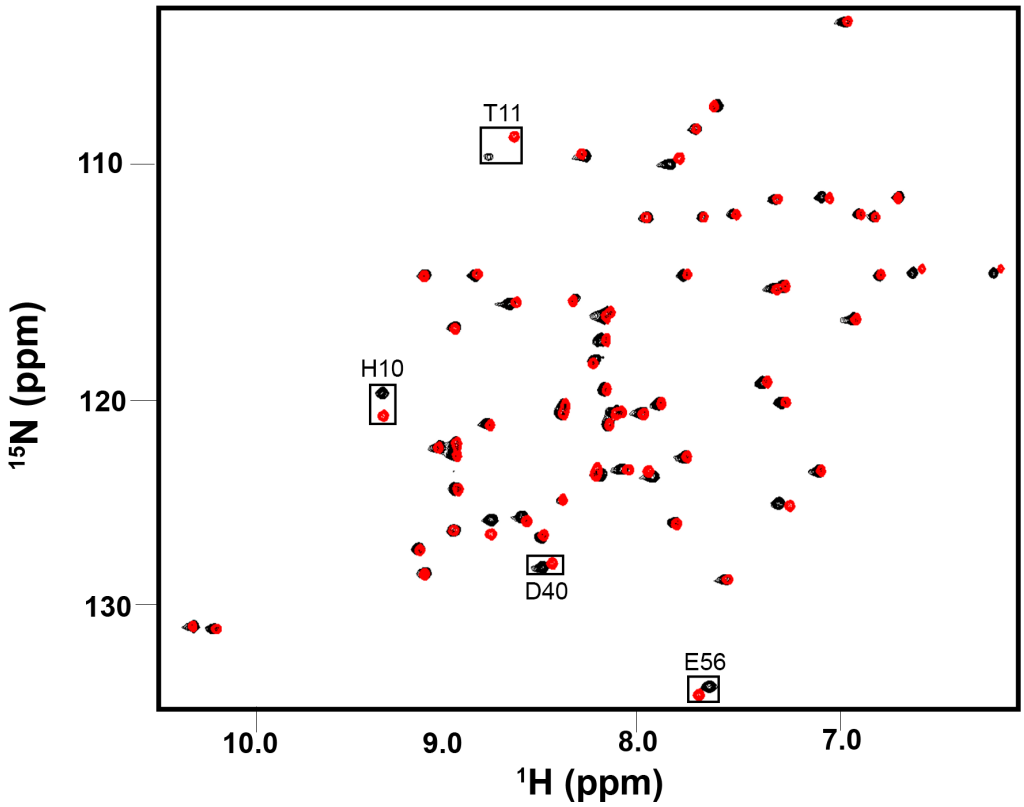

Figure 2.6. ^{15}N-^{1}H HSQC spectra of wildtype GB1 (red) and K10H (red). Lyophilized protein was resuspended in 50 mM citrate, 50 mM HEPES, 50 mM bis-tris propane, 50 mM borate, 5% D2O, pH 7.4 to a final concentration of 500 μM. Spectra were acquired (37 °C) on a Varian Inova 600 MHz spectrometer using sweep widths of 12,000 Hz and 2,500 Hz in the ^{1}H and ^{15}N dimensions, respectively. Each spectrum comprised 64 increments in the 15N dimension with eight scans per increment. Residues that display changes in chemical shift in the K10H variant are boxed. D40 and E56 are involved in a hydrogen-bonding network with residue 10,[196] and T11 is the adjacent residue.[166]

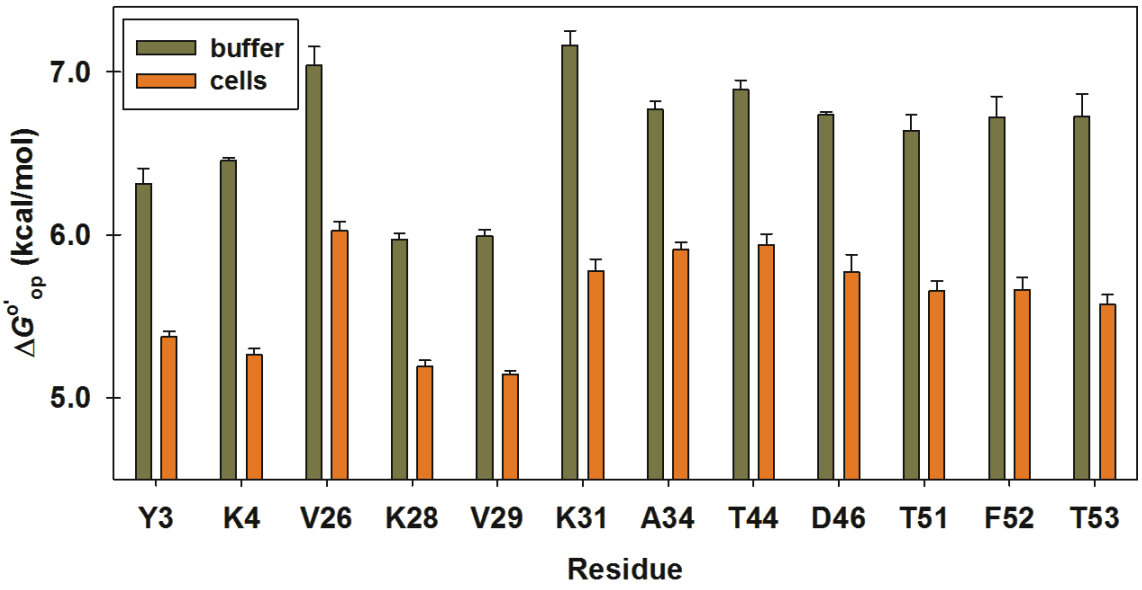

Figure 2.7 $\Delta G^{\circ'}_{op}$ values for wild-type GB1 in buffer and in cells at 37 °C (PBS, 95% D2O, pH 5.8). Error bars represent the standard deviation of the mean from three trials.

2.7 Tables

Table 2.1 Backbone amide proton exchange rates (k_{obs}, s^{-1}) for GB1 in buffer (20 mM citrate, 150 mM NaCl, 37 °C) at pH 4.4, 5.7, 5.8, 6.0, 6.7. Rates for GB1 at pH 7.6 have been reported.[136]

residue	$k_{obs,\ pH\ 4.4}$	$k_{obs,pH\ 5.7}$	$k_{obs,\ pH\ 5.8}$	$k_{obs,\ pH\ 6.0}$	$k_{obs,\ pH\ 6.7}$
Y3	2.23E-05	3.40E-05	2.65E-05	4.35E-05	8.75E-05
K4	7.10E-06	--	2.93E-05	4.87E-05	1.81E-04
V26	1.35E-05	2.06E-05	1.54E-05	2.29E-05	3.21E-05
K28	1.94E-05	3.90E-05	4.51E-05	6.04E-05	1.98E-04
V29	1.14E-05	1.18E-05	1.86E-05	2.49E-05	1.08E-04
K31	1.27E-05	9.00E-06	7.40E-06	1.74E-05	3.58E-05
A34	1.62E-05	1.80E-05	2.02E-05	2.65E-05	5.86E-05
T44	9.90E-06	--	1.09E-05	1.34E-05	2.34E-05
D46	2.18E-05	9.00E-06	1.20E-05	1.83E-05	3.46E-05
T51	1.52E-05	2.57E-05	2.32E-05	3.00E-05	3.89E-05
F52	1.06E-05	9.80E-06	2.54E-05	2.62E-05	3.33E-05
T53	1.25E-05	2.10E-05	1.25E-05	2.98E-05	3.51E-05

CHAPTER 3: ELECTROSTATIC CONTRIBUTIONS TO PROTEIN QUINARY STRUCTURE[1]

3.1 Introduction

Protein 'quinary structure' was independently described three times: by Vaĭnshteĭn in 1973, by Edelstein in 1980 and by McConkey in 1982.[209-211] The latter contribution motivated the work described here. McConkey observed that evolutionarily distant protein homologs have similar overall charges; this is completely unexpected if their exteriors need only be hydrophilic, and suggests that the cellular interior is highly organized. In 1983, Srere consolidated numerous findings that Krebs' cycle enzymes can be isolated together and coined the term metabolon to describe the organizing effect of quinary interactions.[212] The transient nature of quinary interactions allows cells to alter their metabolism in response to the environment and is essential for survival.[213] However, there is no quantitative information regarding the strength of the interactions that comprise quinary structure because they can only be studied in living cells. Here, we quantify the potential contribution of charge-charge interactions.

Efforts to understand protein behavior under physiologically relevant conditions began with studies using high concentrations of uncharged synthetic polymers.[214,215] These results were interpreted mostly in terms of hard-core repulsions, which occur because two atoms cannot occupy the same space at the same time.[168] As expected, these polymer solutions often stabilize proteins.[216] Synthetic polymers, however, do not

[1]The material for this chapter has been published in JACS. The original citation is as follows: Cohen RD, Pielak GJ. (2016) Electrostatic contributions to protein quinary structure *J Am Chem Soc* **138** 13139-13142. RD Cohen and GJ Pielak wrote the paper.

adequately mimic the cellular interior, because biologically relevant crowding molecules also interact chemically with proteins.[147,158,217,218] These so-called soft interactions, which define quinary structure, modulate the effect of hard-core repulsions and can stabilize or destabilize proteins, depending on whether interactions between the crowder and the test protein are repulsive or attractive, respectively.[148]

The B1 domain of streptococcal protein G (GB1, 6.2 kDa, pI 4.6)[166,219] is the quintessential test protein to probe the potential of quinary structure in *E. coli* because GB1 has been extensively characterized, it is a two-state folder and it is not native to *Escherichia coli,* which minimizes the likelihood of specific interactions.[189,196,220,221] In addition, GB1 can be studied in cells using NMR because GB1 and the majority of proteins in *E. coli* are polyanions at physiological pH,[177] and the consequent net charge-charge repulsions facilitate the acquisition of high quality in-cell HSQC spectra.[104]

There are no histidines in wild-type GB1. By installing a histidine at position 10 (K10H GB1) to measure the intracellular pH, we developed a buffer to control the intracellular pH of *E. coli,* and showed that the cytosolic pH affects the quality of in-cell ^{15}N-^{1}H HSQC spectra.[140] Specifically, as the pH is decreased, the accumulation of positive charge on the surrounding *E. coli* proteins increases the attractive interactions with polyanionic GB1, slowing GB1 tumbling and broadening its crosspeaks into the background. These results qualitatively demonstrated that the intracellular pH modulates quinary structure. Here, we used NMR-detected amide proton exchange to quantify the pH-dependence of K10H GB1 stability in buffer and in cells.

Protein stability, $\Delta G_U^{\circ'}$, is the Gibb's free energy of the unfolded (U) state minus that of the folded (F) state, such that

$$\Delta G_U^{\circ'} = -RT \ln\left(K_U\right) = -RT \ln\left(\frac{[U]}{[F]}\right)$$

where R is the universal gas constant and T is the absolute temperature.

Amide proton exchange is a powerful tool for quantifying protein stability in *vitro*[222,223] and in living cells.[136,139,224] For many proteins, the free energy required to open the protein and expose a particular amide proton to solvent, $\Delta G_{op}^{\circ'}$, can be determined by dividing the observed rate of exchange, k_{obs}, by the intrinsic rate of exchange, k_{int}, such that

$$\Delta G_{op}^{\circ'} = -RT \ln\left(\frac{k_{obs}}{k_{int}}\right)$$

One might expect all amide protons that are exposed only upon complete unfolding to give the same $\Delta G_{op}^{\circ'}$. However, this is not the case because of the inherent uncertainty in k_{int} values, which are derived from model peptides, not the particular protein being studied. Nevertheless, the method is valid because exchange data from 20 proteins show that global unfolding residues yield $\Delta G_{op}^{\circ'}$ values within 1 kcal/mol of $\Delta G_U^{\circ'}$ from thermal or co-solute denaturation.[225] Most importantly for this study, the method is valid for GB1 because stability measurements from amide proton exchange data have been confirmed by differential scanning calorimetry.[136,139]

We define the strength of quinary interactions, $\Delta\Delta G_{quin}^{\circ'}$, as the stability of GB1 in cells minus its stability in buffer at the same pH and temperature (37 °C).

$$\Delta\Delta G_{quin}^{\circ'} = \Delta G_{op,cells}^{\circ'} - \Delta G_{op,buff}^{\circ'}$$

Negative $\Delta\Delta G_{quin}^{\circ'}$ values reflect protein destabilization in cells and suggest an increase in attractive quinary interactions.

3.2 Results

We showed previously that replacing a negatively charged residue with more positively charged amino acids destabilizes GB1 in cells in a charge-dependent manner.[139] Other groups have shown that protein surface charge plays a role in quinary interactions.[104,105,118] To further investigate the role of electrostatic interactions in quinary structure without perturbing the protein sequence, we quantified GB1 stability at pH 7.4, 6.0 and 5.0.

At pH 7.4, K10H GB1 and the majority of *E. coli* proteins (pI < 7)[177] are polyanions. K10H GB1 remains polyanionic between pH 7.4 and 5.0, while many *E. coli* proteins change from polyanions to polycations as the pH is decreased below their pI. *E. coli* are known to survive these slightly acidic conditions.[199,226] We hypothesized (Figure 3.1) that the accumulation of positive charge on *E. coli* proteins as the cytosolic pH is lowered would preferentially increase attractive charge-charge interactions with the unfolded ensemble of GB1, because the unfolded protein has more accessible surface, and thus decrease GB1 stability in cells compared to buffer.

The number of quantifiable residues is limited by crosspeak overlap and large k_{obs} values. We were able to quantify $\Delta G^{\circ'}_{op}$ for the twelve residues that have k_{obs} values of < 9.7 x 10^{-4} s^{-1}. Quantifiable residues are well distributed in the secondary structure: β1 (Y3, K4), α1 (A26, K28, V29, K31, A34), β3 (T44, D46) and β4 (T51, F52, T53 These twelve residues are known to exchange upon complete GB1 unfolding; therefore, the mean $\Delta G^{\circ'}_{op}$ approximates $\Delta G^{\circ'}_{U}$.[136,189] Backbone amide protons in β2 exchange too quickly to quantify.[189] At pH 7.4, we can quantify $\Delta G^{\circ'}_{op}$ for K4, A26, T51 and T53 in buffer, but the crosspeaks decay too quickly for quantification in cells. Hydrogen

exchange is base-catalyzed,[223] and we attribute the limited data and the larger uncertainties at the highest pH to faster exchange.

At pH 7.4, the average $\Delta G^{\circ'}_{op}$ in buffer is 6.80 ± 0.07 kcal/mol (Figure 3.4), where the uncertainty is the standard deviation of the mean. This value is approximately equal to that in cells, 6.65 ± 0.09 kcal/mol (Figure 3.4). The $\Delta\Delta G^{\circ'}_{quin}$ values (Figure 3.2) are modestly positive or negative. Their average value, 0.1 ± 0.1 kcal/mol (the uncertainty is the standard deviation of the mean), is insignificant, indicating that the net interaction between GB1 and the cellular milieu is negligible at this pH. Lowering the pH to 6.0 decreases the stability of GB1 in buffer and cells (Figure 3.5), but the destabilization is more dramatic in cells; the average $\Delta\Delta G^{\circ'}_{quin}$ is -0.80 ± 0.05 kcal/mol. At pH 5.0, the destabilization is even more dramatic in cells (Figure 3.6), and the average $\Delta\Delta G^{\circ'}_{quin}$ is -1.12 ± 0.06 kcal/mol. These data indicate a significant role for electrostatic interactions in quinary structure (Figure 3.3).

3.3 Discussion

Studies of globular proteins have long emphasized the requirement for a well-packed and hydrophobic interior[227] with nearly complete formation of internal hydrogen bonds.[228] Despite the fact that these studies were conducted in simple buffered solutions, they provided physiologically relevant information about protein interiors because interior atoms experience the same environment in cells and in buffer: they are surrounded by atoms belonging to that same protein.

Despite the ideas of McConkey,[211] interest in protein surfaces remained stagnant: defining them as hydrophilic was sufficient. Our data indicate that exteriors are as important as interiors for understanding protein chemistry in cells, but this

physiologically-relevant information was hidden because studies were performed in simple buffered solutions. The key idea is that exterior atoms are exposed to mostly water in simple buffered solutions, but in cells, these atoms are exposed to the complex, crowded and dynamic cytoplasm.[229,230] There is now an emerging realization that exteriors play an important role in cells by forming the transient interactions that comprise protein quinary structure.[104,105,117,165]

Quinary interactions, by definition, are absent in dilute solutions, but they play an important role in the dense cellular environment. Traditional crowding theory predicts that the cellular environment should stabilize proteins strictly due to the effects of hard-core repulsions. Although such steric effects must be present, they are modulated by transient chemical interactions that either reinforce or oppose them.[147,148,160,231] Several studies[139,146,147] have shown that the cellular interior destabilizes proteins, but the nature of quinary interactions remains mostly unchartered territory.

As shown here, at pH 7.4, the stabilizing effect of hard-core repulsions is balanced by attractive quinary interactions to the extent that GB1 stability is approximately equal in cells and in buffer. In other words, near neutral pH, attractive interactions between proteins are tempered because both GB1 and the majority of *E. coli* proteins are negatively charged.[177] In buffer, GB1 stability decreases by 1.3 kcal/mol when the pH is lowered from 7.4 to 5.0. The pH-induced destabilization, however, is much more dramatic in cells: GB1 is destabilized by almost 2.5 kcal/mol when the pH in cells is decreased from 7.4 to 5.0. The extent to which pH modulates quinary structure is remarkable (Figure 3.3), causing a 30% decrease in stability.

Additional sources of the stability changes in cells compared to buffer must also be considered. For instance, the pH changes in cells could unfold other proteins or change the unfolded state ensemble of GB1, exposing additional polar and hydrophobic groups. Such interactions,[139] and changes in metabolite ionization, may also contribute to our observations. Changes in chaperone activity are not likely to contribute because as enzymes, chaperones do not affect equilibria, which are the basis for the measurements described here.

3.4 Summary

Ultimately, the changes in interactions that alter GB1 stability in cells compared to buffer arise from manipulating the intracellular pH. In buffer, protein stability is governed by hydrophobic interactions and specific intramolecular interactions.[228,232] Despite their predicted importance,[209-211] the role of protein surfaces, with few exceptions,[233,234] has been overlooked, and the contribution of intermolecular interactions in cells has been ignored. Our results provide quantitative evidence that charge-charge interactions are important factors in quinary structure. In summary, surfaces cannot be ignored when proteins are studied in their native environment because there is more to protein stability than a well-packed, hydrophobic core.

3.5 Materials and Methods

The pET-11a plasmid harboring the K10H variant,[235] the isolation and purification of GB1 [140,166] and the protocol for NMR-detected amide proton exchange in cells and in buffer have been described.[136] The construct also carries the T2Q variant to prevent N-terminal deamidation.[235] For dilute solution experiments, 2.2 mg of lyophilized protein

was resuspended in 500 µL of 75 mM HEPES/75 mM bis-tris propane/75 mM citrate, 99.9% D_2O at the desired pH, and ^{15}N-^{1}H HSQC spectra were acquired serially.

The quenched lysate method is an established approach for quantifying protein stability in cells.[136,224] Briefly, expression of the K10H variant was induced with 1 mM (final concentration) of isopropyl-β-D-thiogalactopyranoside when *E. coli* cells reached an optical density of 0.6 at 600 nm. After 2 h, chloramphenicol was added to stop protein expression and the cells were harvested by centrifugation. A timer initiated after resuspension in the D_2O-containing buffer. Individual aliquots were removed at discrete time points and hydrogen exchange was quenched while cells were simultaneously lysed by vortexing with glass beads.[136] After centrifugation, the supernatant was transferred to an NMR tube and a spectrum acquired. At pH 7.4 and pH 6.0, aliquots were removed within 2 h after exchange was initiated. At pH 5.0, aliquots were removed ~ 1 h, 2 h, 3 h, 8 h and 15 h after initiating exchange.

Exchange rates were converted to $\Delta G_{op}^{\circ'}$ values using intrinsic rates of exchange from SPHERE (37 °C, alanine oligopeptide basis).[182,185] Figures 3.4-3.6 show $\Delta G_{op}^{\circ'}$ values as a function of residue number at each pH value and Tables 3.1-3.3 list k_{obs} and $\Delta G_{op}^{\circ'}$ values. The stability measurements were confirmed by differential scanning calorimetry using a Microcal VP-DSC calorimeter (Table 3.4). For calorimetry, the temperature was increased from 20 °C to 95 °C at a rate of 60 °C/h for three scans to yield $\Delta H_{cal}^{\circ'}$ and T_m. Analysis was performed as described by Becktel and Schellman[236] to obtain $\Delta G_D^{\circ'}$.

3.6 Figures

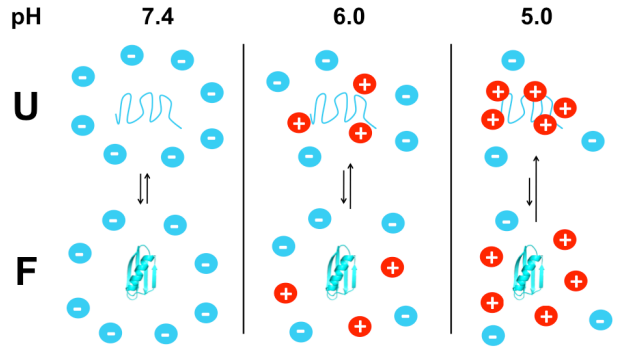

Figure 3.1. Interactions between E. coli proteins (red and blue circles) and the unfolded (U) and folded (F) states of GB1 (PDB ID 3PGB) as a function of pH. As the pH is decreased, *E. coli* proteins become more positively charged, and GB1 is destabilized. The relative lengths of the arrows are not to scale; $\Delta G_U^{\circ'}$ is > 0 at all three pH values (Figures 3.4−3.6).

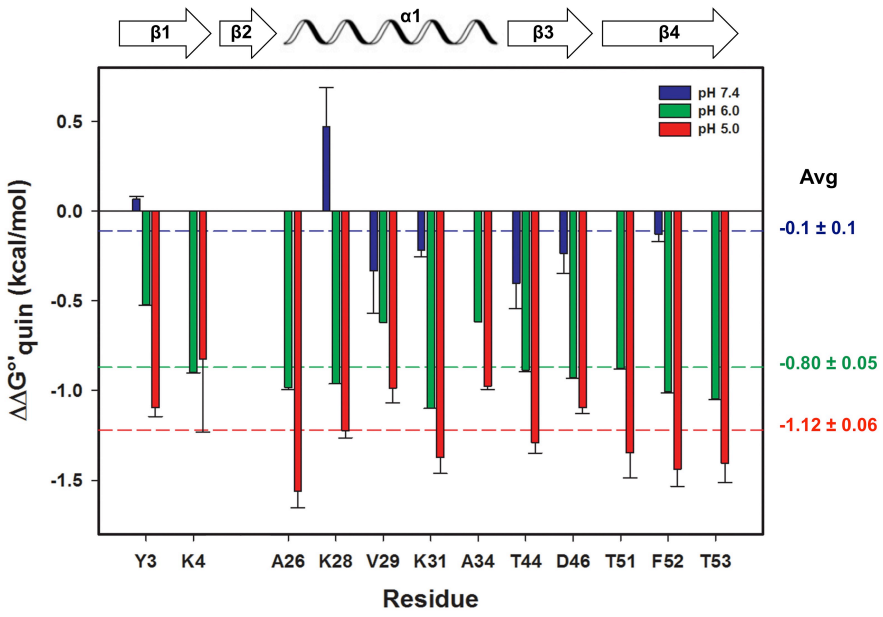

Figure 3.2. $\Delta\Delta G^{\circ'}_{quin}$ values for K10H GB1 at pH 7.4 (blue), pH 6.0 (green) and pH 5.0 (red) at 37 °C. Error bars represent the uncertainties propagated from triplicate (dilute solution) and duplicate (cells) experiments. Dashed lines represent the average at each pH. The uncertainties are the standard deviation of the mean.

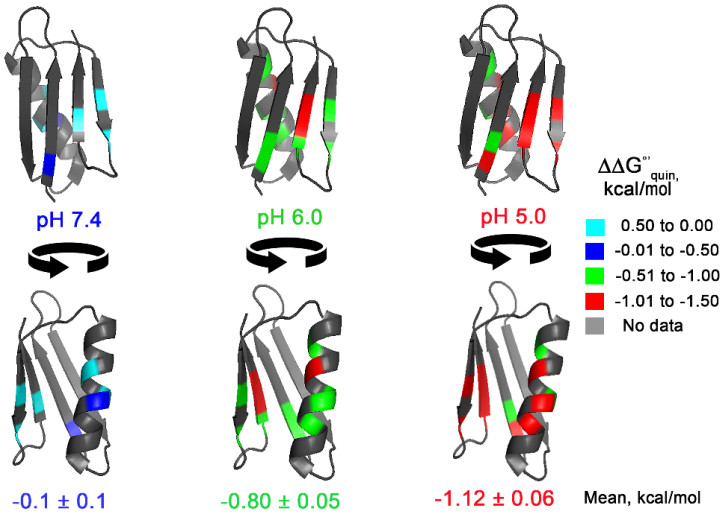

Figure 3.3. Structure of GB1 (PDB ID 3PGB) with residues colored by the magnitude of quinary interactions at each pH value.

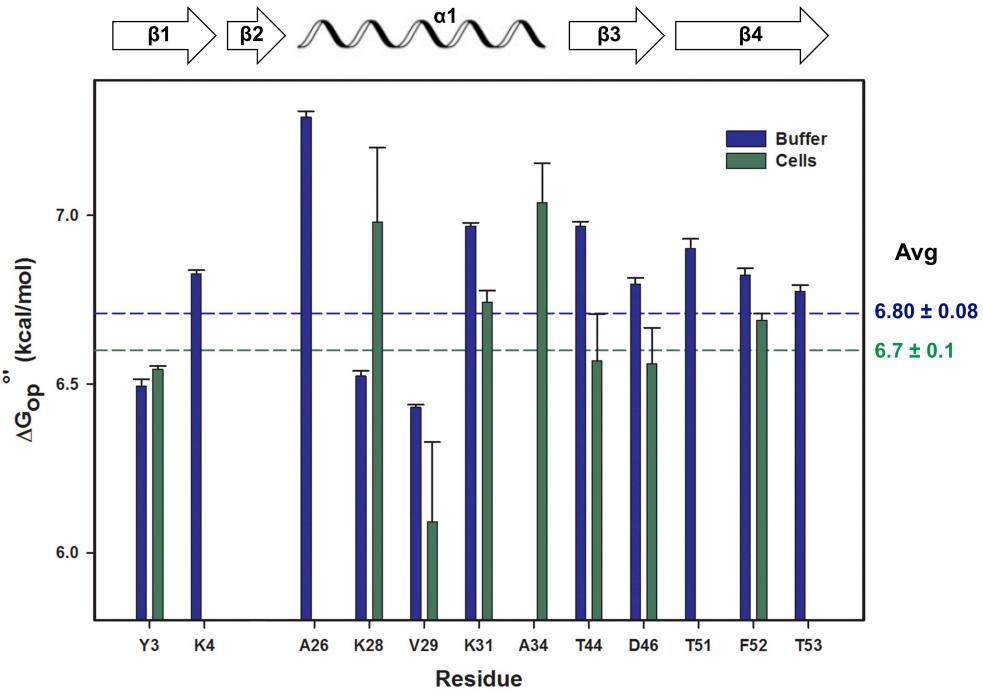

Figure 3.4. $\Delta G^{\circ'}_{op}$ values for GB1 K10H in cells (green) and buffer (blue) at pH 7.4. Error bars represent the standard deviation of the mean for two (cells) and three (buffer) trials. The dashed line indicates the average $\Delta G^{\circ'}_{op}$ value for each condition.

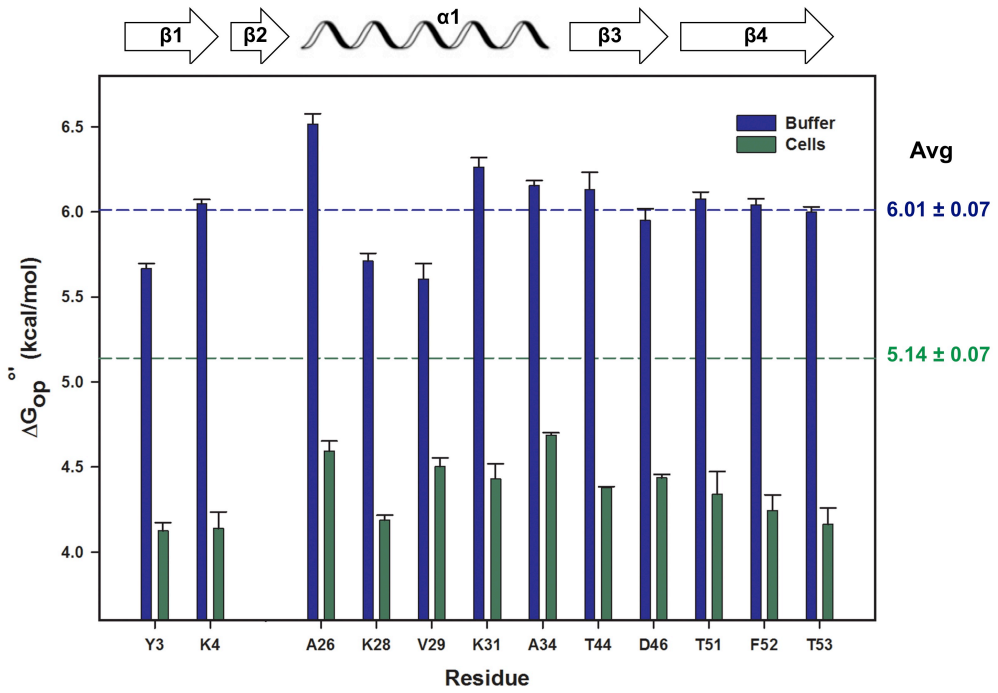

Figure 3.5. $\Delta G^{\circ'}_{op}$ values for GB1 K10H in cells (green) and buffer (blue) at pH 6.0. Error bars represent the standard deviation of the mean for two (cells) and three (buffer) trials. The dashed line indicates the average $\Delta G^{\circ'}_{op}$ value for each condition.

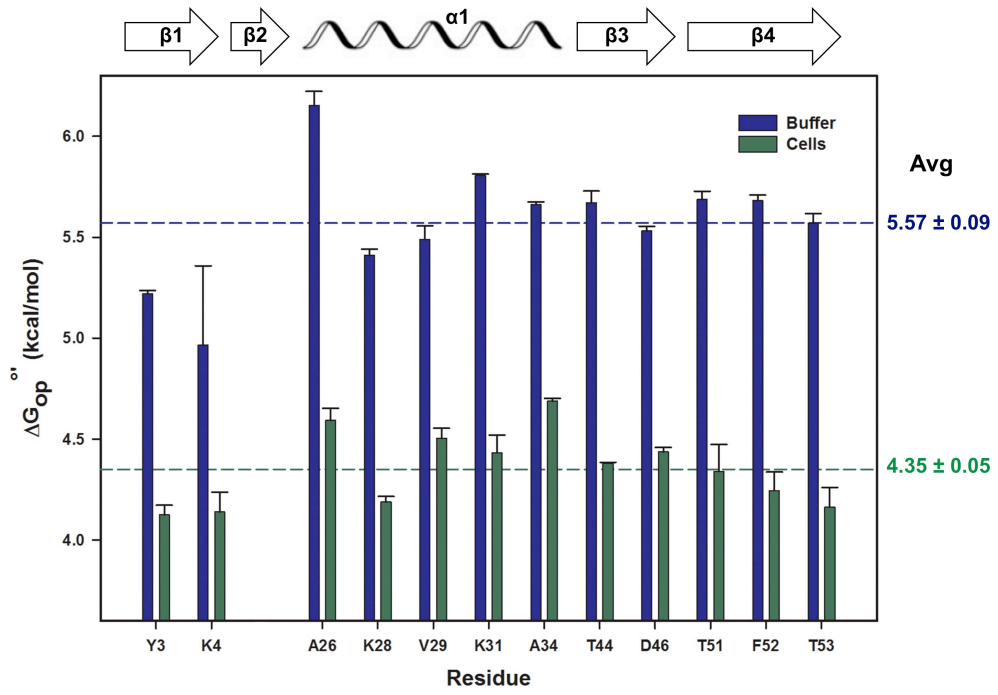

Figure 3.6. $\Delta G^{\circ'}_{op}$ values for GB1 K10H in cells (green) and buffer (blue) at pH 5.0. Error bars represent the standard deviation of the mean for two (cells) and three (buffer) trials. The dashed line indicates the average $\Delta G^{\circ'}_{op}$ value for each condition.

3.7 Tables

Table 3.1. Backbone amide proton exchange rates (k_{obs}, s^{-1}) and corresponding $\Delta G^{\circ'}_{op}$ (kcal/mol) values for K10H GB1 in cells and in buffer (75 mM HEPES, 75 mM bis-tris propane, 75 mM citrate) at pH 7.4, 37°C. pH is corrected for the isotope effect. For glass electrodes in D_2O, pH = pH_{read} + 0.4.[206] [a]SDM: standard deviation of the mean from three trials. [b] $\Delta\Delta G^{\circ'}_{quin} = \Delta G^{\circ'}_{op,cells} - \Delta G^{\circ'}_{op,buff}$. [c]From propagation of error.

residue	$k_{obs,cells}$	SDM[a]	$k_{obs,dilute}$	SDM[a]
Y3	8.5E-04	±1E-05	9.3E-04	±3E-05
K4	--	--	6.5E-04	±1E-05
A26	--	--	4.7E-04	±1E-05
K28	3E-04	±1E-04	6.7E-04	±2E-05
V29	6E-04	±2E-04	3.20E-04	±4E-06
K31	7.6E-04	±4E-05	5.30E-04	±9E-06
A34	5E-04	±1E-04	--	--
T44	7E-04	±1E-04	3.30E-04	±7E-06
D46	6E-04	±1E-04	3.8E-04	±1E-05
T51	--	--	6.3E-04	±3E-05
F52	7.2E-04	±2E-05	5.8E-04	±2E-05
T53	--	--	6.8E-04	±2E-05

residue	$\Delta G^{\circ'}_{op,cells}$	SDM[a]	$\Delta G^{\circ'}_{op,buff}$	SDM[a]	$\Delta\Delta G^{\circ'}_{quin}$ [b]	uncertainty[c]
Y3	6.54	±0.01	6.49	±0.02	0.07	±0.02
K4	--	--	6.83	±0.01	--	--
A26	--	---	7.29	±0.02	--	--
K28	7.0	±0.2	6.52	±0.01	0.5	±0.2
V29	6.1	±0.2	6.43	±0.01	-0.3	±0.2
K31	6.74	±0.04	6.97	±0.01	-0.22	±0.04
A34	7.0	±0.1	--	--	--	--
T44	6.6	±0.1	6.97	±0.01	-0.4	±0.1
D46	6.6	±0.1	6.80	±0.02	-0.2	±0.1
T51	--	--	6.90	±0.03	--	--
F52	6.69	±0.02	6.82	±0.02	-0.13	±0.03
T53	--	--	6.77	±0.02	--	--

Table 3.2. Backbone amide proton exchange rates (kobs, s^{-1}) and corresponding $\Delta G^{\circ'}_{op}$ (kcal/mol) values for K10H GB1 in cells and in buffer (75 mM HEPES, 75 mM bis-tris propane, 75 mM citrate) at pH 6.0, 37°C. pH is corrected for the isotope effect. For glass electrodes in D_2O, pH = pH_{read} + 0.4.[206] [a]SDM: standard deviation of the mean from three trials. [b] $\Delta\Delta G^{\circ'}_{quin} = \Delta G^{\circ'}_{op,cells} - \Delta G^{\circ'}_{op,buff}$. [c]From propagation of error.

residue	$k_{obs,cells}$	SDM	$k_{obs,dilute}$	SDM
Y3	3.4E-04	±6E-05	1.40E-04	±7E-06
K4	4.4E-04	±9E-05	1.00E-04	±5E-06
A26	3E-04	±1E-04	6.60E-05	±7E-06
K28	4.9E-04	±2E-05	1.03E-04	±8E-06
V29	1.3E-04	±1E-05	4.9E-05	±8E-06
K31	3.9E-04	±2E-05	6.6E-05	±6E-06
A34	2.3E-04	±1E-05	8.4E-05	±5E-06
T44	2.2E-04	±3E-05	5E-05	±1E-05
D46	2.9E-04	±2E-05	6.4E-05	±8E-06
T51	4.0E-04	±4E-05	9.5E-05	±7E-06
F52	4E-04	±1E-04	8.2E-05	±5E-06
T53	5.3E-04	±8E-05	9.4E-05	±5E-06

residue	$\Delta G^{\circ'}_{op,cells}$	SDM[a]	$\Delta G^{\circ'}_{op,buff}$	SDM[a]	$\Delta\Delta G^{\circ'}_{quin}$ [b]	uncertainty[c]
Y3	5.2	±0.1	5.67	±0.03	-0.52	±0.003
K4	5.2	±0.1	6.05	±0.03	-0.90	±0.003
A26	5.5	±0.2	6.52	±0.06	-0.98	±0.01
K28	4.75	±0.02	5.71	±0.05	-0.96	±0.001
V29	4.99	±0.01	5.61	±0.09	-0.62	±0.001
K31	5.17	±0.02	6.26	±0.05	-1.10	±0.001
A34	5.54	±0.01	6.15	±0.03	-0.62	±0.003
T44	5.25	±0.08	6.1	±0.1	-0.89	±0.008
D46	5.02	±0.04	5.95	±0.07	-0.93	±0.003
T51	5.20	±0.07	6.08	±0.04	-0.88	±0.003
F52	5.0	±0.2	6.04	±0.03	-1.01	±0.007
T53	4.95	±0.09	6.00	±0.03	-1.05	±0.003

Table 3.3. Backbone amide proton exchange rates (k_{obs}, s^{-1}) and corresponding $\Delta G^{\circ'}_{op}$ (kcal/mol) values for K10H GB1 in cells and in buffer (75 mM HEPES, 75 mM bis-tris propane, 75 mM citrate) at 5.0, 37°C. pH is corrected for the isotope effect. For glass electrodes in D_2O, pH = pH_{read} + 0.4.[206] [a]SDM: standard deviation of the mean from three trials. [b] $\Delta\Delta G^{\circ'}_{quin} = \Delta G^{\circ'}_{op,cells} - \Delta G^{\circ'}_{op,buff}$. [c]From propagation of error.

residue	$k_{obs,cells}$	SDM	$k_{obs,dilute}$	SDM
Y3	1.8E-04	±1E-05	2.97E-05	±8E-07
K4	2.1E-04	±3E-05	2.83E-05	±8E-07
A26	1.5E-04	±1E-05	1.2E-05	±1E-06
K28	1.62E-04	±7E-06	2.2E-05	±1E-06
V29	3.0E-05	±2E-06	6.2E-06	±7E-07
K31	1.3E-04	±2E-05	1.40E-05	±2E-07
A34	9.2E-05	±2E-06	1.89E-05	±4E-07
T44	8.98E-05	±6E-07	1.1E-05	±1E-06
D46	1.21E-04	±4E-06	2.04E-05	±7E-07
T51	1.7E-04	±4E-05	1.8E-05	±1E-06
F52	1.6E-04	±2E-05	1.50E-05	±6E-07
T53	1.9E-04	±3E-05	2.0E-05	±2E-06

residue	$\Delta G^{\circ'}_{op,cells}$	SDM[a]	$\Delta G^{\circ'}_{op,buff}$	SDM[a]	$\Delta\Delta G^{\circ'}_{quin}$ [b]	uncertainty[c]
Y3	4.13	±0.05	5.22	±0.02	-1.10	±0.05
K4	4.1	±0.1	5.0	±0.4	-0.8	±0.4
A26	4.59	±0.06	6.15	±0.07	-1.56	±0.09
K28	4.19	±0.03	5.41	±0.03	-1.22	±0.04
V29	4.50	±0.05	5.49	±0.07	-0.99	±0.08
K31	4.43	±0.09	5.81	±0.01	-1.37	±0.09
A34	4.69	±0.01	5.66	±0.01	-0.97	±0.02
T44	4.38	±0.00	5.67	±0.06	-1.29	±0.06
D46	4.44	±0.02	5.53	±0.02	-1.10	±0.03
T51	4.3	±0.1	5.69	±0.04	-1.4	±0.1
F52	4.24	±0.09	5.68	±0.03	-1.4	±0.1
T53	4.2	±0.1	5.57	±0.05	-1.4	±0.1

Table 3.4. Average $\Delta G_U^{\circ'}$ (kcal/mol) values for K10H GB1 in buffer (75 mM HEPES, 75 mM bis-tris propane, 75 mM citrate) quantified by amide-proton exchange (HDX) and differential scanning calorimetry (DSC) at pH 7.4, 6.0 and 5.0. pH is corrected for the isotope effect. For glass electrodes in D_2O, $pH = pH_{read} + 0.4$.[206] [a]SDM: standard deviation of the mean from three trials. $\Delta G_{U,DSC}^{\circ'}$ was calculated by using the modified Gibbs-Helmholtz equation using $\Delta H^{\circ'}_{cal}$ and T_M from calorimetry.[236]

pH	$\Delta G_{U,HDX}^{\circ'}$	SDM[a]	$\Delta G_{U,DSC}^{\circ'}$	SDM[a]
7.4	6.80	±0.08	7.1	±0.2
6.0	6.01	±0.07	6.2	±0.2
5.0	5.57	±0.09	5.2	±0.4

CHAPTER 4: QUINARY STRUCTURE REMODELS THE UNFOLDED PROTEIN ENSEMBLE[1]

4.1 Introduction

Protein folding has fascinated biological chemists for over a century.[237-241] In 1961, Anfinsen proposed that protein tertiary structure is encoded entirely by its amino acid sequence. His thermodynamic hypothesis states that "the three-dimensional structure of a native protein in its normal physiological milieu… is the one in which the Gibbs free energy of the whole system is lowest."[10] The Gibbs free energy of unfolding ($\Delta G_U^{\circ'}$) is the free energy of the unfolded ensemble (U) minus that of the folded state (F)

$$\Delta G_U^{\circ'} = G_U^{\circ'} - G_F^{\circ'}$$

The equilibrium thermodynamics of protein stability has been studied extensively in dilute solution. Although such experiments yield valuable information, these simple solutions do not capture the effects arising from quinary structure,[12-14,38] which comprises the transient interactions between macromolecules that organize the cellular interior. Repulsive interactions stabilize a test protein by minimizing the volume available, which favors the smaller folded state. Attractive interactions favor the unfolded state. The relative strengths of these interactions determine the net effect on

[1]The material in this chapter is being submitted for publication. If accepted, the original citation will be as follows: Cohen RD, Pielak GJ (2016) Quinary structure remodels the unfolded protein ensemble. RD Cohen and GJ Pielak wrote the paper.

protein stability.[148] Quinary structure's net effect on stability is determined by the intrinsic properties of the test protein: some proteins are stabilized in cells relative to buffer,[149-151] but others are destabilized.[136,139,146,151,242]

We chose the B1 domain of protein G (GB1, 6.2 kDa, pI 4.5)[166] as a model system for understanding quinary structure effects on protein stability. GB1 undergoes a two-state folding equilibrium[220] and has been extensively characterized in buffer[136,139,166,189,196,219-221,235] and in cells.[104-106,136,139,141] We have shown that quinary interactions do not change the structure of folded GB1[136] and have a negligible effect on its stability at pH 7.4.[141] Importantly, GB1 is foreign to *E. coli,* minimizing the risk of complicating specific interactions. We refer to the T2Q;K10H variant as the wild-type protein, because the T2Q mutation prevents N-terminal degradation[235] and the K10H mutation provides an intracellular pH probe.[140]

We recently used this system to demonstrate qualitatively[140] and quantitatively[141] that electrostatic interactions involving surface residues of the folded state contribute to protein quinary structure.[141] Now, we focus on quinary structure effects on the unfolded ensemble. Direct characterization of the ensemble is not practical because only about one in 10^5 GB1 molecules is unfolded under native conditions. However, we can gain information about the unfolding equilibrium by using amide proton exchange.[243]

NMR-detected amide proton exchange[137,138,179] is a powerful method to quantify equilibrium protein stability. For globular proteins, each backbone amide proton is in equilibrium between the closed, folded state and the open, exposed state. In D_2O, amide protons in the open state exchange for deuterons

$$\text{closed}-\text{NH} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{open}-\text{NH} \xrightarrow{k_{\text{int}}} \text{open}-\text{ND}$$

where k_{op} and k_{cl} are the rates of opening and closing, respectively, and the intrinsic exchange rate, k_{int}, is the rate of exchange in an unstructured peptide. The reaction is irreversible because the experiment is performed in D_2O.

Under the conditions used here, k_{int} is rate-determining for GB1[136,221] and the observed rate of exchange, k_{obs}, is proportional to the equilibrium constant of opening, K_{op}.

$$k_{obs} = \frac{k_{op}}{k_{cl}} k_{int} = K_{op} k_{int}$$

K_{op} is used to quantify local protein stability

$$\Delta G^{\circ'}_{op} = -RT \ln K_{op}$$

where R is the universal gas constant, and T is the absolute temperature, which is 310 K for all the work described here.

For any particular protein, individual residues that become exposed only upon global unfolding do not yield identical $\Delta G^{\circ'}_{op}$ values because the intrinsic rates of exchange are derived from model peptides, not the protein being studied. Importantly, we know from studies of numerous proteins that residues that exchange by global unfolding yield $\Delta G^{\circ'}_{op}$ values within 1 kcal/mol of $\Delta G^{\circ'}_{U}$ obtained from thermal denaturation.[225] Residues that yield $\Delta G^{\circ'}_{op}$ values under the conditions described here are known to exchange only upon complete GB1 unfolding.[136,189] For these reasons, we report the average $\Delta G^{\circ'}_{op}$ values as $\Delta G^{\circ'}_{U}$, along with an uncertainty derived from the standard deviation of the mean.

GB1 is one of only several proteins whose unfolded ensemble has been studied. Specifically, studies of GB1 peptides show that a non-native hydrophobic staple,[244] an α-helical motif formed between two hydrophobic residues at positions N' and N4, forms between V21 and A26.[245] Analysis of NOESY spectra show that the V21 side chain interacts with A26 and T25 in the staple.[245] Although the staple is absent in the folded structure, it stabilizes the unfolded state *via* a reverse hydrophobic effect because the hydrophobic valine is less solvent exposed in the unfolded state than in the native state.[244,246] To assess whether quinary interactions affect the unfolded ensemble, we quantified the stability of GB1 in buffer and in cells of four variants (V21A, V21I, V21T and V21T;T25V) expected to affect the hydrophobic staple.

4.2 Results

4.2.1 Disrupting the hydrophobic staple stabilizes GB1 in buffer

The $\Delta G^{\circ'}_{op}$ values for individual residues in buffer are shown in Figure 4.1. The values are well distributed across the secondary structure [β1 (Y3, K4), α1 (A26, K28, V29, K31, A34), β3 (T44, D46) and β4 (T51, F52, T53)], but crosspeak overlap and large k_{obs} values limit the number of quantifiable residues, especially in the β2 region.[189]

The unfolding free energy for wild type GB1 in 75 mM HEPES/75 mM bis-tris propane/75 mM citrate, pH 7.4 at 37 °C has been reported (6.80 ± 0.07 kcal/mol).[140] In buffer, the V21A and V21I mutations both stabilize GB1, yielding $\Delta G^{\circ'}_{U}$ values of 8.92 ± 0.08 and 8.86 ± 0.08 kcal/mol, respectively. V21T was too destabilized to quantify. Its stability is ≤6.4 kcal/mol, the smallest quantifiable value under these conditions. The double mutant, V21T;T25V, was more stable than the wild-type protein and yielded a $\Delta G^{\circ'}_{U}$ value of 8.3 ± 0.1 kcal/mol.

4.2.2 Disrupting the hydrophobic staple has much smaller effect in cells

In cells, the stability of each of the five variants is lower than their stability in buffer (Figure 4.2). This general destabilization is expected because attractive quinary interactions are present in cells.[104,105,117,118,136,139,140,146,147,151,165,247] Wild-type GB1 has a stability of 6.6 ± 0.1 kcal/mol in cells. The V21A and V21I variants have $\Delta G_U^{\circ'}$ values of 7.0 ± 0.2 and 7.33 ± 0.05 kcal/mol, respectively. The V21T and V21T;T25V variants were too destabilized for quantification, but their stabilities must be less than 6.1 kcal/mol, the smallest value we could measure in cells under these conditions.

4.3 Discussion

Changing surface residues that are not involved in secondary structure should not affect protein stability.[248] The surface residue V21 lies in an exposed loop, and its amide does not participate in intramolecular hydrogen bonds.[166,189] Contrary to expectation, but consistent with the effect of disrupting the hydrophobic staple in the unfolded state, hydrophobic residues at position 21 increase GB1 stability in buffer (Figure 4.3). The mutations V21A, which removes two methyl groups, and V21I, which adds one methyl group, stabilize GB1 by 2.11 ± 0.04 and 2.06 ± 0.04 kcal/mol, respectively.

On the other hand, threonine strengthens the staple and decreases GB1 stability in buffer. The V21T mutation, which replaces a methyl group with a hydroxyl group, destabilizes GB1 to such an extent that we cannot quantify its stability, but given our limit of detection, the decrease must be more than 0.8 kcal/mol.[136] We hypothesized that the V21T change stabilizes the unfolded ensemble because the side chain oxygen of T21 accepts a hydrogen bond from the amide proton of A26 or T25. We tested this

idea by introducing a second mutation that removes the additional hydrogen-bond acceptor. As expected, the V21T;T25V mutation disrupts the staple, consequently stabilizing the protein in buffer by 1.55 ± 0.08 kcal/mol.

We considered the possibility that the amino acid changes altered the folded state, but analysis of ^{15}N-^{1}H HSQC spectra shows that the mutations do not affect the protein (Figure 4.4). There is precedence for the idea that amino acid changes can affect unfolded states. For instance, mutations that have no effect on native contacts in staphylococcal nuclease change the stability of the unfolded state ensemble by impacting non-native interactions[249-251] and introducing a disulfide bond into cytochrome *c* stabilizes a compact unfolded state, destabilizing the protein.[252]

They key point is that the effects of changes to the staple are much less impactful in cells than in buffer (Figure 4.5). Whereas V21A and V21I stabilize GB1 in buffer by more than 2 kcal/mol, they stabilize GB1 in cells by only 0.2 ± 0.1 and 0.6 ± 0.1, respectively. Despite increasing the stability of GB1 in buffer, the V21T;T25V variant was too destabilized to quantify in cells. The structure of the folded state remains the same in buffer and in cells for all variants, as indicated by the similar patterns observed in HSQC spectra and in opening free energies along the primary structure. Therefore, the attenuation of stability changes from disrupting the hydrophobic staple in cells arises from changes in the unfolded ensemble. Specifically, quinary interactions decrease the stability of the hydrophobic staple relative to buffer, thus remodeling the unfolded ensemble. Consistent with this idea, Guzman *et al.* showed that the cellular environment has opposing effects on the stability of two proteins with similar size and surface properties: one is stabilized while the other is

destabilized.[151] The authors suggest that quinary interactions with the unfolded ensemble may be responsible for the differences in stabilization.

4.4 Summary

The cellular interior is a dynamic environment teeming with hundreds of thousands of different macromolecules. Quinary structure imposes the organization essential for governing the myriad of biochemical reactions that sustain life. While most biochemistry is studied in dilute, buffered solutions, we now know that quinary structure modulates protein stability through interactions with the folded state.[104,105,136,139-141,146] Our observations here indicate that quinary interactions influence the ensemble of unfolded states, further reinforcing the idea that data from experiments in dilute solution yield important fundamental information about proteins, but can fail to capture the impact of chemical interactions with the dynamic, heterogeneous environment of the cell. The source of these interactions lies in quinary structure, which can only be uncovered when proteins are studied in their physiological environment.[12-14] This knowledge will yield insight into the energetics that control protein conformation in cells, providing new possibilities for understanding the physiological roles of chaperones, the proteasome, and disease-causing protein mutations, particularly aggregation-based diseases, as they all directly involve the properties of unfolded protein ensembles in cells.[253-255]

4.5 Materials and Methods

The pET11a plasmids harboring the T2Q and T2Q;K10H GB1 gene have been described.[140,235] Mutations were installed using site-directed mutagenesis (Quikchange; Agilent) with the following primers and their reverse complements (mutation bolded):

V21A, 5'-CC ACC GAA GCT **GCT** GAC GCT GCT ACC GCG -3'; V21I, 5'-CC ACC GAA GCT **ATT** GAC GCT GCT ACC GCG -3'; V21T, 5'-CC ACC GAA GCT **ATC** GAC GCT GCT ACC GCG -3'; and V21T;T25V, 5'-GCT **ATC** GAC GCT GCT **GCT** GCG GAA AAA G-3'. Mutations were confirmed by DNA sequence analysis (Eton Bioscience, Research Triangle Park, NC).[256]

Proteins were expressed in *Escherichia coli* BL21(DE3) cells and purified as described.[136,166] For experiments in dilute solution, 2.2 mg of purified protein was resuspended in 500 µL 75 mM bis-tris propane/75 mM HEPES/75 mM citrate, 99.9 % D_2O, pH 7.4 and ^{15}N-^{1}H HSQC spectra were acquired serially. pH readings are uncorrected for the deuterium isotope effect ($pH = pH_{read} + 0.4$).[206]

Free energies of opening were quantified in cells using the quenched lysate method.[136] Briefly, cells were grown at 37 °C in M9 minimal media containing $^{15}NH_4Cl$ until they reached an optical density of 0.6 at 600 nm. Overexpression was induced with 1 mM (final concentration) of isopropyl-β-D-thiogalactopyranoside. After 2 h, cells were harvested by centrifugation and a timer was initiated upon resuspension in D_2O-containing buffer. Individual aliquots were removed at discrete time points and hydrogen exchange was quenched and cells were lysed. After centrifugation, the quenched lysate was transferred to an NMR tube and a spectrum acquired.

Exchange rates were converted to free energies using intrinsic exchange rates from SPHERE (37 °C, alanine oligopeptide basis, pH 7.4).[185] Intrinsic exchange rates remain unchanged in cells[184] and measurements obtained from this method have been confirmed.[136,139]

4.6 Figures

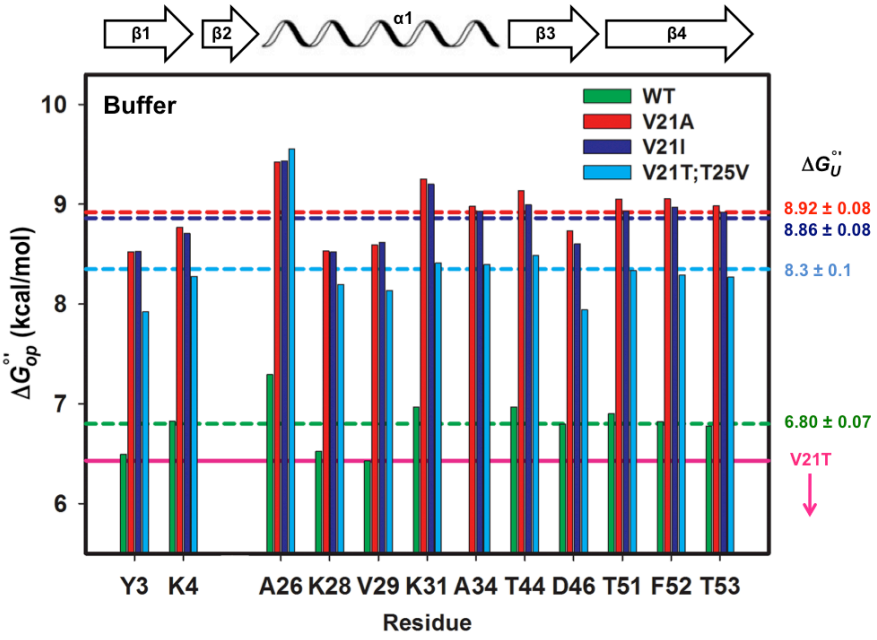

Figure 4.1. Stability of GB1 variants in buffer. $\Delta G^{\circ'}_{op}$ values for the wild-type protein (green), and the V21A (red), V21I (dark blue) and V21T;T25V (light blue) variants in buffer. V21T (purple) was not stable enough to quantify and the arrow indicates that the $\Delta G^{\circ'}_{op}$ value is less than the least stable quantifiable residue (V29 in the wild-type protein). The dashed lines indicate the average $\Delta G^{\circ'}_{op}$ value, which equals $\Delta G^{\circ'}_{U}$, for each variant. $\Delta G^{\circ'}_{U}$ values are shown on the right and the uncertainty represents the standard deviation of the mean.

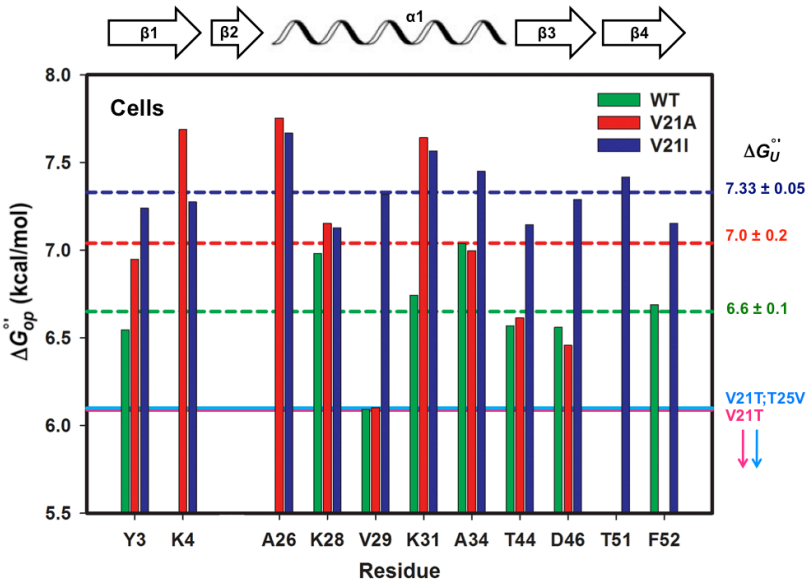

Figure 4.2. Stability of GB1 variants in cells. $\Delta G_{op}^{\circ'}$ values for wild-type (green), V21A (red) and V21I (blue) in cells. Dashed lines indicate the $\Delta G_U^{\circ'}$ values, equal to the value and their standard deviations of the mean for each variant shown on the right. V21T (purple) and V21T;T25V (cyan) were too destabilized to quantify. The arrows indicate that the $\Delta G_{op}^{\circ'}$ values are less than the least stable residue (V29) in wild-type.

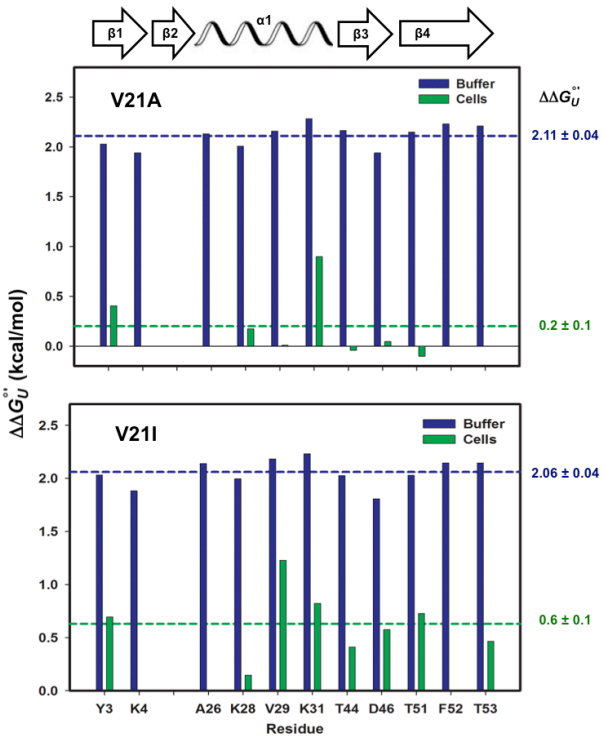

Figure 4.3. Mutations increasing GB1 stability in buffer have small effects in cells. Changes in opening free energies, $\Delta\Delta G^{\circ'}_{op}$ ($\Delta G^{\circ'}_{op,\mathrm{var}} - \Delta G^{\circ'}_{op,wt}$) for GB1 V21A (top panel) and GB1 V21I (bottom panel) in buffer (blue) and in cells (green). Dashed lines indicate the $\Delta\Delta G^{\circ'}_{U}$ value derived from the average $\Delta\Delta G^{\circ'}_{op}$ values and their standard deviation.

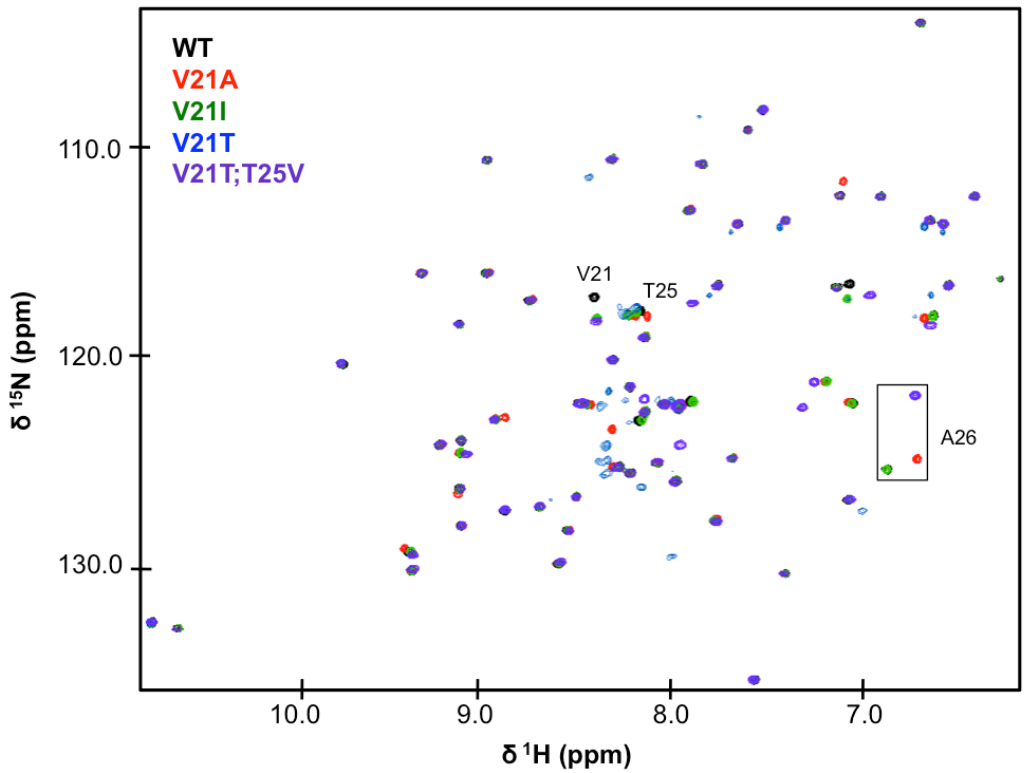

Figure 4.4. Overlaid ^{15}N-^{1}H HSQC spectra of WT (black), V21A (red), V21I (green), V21T (blue) and V21T;T25V (purple) GB1 in 20 mM $NaPO_i$, 150 mM NaCl, 25 °C. Residues participating in the hydrophobic staple (V21, T25, A26) are labeled for the wildtype protein.

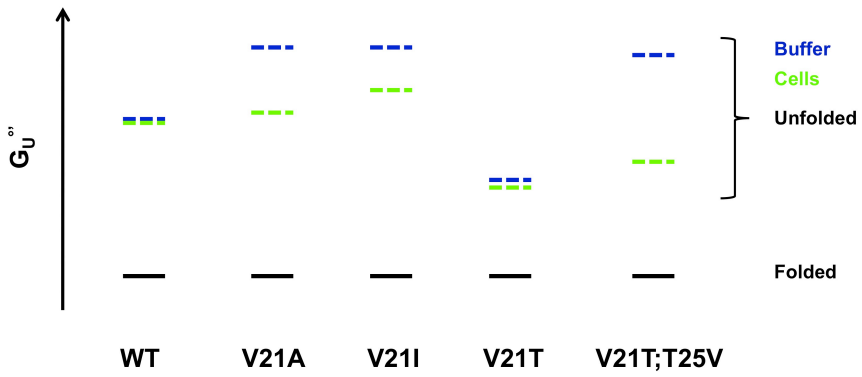

Figure 4.5. Quinary interactions modulate the unfolded state ensemble. Free energy diagram depicting the relative free energies for the folded (straight line) and unfolded states (dashed lines) in buffer (blue) and in cells (green) for the GB1 variants.

4.7 Tables

Table 4.1. Backbone amide proton exchange rates (k_{obs}, s^{-1}) and corresponding $\Delta G^{\circ'}_{op}$ (kcal/mol) values for V21A GB1 in cells and in buffer (75 mM HEPES, 75 mM bis-tris propane, 75 mM citrate, pH 7.4) at 37°C. pH is corrected for the isotope effect. For glass electrodes in D_2O, pH = pH_{read} + 0.4. $\Delta\Delta G^{\circ'}_{var} = \Delta G^{\circ'}_{var,op} - \Delta G^{\circ'}_{wt,op}$. From previous experiments, the uncertainties for ΔG (kcal/mol) are approximately ± 0.05 (buffer) and ±0.1 (cells).[136,139,141]

residue	$k_{obs,cells}$	$k_{obs,buff}$	$\Delta G^{\circ'}_{cells,op}$	$\Delta\Delta G^{\circ'}_{var,cells}$	$\Delta G^{\circ'}_{buff,op}$	$\Delta\Delta G^{\circ'}_{var,buff}$
Y3	4.43E-04	3.45E-05	6.95	0.40	8.52	2.03
K4	1.60E-04	2.78E-05	7.69	--	8.77	1.94
A26	2.23E-04	1.50E-05	7.75	--	9.42	2.13
K28	2.42E-04	2.59E-05	7.15	0.17	8.53	2.01
V29	5.38E-04	9.55E-06	6.10	0.01	8.59	2.16
K31	1.77E-04	1.31E-05	7.64	0.90	9.25	2.28
A34	5.40E-04	2.17E-05	7.00	-0.04	8.98	--
T44	5.91E-04	9.94E-06	6.61	0.04	9.13	2.16
D46	6.53E-04	1.63E-05	6.46	-0.10	8.74	1.94
T51	--	1.93E-05	--	--	9.05	2.15
F52	--	1.57E-05	--	--	9.05	2.23
T53	--	1.88E-05	--	--	8.98	2.21

Table 4.2. Backbone amide proton exchange rates (k_{obs}, s^{-1}) and corresponding $\Delta G^{\circ'}_{op}$ (kcal/mol) values for V21I GB1 in cells and in buffer (75 mM HEPES, 75 mM bis-tris propane, 75 mM citrate, pH 7.4) at 37°C. pH is corrected for the isotope effect. For glass electrodes in D_2O, pH = pH_{read} + 0.4. $\Delta\Delta G^{\circ'}_{var} = \Delta G^{\circ'}_{var,op} - \Delta G^{\circ'}_{wt,op}$. From previous experiments, the uncertainties for ΔG (kcal/mol) are approximately ± 0.05 (buffer) and ±0.1 (cells).[136,139,141]

residue	$k_{obs,cells}$	$k_{obs,buff}$	$\Delta G^{\circ'}_{cells,op}$	$\Delta\Delta G^{\circ'}_{var,cells}$	$\Delta G^{\circ'}_{buff,op}$	$\Delta\Delta G^{\circ'}_{var,buff}$
Y3	4.43E-04	3.45E-05	7.24	0.40	8.53	2.03
K4	1.60E-04	2.78E-05	7.27	--	8.71	1.94
A26	2.23E-04	1.50E-05	7.67	--	9.43	2.13
K28	2.42E-04	2.59E-05	7.13	0.17	8.52	2.01
V29	5.38E-04	9.55E-06	7.32	0.01	8.62	2.16
K31	1.77E-04	1.31E-05	7.57	0.90	9.20	2.28
A34	5.40E-04	2.17E-05	7.45	-0.04	8.93	--
T44	5.91E-04	9.94E-06	7.14	0.04	8.99	2.16
D46	6.53E-04	1.63E-05	7.29	-0.10	8.60	1.94
T51	--	1.93E-05	7.42	--	8.93	2.15
F52	--	1.57E-05	7.15	--	8.97	2.23
T53	--	1.88E-05	--	--	8.92	2.21

Table 4.3. Backbone amide proton exchange rates (k_{obs}, s^{-1}) and corresponding $\Delta G^{\circ'}_{op}$ (kcal/mol) values for V21T;T25V GB1 in buffer (75 mM HEPES, 75 mM bis-tris propane, 75 mM citrate, pH 7.4) at 37°C. pH is corrected for the isotope effect. For glass electrodes in D_2O, pH = pH_{read} + 0.4. $\Delta\Delta G^{\circ'}_{var} = \Delta G^{\circ'}_{var,op} - \Delta G^{\circ'}_{wt,op}$. From previous experiments, the uncertainties for ΔG (kcal/mol) are approximately ± 0.05 (buffer).[136,139,141]

residue	$k_{obs,buff}$	$\Delta G^{\circ'}_{buff,op}$	$\Delta\Delta G^{\circ'}_{var,buff}$
Y3	1.76E-04	7.92	1.43
K4	2.08E-04	8.28	1.45
A26	1.52E-04	9.55	2.26
K28	1.62E-04	8.20	1.67
V29	3.02E-05	8.14	1.71
K31	1.32E-04	8.41	1.44
A34	9.22E-05	8.40	--
T44	8.98E-05	8.49	1.52
D46	1.21E-04	7.94	1.15
T51	1.66E-04	8.34	1.43
F52	1.56E-04	8.29	1.47
T53	1.90E-04	8.27	1.50

REFERENCES

1 Linderstrøm-Lang, K. U. *Proteins and enzymes* in *Lane Medical Lectures* Vol. 6 (Stanford University Publications, 1952).

2 Fischer, E. Über die Hydrolyse der Proteinstoffe. *Chemiker Zeitung* **26**, 939-940 (1902).

3 Hofmeister, F. Ueber den bau des Eiweissmoleküla. *Naturwissenschaftliche Rundschau* **17**, 529-545 (1902).

4 Pauling, L., Corey, R. B. & Branson, H. R. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* **37**, 235-240 (1951).

5 Wrinch, D. The structure of the globular proteins. *Nature* **143**, 482-483 (1939).

6 Kauzmann, W. Structural factors in protein denaturation. *J. Cell Physiol. Suppl.* **47**, 113-131 (1956).

7 Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14**, 1-63 (1959).

8 Kauzmann, W. The three dimensional structures of proteins. *Biophys. J.* **4**, 43-54 (1964).

9 Bernal, J. D. General introduction structure arrangements of macromolecules. *Discuss. Faraday Soc.* **25**, 7-18 (1958).

10 Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223-230 (1973).

11 McBrien, D. C. H. & Moses, V. Compartmentation of the metabolism of lactose, galactose and glucose in *Escherichia coli*. *J. Gen. Microbiol.* **51**, 159-172 (1968).

12 Vaïnshteïn, B. K. Three-dimensional electron microscopy of biological macromolecules. *Physics-Uspekhi* **16**, 185-206 (1973).

13 Edelstein, S. J. Patterns in the quinary structures of proteins. *Biophys. J.* **32**, 347-360 (1980).

14 McConkey, E. H. Molecular evolution, intracellular organization, and the quinary structure of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 3236-3240 (1982).

15 Krebs, H. A. & Johnson, W. A. Acetopyruvic acid (αγ-diketovaleric acid) as an intermediate metabolite in animal tissues. *Biochem. J.* **31**, 772-779 (1937).

16 Krebs, H. A. & Johnson, W. A. Metabolism of ketonic acid in animal tissues. *Biochem. J.* **31**, 645-660 (1937).

17 An, S., Kumar, R., Sheets, E. D. & Benkovic, S. J. Reversible compartmentalization of *de novo* purine biosynthetic complexes in living cells. *Science* **320**, 103-106 (2008).

18 Gu, J. *et al.* The architecture of the mammalian respirasome. *Nature* **537**, 639-643 (2016).

19 Letts, J. A., Fiedorczuk, K. & Sazanov, L. A. The architecture of respiratory supercomplexes. *Nature* **537**, 644-648 (2016).

20 Green, D. E., Loomis, W. F. & Auerbach, V. H. Studies of the cyclophorase system: The complete oxidation of pyruvic acid to carbon dioxide and water. *J. Biol. Chem.* **172**, 389-404 (1948).

21 Green, D. E. The cyclophorase complex of enzymes. *Biol. Rev.* **26**, 410-455 (1951).

22 Floridi, A. Regulation processes in biological systems. II. Role of compartmentation and of organized multienzyme systems. *Acta Vitaminol Enzymol* **32**, 67-109 (1978).

23 veer Reddy, G. P. & Pardee, A. B. Multienzyme complex for metabolic channeling in mammalian DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* **77**, 3312-3316 (1980).

24 Ovádi, J. Physiological significance of metabolic channelling. *J. Theor. Biol.* **152**, 1-22 (1991).

25 Cascante, M., Sorribas, A. & Canela, E. I. Enzyme-enzyme interactions and metabolite channelling: Alternative mechanisms and their evolutionary significance. *Biochem. J.* **298**, 313-320 (1994).

26 Winkel, B. S. J. Metabolic channeling in plants. *Annu. Rev. Plant Biol.* **55**, 85-107 (2004).

27 Hyde, C., Ahmed, S., Padlan, E. A., Miles, E. W. & Davies, D. R. Three-dimensional structure of the tryptophan synthase $\alpha_2\beta_2$ multienzyme complex from *Salmonella typhimurium*. *J. Biol. Chem.* **263**, 17857-17871 (1988).

28 Mozzarelli, A., Peracchi, A., Rossi, G. L., Ahmed, S. A. & Miles, E. W. Microspectrophotometric studies on single crystals of the tryptophan synthase $\alpha_2\beta_2$ complex demonstrate formation of enzyme-substrate intermediates. *J. Biol. Chem.* **264**, 15774-15780 (1989).

29 Hyde, C. C. & Miles, E. W. The tryptophan synthase multienzyme complex: Exploring structure-function relationships with X-ray crystallography and mutagenesis. *Nature Biotechnol.* **8**, 27-32 (1990).

30 Srere, P. A., Mattiasson, B. & Mosbach, K. An immobilized three-enzyme system: A model for microenvironmental compartmentation in mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 2534-2538 (1973).

31 Halper, L. A. & Srere, P. A. Interaction between citrate synthase and mitochondrial malate dehydrogenase in the presence of polyethylene glycol. *Arch. Biochem. Biophys.* **184**, 529-534 (1977).

32 Beeckmans, S. & Kanarek, L. Demonstration of physical interactions between consecutive enzymes of the citric acid cycle and of the aspartate-malate shuttle: A study involving fumarase, malate dehydrogenase, citrate synthase and aspartate aminotransferase. *Eur. J. Biochem.* **117**, 527-535 (1981).

33 Greksak, M., Lopes-Cardozo, M. & Van den Bergh, S. G. Citrate synthesis in intact rat-liver mitochondria is irreversible. *Eur. J. Biochem.* **122**, 423-427 (1982).

34 Srere, P. A. The structure of the mitochondrial inner membrane-matrix. *Trends Biochem. Sci.* **7**, 375-378 (1982).

35 Barnes, S. J. & Weitzman, P. D. J. Organization of citric acid cycle enzymes into a multienzyme cluster. *FEBS Lett.* **201**, 267-270 (1986).

36 Haggie P.M. & Brindle, K. M. Mitochondrial citrate synthase is immobilized *in vivo*. *J. Biol. Chem.* **274**, 3941-3945 (1999).

37 Srere, P. A. The metabolon. *Trends Biochem. Sci.* **10**, 109-110 (1985).

38 Vélot, C., Mixon, M. B., Teige, M. & Srere, P. A. Model of a quinary structure between Krebs TCA cycle enzymes: A model for the metabolon. *Biochemistry* **36**, 14271-14276 (1997).

39 Morgunov, I. & Srere, P. A. Interaction between citrate synthase and malate dehydrogenase: Substrate channeling of oxaloacetate. *J. Biol. Chem.* **273**, 29540-29544 (1998).

40 Miles, E. W., Rhee, S. & Davies, D. R. The molecular basis of substrate channeling. *J. Biol. Chem.* **274**, 12193-12196 (1999).

41 Robinson, J. B., Inman, L., Sumegi, B. & Srere, P. A. Further characterization of the Krebs tricarboxylic acid cycle metabolon. *J. Biol. Chem.* **262**, 1786-1790 (1987).

42 Elcock, A. H. & McCammon, J. A. Evidence for electrostatic channeling in a fusion protein of malate dehydrogenase and citrate synthase. *Biochemistry* **35**, 12652-12658 (1996).

43 Elcock, A. H., Huber, G. A. & McCammon, J. A. Electrostatic channeling of substrates between enzyme active sites: Comparison of simulation and experiment. *Biochemistry* **36**, 16049-16058 (1997).

44 Shatalin, K., Lebreton, S., Rault-Leonardon, M., Vélot, C. & Srere, P. A. Electrostatic channeling of oxaloacetate in a fusion protein of porcine citrate synthase and porcine mitochondrial malate dehydrogenase. *Biochemistry* **38**, 881-889 (1999).

45 Wu, F. & Minteer, S. Krebs cycle metabolon: Structural evidence of substrate channeling revealed by cross-linking and mass spectrometry. *Angew. Chem., Int. Ed.* **54**, 1851-1854 (2015).

46 Bulutoglu, B., Garcia, K. E., Wu, F., Minteer, S. D. & Banta, S. Direct evidence for metabolon formation and substrate channeling in recombinant TCA cycle enzymes. *ACS Chem. Biol.* **11**, 2847-2853 (2016).

47 Buchanan, J. M. Biosynthesis of the purines. *J. Cell Physiol. Suppl.* **38**, 143-171 (1951).

48 Caperelli, C. A., Chettur, G., Lin, L. Y. & Benkovic, S. J. Purification of glycineamide ribonucleotide transformylase. *Biochem. Biophys. Res. Commun.* **82**, 403-410 (1978).

49 Caperelli, C. A., Benkovic, P. A., Chettur, G. & Benkovic, S. J. Purification of a complex catalyzing folate cofactor synthesis and transformulation in *de novo* purine biosynthesis. *J. Biol. Chem.* **255**, 1885-1890 (1980).

50 Schendel, F. J., Cheng, Y. S., Otvos, J. D., Wehrli, S. & Stubbe, J. Characterization and chemical properties of phosphoribosylamine, an unstable intermediate in the *de novo* purine biosynthetic pathway. *Biochemistry* **27**, 2614-2623 (1988).

51 Henikoff, S. *et al.* Multiple purine pathway enzymes activites are encoded at a single genetic locus in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 720-724 (1986).

52 Chan, C. Y. *et al.* Purinosome formation as a function of the cell cycle. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 1368-1373 (2015).

53 Verrier, F. *et al.* GPCRs regulate the assembly of a multienzyme complex for purine biosynthesis. *Nat. Chem. Biol.* **7**, 909-915 (2011).

54 An, S., Kyoung, M., Allen, J. J., Shokat, K. M. & Benkovic, S. J. Dynamic regulation of a metabolic multi-enzyme complex by protein kinase CK2. *J. Biol. Chem.* **285**, 11093-11099 (2010).

55 Sundaram, A. & An, S. Investigating the role of 3-phosphoinositide-dependent protein kinase 1 in the spatiotemporal regulation of the purinosome. *FASEB J* **29** (2015).

56 Huang, B., Jones, S. A., Brandenburg, B. & Zhuang, X. Whole-cell 3D STORM reveals interactions between cellular structures with nanometer-scale resolution. *Nat. Methods* **5**, 1047-1052 (2008).

57 Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810-813 (2008).

58 Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793-795 (2006).

59 An, S., Deng, Y., Tomsho, J. W., Minjoung, K. & Benkovic, S. J. Microtubule-assisted mechanism for functional metabolic macromolecular complex formation. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 12872-12876 (2010).

60 French, J. B. *et al.* Spatial colocalization and functional link of purinosomes and mitochondria. *Science* **351**, 733-737 (2016).

61 Barnea, G. *et al.* The genetic design of signaling cascades to record receptor activation. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 64-69 (2008).

62 Deng, Y. *et al.* Mapping protein-protein proximity in the purinosome. *J. Biol. Chem.* **287**, 36201-36207 (2012).

63 Kyoung, M., Russell, S. J., Kohnhorst, C. L., Esemoto, N. N. & An, S. Dynamic architecture of the purinosome involved in human *de novo* purine biosynthesis. *Biochemistry* **54**, 870-880 (2015).

64 Stavreva, D. A. & McNally, J. G. Fluorescence recovery after photobleaching (FRAP) methods for visualizing protein dynamics in living mammalian cell nuclei. *Methods Enzymol.* **375** (2004).

65 Axelrod, D., Koppel, D. E., Schlessinger, J., Elson, E. & Webb, W. W. Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophys. J.* **16**, 1055-1069 (1976).

66 Keilin, D. & Hartree, E. F. Activity of the cytochrome system in heart muscle preparations. *Biochem. J.* **41**, 500-502 (1947).

67 Chance, B., Estabrook, R. W. & Lee, C. Electron transport in the oxysome. *Science* **140**, 379-380 (1963).

68 Hatefi, Y. & Rieske, J. S. The preparation and properties of DPNH-cytochrome *c* reductase (complex I-III of the respiratory chain). *Methods Enzymol.* **10**, 225-231 (1967).

69 Tisdale, H. D. Preparation and properties of succinic—cytochrome *c* reductase (complex II–III) *Methods Enzymol.* **10**, 213-215 (1967).

70 Blair, P. V. Preparation and properties of repeating units of mitochondrial electron transfer. *Methods Enzymol.* **10**, 208-212 (1967).

71 Fowler, L. & Hatefi, Y. Reconstitution of the electron transport system III. Reconstitution of DPNH oxidase, succinic oxidase, and DPNH, succinic oxidase. *Biochem. Biophys. Res. Commun.* **5**, 203-208 (1961).

72 Hatefi, Y., Haavik, A. G., Fowler, L. R. & Griffiths, D. E. Studies on the electron transfer system: Xlii. Reconstitution of the electron transfer systems. *J. Biol. Chem.* **237**, 2661-2669 (1962).

73 Ragan, C. I. & Heron, C. The interaction between mitochondrial NADH-ubiquinone oxidoreductase and ubiquinol-cytochrome *c* oxidoreductase. Evidence for stoichiometric association. *Biochem. J.* **174**, 783-790 (1978).

74 Hatefi, Y., Haavik, A. & Griffiths, D. Reconstitution of the electron transport system: I. Preparation and properties of the interacting enzymes complexes. *Biochem. Biophys. Res. Commun.* **4**, 441-446 (1961).

75 Hatefi, Y., Haavik, A. G. & Griffiths, D. E. Studies on the electron transfer system: Xl. Preparation and properties of mitochondrial DPNH-coenzyme Q reductase. *J. Biol. Chem.* **237**, 1676-1680 (1962).

76 Hatefi, Y. & Rieske, J. S. Preparation and properties of DPNH-coenzyme Q reductase (complex I of the respiratory chain). *Methods Enzymol.* **10**, 235-239 (1967).

77 Hackenbrock, C. R., Chazotte, B. & Gupte, S. S. The random collision model and a critical assessment of diffusion and collision in mitochondrial electron transport. *J. Bioenerg. Biomembr.* **18**, 331-368 (1986).

78 Gupte, S. *et al.* Relationship between lateral diffusion, collision frequency, and electron transfer of mitochondrial inner membrane oxidation-reduction components. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 2606-2610 (1984).

79 Gupte, S. & Hackenbrock, C. R. Multidimensional diffusion modes and collision frequencies of cytochrome *c* and its redox partners. *J. Biol. Chem.* **263**, 5241-5247 (1988).

80 Hochman, J., Ferguson-Miller, S. & Schindler, M. Mobility in the mitochondrial electron transport chain. *Biochemistry* **24**, 2509-2516 (1985).

81 Schägger, H. & Pfeiffer, K. Supercomplexes in the respiratory chains of yeast and mammalian mitochondria. *EMBO J* **19**, 1777-1783 (2000).

82 Schägger, H. & Pfeiffer, K. The ratio of oxidative phosphorylation complexes I–V in bovine heart mitochondria and the composition of respiratory chain supercomplexes. *J. Biol. Chem.* **276**, 37861-37867 (2001).

83 Acín-Pérez, R., Fernández-Silva, P., Peleato, M. L., Pérez-Martos, A. & Enriquez, J. A. Respiratory active mitochondrial supercomplexes. *Mol. Cell* **32**, 529-539 (2008).

84 Porras, C. A. & Bai, Y. Respiratory supercomplexes: Plasticity and implications. *Front. Biosci.* **20**, 621-634 (2015).

85 Acín-Pérez, R. & Enriquez, J. A. The function of the respiratory supercomplexes: The plasticity model. *Biochim. Biophys. Acta* **1837**, 444-450 (2014).

86 Wittig, I., Carrozzo, R., Santorelli, F. M. & Schägger, H. Supercomplexes and subcomplexes of mitochondrial oxidative phosphorylation. *Biochim. Biophys. Acta* **1757**, 1066-1072 (2006).

87 Bultema, J. B., Braun, H.-P., Boekema, E. J. & Kouřil, R. Megacomplex organization of the oxidative phosphorylation system by structural analysis of

respiratory supercomplexes from potato. *Biochim. Biophys. Acta* **1787**, 60-67 (2009).

88 Althoff, T., Mills, D. J., Popot, J.-L. & Kühlbrandt, W. Arrangement of electron transport chain components in bovine mitochondrial supercomplex $I_1III_2IV_1$. *EMBO J* **30**, 4652-4664 (2011).

89 Schäfer, E., Dencher, N. A., Vonck, J. & Parcej, D. N. Three-dimensional structure of the respiratory chain supercomplex $I_1III_2IV_1$ from bovine heart mitochondria. *Biochemistry* **46**, 12579-12585 (2007).

90 Dudkina, N. V., Kudryashev, M., Stahlberg, H. & Boekema, E. J. Interaction of complexes I, III, and IV within the bovine respirasome by single particle cryoelectron tomography. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 15196-15200 (2011).

91 Schafer, E. *et al.* Architecture of active mammalian respiratory chain supercomplexes. *J. Biol. Chem.* **281**, 15370-15375 (2006).

92 Mileykovskaya, E. & Dowhan, W. Cardiolipin-dependent formation of mitochondrial respiratory supercomplexes. *Chem. Phys. Lipids* **179**, 42-48 (2014).

93 Zimmerman, S. B. & Trach, S. O. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J. Mol. Biol.* **222**, 599-620 (1991).

94 Conlon, I. & Raff, M. Differences in the way a mammalian cell and yeast cells coordinate cell growth and cell-cycle progression. *J Biol* **2**, 7 (2003).

95 Zeskind, B. J. *et al.* Nucleic acid and protein mass mapping by live-cell deep-ultraviolet microscopy. *Nat Methods* **4**, 567-569 (2007).

96 Cheung, M. C. *et al.* Intracellular protein and nucleic acid measured in eight cell types using deep-ultraviolet mass mapping. *Cytometry A* **83**, 540-551 (2013).

97 Clegg, J. S. Properties and metabolism of the aqueous cytoplasm and its boundaries. *Am. J. Physiol* **246**, 133-151 (1984).

98 Serber, Z. *et al.* High-resolution macromolecular NMR spectroscopy in living cells. *J. Am. Chem. Soc.* **123**, 2446-2447 (2001).

99 Li, C. *et al.* Differential dynamical effects of macromolecular crowding on an intrinsically disordered protein and a globular protein: Implications for in-cell NMR spectroscopy. *J. Am. Chem. Soc.* **130**, 6310-6311 (2008).

100 Serber, Z., Ledwidge, R., Miller, S. M. & Dötsch, V. Evaluation of parameters critical to observing proteins inside living *Escherichia coli* by in-cell NMR spectroscopy. *J. Am. Chem. Soc.* **123**, 8895-8901 (2001).

101 Barnes, C. O., Monteith, W. B. & Pielak, G. J. Internal and global protein motion assessed with a fusion construct and in-cell NMR spectroscopy. *ChemBioChem* **12**, 390-391 (2011).

102 Bertini, I., Cavallaro, G. & Rosato, A. Cytochrome *c*: Occurrence and functions. *Chem. Rev.* **106**, 90-115 (2006).

103 Pielak, G., Auld, D., Betz, S., Hilgen-Willis, S. & Garcia, L. *Nuclear magnetic resonance studies of class I cytochrome c* in *Cytochrome c: A multidisciplinary approach* (eds RA Scott & AG Mauk) 203-284 (Sausalito: University Science Books, 1996).

104 Crowley, P. B., Chow, E. & Papkovskaia, T. Protein interactions in the *Escherichia coli* cytosol: An impediment to in-cell NMR spectroscopy. *ChemBioChem* **12**, 1043-1048 (2011).

105 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* **50**, 9225-9236 (2011).

106 Reardon, P. N. & Spicer, L. D. Multidimensional NMR spectroscopy for protein characterization and assignment inside cells. *J. Am. Chem. Soc.* **127**, 10848-10849 (2005).

107 Sakakibara, D. *et al.* Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* **458**, 102-105 (2009).

108 Richards, F. M. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151-176 (1977).

109 Dedmon, M. M., Patel, C. N., Young, G. B. & Pielak, G. J. FlgM gains structure in living cells. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12681-12684 (2002).

110 Alderson, T. R. & Bax, A. Parkinson's disease: Disorder in the court. *Nature* **530**, 38-39 (2016).

111 Binolfi, A., Theillet, F. X. & Selenko, P. Bacterial in-cell NMR of human alpha-synuclein: A disordered monomer by nature? *Biochem. Soc. Trans* **40**, 950-954 (2012).

112 Theillet, F. *et al.* Structural disorder of monomeric α-synuclein persists in mammalian cells. *Nature* **530**, 45-50 (2016).

113 Serber, Z. *et al.* Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. *Nat. Protocols* **1**, 2701-2709 (2006).

114 Selenko, P., Serber, Z., Gadea, B., Ruderman, J. & Wagner, G. Quantitative NMR analysis of the protein G B1 domain in *Xenopus laevis* egg extracts and intact oocytes. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11904-11909 (2006).

115 Barbieri, L., Luchinat, E. & Banci, L. Characterization of proteins by in-cell NMR spectroscopy in cultured mammalian cells. *Nat. Protocols* **11**, 1101-1111 (2016).

116 Inomata, K. *et al.* High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* **458**, 106-109 (2009).

117 Majumder, S. *et al.* Probing protein quinary interactions by in-cell nuclear magnetic resonance spectroscopy. *Biochemistry* **54**, 2727-2738 (2015).

118 Barbieri, L., Luchinat, E. & Banci, L. Protein interaction patterns in different cellular environments are revealed by in-cell NMR. *Sci. Rep.* **5**, 14456 (2015).

119 Gerig, J. Fluorine NMR of proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* **26**, 293-370 (1994).

120 Danielson, M. A. & Falke, J. J. Use of ^{19}F NMR to probe protein structure and conformational changes. *Annu. Rev. Biophys. Biomol. Struct.* **25**, 163-195 (1996).

121 Williams, S. P., Fulton, A. M. & Brindle, K. M. Estimation of the intracellular free ADP concentration by ^{19}F NMR studies of fluorine-labeled yeast phosphoglycerate kinase *in vivo*. *Biochemistry* **32**, 4895-4902 (1993).

122 Williams, S. P., Haggie, P. M. & Brindle, K. M. ^{19}F NMR measurements of the rotational mobility of protein *in vivo*. *Biophys. J.* **72**, 490-498 (1997).

123 Jackson, J. C., Hammill, J. T. & Mehl, R. A. Site-specific incorporation of a ^{19}F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *J. Am. Chem. Soc.* **129**, 1160-1166 (2007).

124 Li, C. *et al.* Protein ^{19}F NMR in *Escherichia coli*. *J Am Chem Soc* **132**, 321-327 (2010).

125 Harper, D. B. & O'Hagan, D. The fluorinated natural products. *Natural Product Reports* **11**, 123-133 (1994).

126 O'Hagan, D. & B. Harper, D. Fluorine-containing natural products. *J. Fluorine Chem.* **100**, 127-133 (1999).

127 Crowley, P. B., Kyne, C. & Monteith, W. B. Simple and inexpensive incorporation of ^{19}F-tryptophan for protein NMR spectroscopy. *Chem. Commun.* **48**, 10681-10683 (2012).

128 Entress, R. M. H. *et al.* ^{19}F NMR in the measurement of binding affinities of chloroeremomycin to model bacterial cell-wall surfaces that mimic VanA and VanB resistance. *Chem. Biol.* **5**, 329-337 (1998).

129 Richards, K. L., Rowe, M. L., Hudson, P. B., Williamson, R. A. & Howard, M. J. Combined ligand-observe ^{19}F and protein-observe ^{15}N,^{1}H-HSQC NMR suggests phenylalanine as the key δ-somatostatin residue recognized by human protein disulfide isomerase. *Sci. Rep.* **6**, 19518 (2016).

130 Gee, C. T. *et al.* Protein-observed ^{19}F-NMR for fragment screening, affinity quantification and druggability assessment. *Nat. Protocols* **11**, 1414-1427 (2016).

131 Jordan, J. B. *et al.* Fragment-linking approach using ^{19}F NMR spectroscopy to obtain highly potent and selective inhibitors of β-secretase. *J. Med. Chem.* **59**, 3732-3749 (2016).

132 Norton, R. S., Leung, E. W., Chandrashekaran, I. R. & MacRaild, C. A. Applications of ^{19}F NMR in fragment-based drug discovery. *Molecules* **21**, 860 (2016).

133 Gee, C. T., Koleski, E. J. & Pomerantz, W. C. K. Fragment screening and druggability assessment for the CBP/p300 KIX domain through protein-observed ^{19}F NMR spectroscopy. *Angew. Chem., Int. Ed.* **54**, 3735-3739 (2015).

134 Ghaemmaghami, S., Fitzgerald, M. C. & Oas, T. G. A quantitiative, high-throughput screen for protein stability. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8296-8301 (2000).

135 Ghaemmaghami, S. & Oas, T. G. Quantitative protein stability measurement *in vivo*. *Nat. Struct. Biol.* **8**, 879-882 (2001).

136 Monteith, W. B. & Pielak, G. J. Residue level quantification of protein stability in living cells. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11336-11340 (2014).

137 Englander, S. W. & Kallenbach, N. R. Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.* **16**, 521-655 (1983).

138 Miklos, A. C., Li, C. & Pielak, G. J. Using NMR-detected backbone amide ^{1}H exchange to assess macromolecular crowding effects on globular-protein stability. *Methods Enzymol.* **466**, 1-16 (2009).

139 Monteith, W. B., Cohen, R. D., Smith, A. E., Guzman-Cisneros, E. & Pielak, G. J. Quinary structure modules protein stability in cells. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 1739-1742 (2015).

140 Cohen, R. D., Guseman, A. J. & Pielak, G. J. Intracellular pH modulates quinary structure. *Protein Sci.* **24**, 1748-1755 (2015).

141 Cohen, R. D. & Pielak, G. J. Electrostatic contributions to protein quinary structure *J. Am. Chem. Soc.* **138**, 13139-13142 (2016).

142 Matthews, B. W. Studies on protein stability with T4 lysozyme. *Adv. Protein Chem.* **46**, 249-278 (1995).

143 Smith, C. K. & Regan, L. Guidelines for protein design: The energetics of β sheet side chain interactions. *Science* **270**, 980 (1995).

144 Lahr, S. J. *et al.* Patterned library analysis: A method for the quantitative assessment of hypotheses concerning the determinants of protein structure. *Proc. Natl. Acad. Sci. USA* **96**, 14860-14865 (1999).

145 Schreiber, G. & Fersht, A. R. Energetics of protein-protein interactions: Analysis of the barnase-barstar interface by single mutations and double mutant cycles. *J. Mol. Biol.* **248**, 478-486 (1995).

146 Danielsson, J. *et al.* Thermodynamics of protein destabilization in live cells. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 12402-12407 (2015).

147 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.* **113**, 1725-1730 (2016).

148 Sarkar, M., Li, C. & Pielak, G. J. Soft interactions and crowding. *Biophys. Rev.* **5**, 187-194 (2013).

149 Dhar, A. *et al.* Protein stability and folding kinetics in the nucleus and endoplasmic reticulum of eucaryotic cells. *Biophys. J.* **101**, 421-430 (2011).

150 Guo, M., Xu, Y. & Gruebele, M. Temperature dependence of protein folding kinetics in living cells. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17863-17867 (2012).

151 Guzman, I., Gelman, H., Tai, J. & Gruebele, M. The extracellular protein VlsE is destabilized inside cells. *J. Mol. Biol.* **426**, 11-20 (2014).

152 Chen, K. *et al.* Simple NMR methods for evaluating higher order structures of monoclonal antibody therapeutics with quinary structure. *J. Pharm. Biomed. Anal.* **128**, 398-407 (2016).

153 Shire, S. J. Formulation and manufacturability of biologics. *Curr. Opin. Biotechnol.* **20**, 708-714 (2009).

154 Pace, C. N., Grimsley, G. R. & Scholtz, J. M. Protein ionizable groups: Pk values and their contribution to protein stability and solubility. *J. Biol. Chem.* **284**, 13285-13289 (2009).

155 London, R. E., Gregg, C. T. & Matwiyoff, N. A. Nuclear magnetic resonance of rotational mobility of mouse hemoglobin labeled with [2-^{13}c] histidine. *Science* **188**, 266-268 (1975).

156 Smith, A. E., Zhang, Z., Pielak, G. J. & Li, C. NMR studies of protein folding and binding in cells and cell-like environments. *Curr. Opin. Struct. Biol.* **30**, 7-16 (2015).

157 Wang, Y., Sarkar, M., Smith, A. E., Krois, A. S. & Pielak, G. J. Macromolecular crowding and protein stability. *J Am Chem Soc* **134**, 16614-16618 (2012).

158 Benton, L. A., Smith, A. E., Young, G. B. & Pielak, G. J. Unexpected effects of macromolecular crowding on protein stability. *Biochemistry* **51**, 9773-9775 (2012).

159 Senske, M. *et al.* Protein stabilization by macromolecular crowding through enthalpy rather than entropy. *J. Am. Chem. Soc.* **136**, 9036-9041 (2014).

160 Sapir, L. & Harries, D. Is the depletion force entropic? Molecular crowding beyond steric interactions. *Curr. Opin. Colloid Interface Sci.* **20**, 3-10 (2015).

161 Srere, P. A. Complexes of squential metabolic enzymes. *Annu. Rev. Biochem.* **56**, 89-124 (1987).

162 Wirth, A. J. & Gruebele, M. Quinary protein structure and the consequences of crowding in living cells: Leaving the test-tube behind. *Bioessays* **35**, 984-993 (2013).

163 Gershenson, A. Deciphering protein stability in cells. *J. Mol. Biol.* **426**, 4-6 (2014).

164 Majumder, S. *et al.* Probing protein quinary interactions by in-cell nuclear magnetic resonance spectroscopy. *Biochemistry* **54**, 2727-2738 (2015).

165 Kyne, C., Ruhle, B., Gautier, V. W. & Crowley, P. B. Specific ion effects on macromolecular interactions in *Escherichia coli* extracts. *Protein Sci.* **24**, 310-318 (2015).

166 Gronenborn, A. M. *et al.* A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* **253**, 657-660 (1991).

167 Gallagher, T., Alexander, P., Bryan, P. & Gilliland, G. L. Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry* **33**, 4721-4729 (1994).

168 Minton, A. P. Excluded volume as a determinant of macromolecular structure and reactivity. *Biopolymers* **20**, 2093-2120 (1981).

169 Linderstrøm-Lang, K. On the ionisation of proteins. *C.R. Trav. Lab Carlsberg* **15**, 1-29 (1924).

170 Matthew, J. B. & Richards, F. M. The pH dependence of hydrogen exchange in proteins. *J. Biol. Chem.* **258**, 3039-3044 (1983).

171 Yang, A. & Honig, B. On the pH dependence of protein stability. *J. Mol. Biol.* **231**, 459-474 (1993).

172 Myer YP & Harbury HA. Optical rotatory dispersion of cytochrome *c*. *Proc. Natl. Acad. Sci. U. S. A.* **54**, 1391-1398 (1965).

173 Myer YP. Conformation of cytochromes. III. Effect of urea, temperature, extrinsic ligands, and pH variation on the conformation of horse heart ferricytochrome *c*. *Biochemistry* **7**, 765-775 (1968).

174 Jeng, M.-F. & Englander, S. W. Stable submolecular folding units in a non-compact form of cytochrome *c*. *J. Mol. Biol.* **221**, 1045-1061 (1991).

175 Fink, A. L., Calciano, L. J., Goto, Y., Kurotsu, T. & Palleros, D. R. Classification of acid denaturation of proteins: Intermediates and unfolded states. *Biochemistry* **33**, 12504-12511 (1994).

176 Yang A & Honig B. Structural origins of pH and ionic strength effects on protein stability: Acid denaturation of sperm whale apomyoglobin. *J. Mol. Biol.* **237**, 602-614 (1994).

177 Spitzer, J. & Poolman, B. The role of biomacromolecular crowding, ionic strength, and physicochemical gradients in the complexities of life's emergence. *Microbiol Mol Biol Rev* **73**, 371-388 (2009).

178 Zhou, P. & Wagner, G. Overcoming the solubility limit with solubility-enhancement tags: Successful applications in biomolecular NMR studies. *J Biomol NMR* **46**, 23-31 (2010).

179 Linderstrøm-Lang, K. *Deuterium exchange and protein structure* in *Symposium on protein structure* (ed A. Neuberger) 22-34 (1958).

180 Hvidt AA & Nielsen SO. Hydrogen exchange in proteins. *Adv. Protein Chem.* **21**, 287-386 (1966).

181 Molday, R. S., Englander, S. W. & Kallen, R. G. Primary structure effects on peptide group hydrogen exchange. *Biochemistry* **11**, 150-158 (1972).

182 Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75-86 (1993).

183 Smith, A. E., Zhou, L. Z. & Pielak, G. J. Hydrogen exchange of disordered proteins in *Escherichia coli*. *Protein Sci.* **24**, 706-713 (2015).

184 Smith AE, Sarkar M, Young GB & Pielak GJ. Amide proton exchange of a dynamic loop in cell extracts. *Protein Sci.* **22**, 1313-1319 (2013).

185 Zhang, Y. Z. *Protein and peptide structure and interactions studied by hydrogen exchange and NMR*., University of Pennsylvania, Philadelphia, (1995).

186 Alexander, P., Orban, J. & Bryan, P. Kinetic analysis of folding and unfolding the 56 amino acid IgG-binding domain of streptococcal protein G. *Biochemistry* **31**, 7243-7248 (1992).

187 Englander, S. W., Sosnick, T. R., Englander, J. J. & Mayne, L. Mechanisms and uses of hydrogen exchange. *Curr. Opin. Struct. Biol.* **6**, 18-23 (1996).

188 McAllister, R. G. & Konermann, L. Challenges in the interpretation of protein h/d exchange data: A molecular dynamics simulation perspective. *Biochemistry* **54**, 2683-2692 (2015).

189 Orban, J., Alexander, P., Bryan, P. & Khare, D. Assessment of stability differences in the protein G B1 and B2 domains from hydrogen-deuterium exchange: Comparison with calorimetric data. *Biochemistry* **34**, 15291-15300 (1995).

190 Shimba, N. *et al.* Quantitative identification of the protonation state of histidines *in vitro* and *in vivo*. *Biochemistry* **42**, 9227-9234 (2003).

191 Davison, T. S. *et al.* Structure and functionality of a designed p53 dimer. *J. Mol. Biol.* **307**, 605-617 (2001).

192 Henderson, L. Concerning the relationship between the strength of acids and their capacity to preserve neutrality. *Am. J. Physiol* **21**, 173-179 (1908).

193 Hasselbalch, K. *Die berechnung der wasserstoffzahl des blutes auf der freien und gebundenen kohlensaure desselben, und die sauerstoffbindung des blutes als funktion der wasserstoffzahl*. *Biochemistry Z* **78**, 112-144 (1916).

194 Kelly, A., Ou, H., Withers, R. & Dötsch, V. Low-conductivity buffers for high-sensitivity NMR measurements. *J. Am. Chem. Soc.* **124**, 12013-12019 (2002).

195 Bryant, J. E., Lecomte, J. T. J., Lee, A. L., Young, G. B. & Pielak, G. J. Protein dynamics in living cells (retracted article. See vol 46, pg 8206, 2007). *Biochemistry* **44**, 9275-9279 (2005).

196 Khare, D. *et al.* pK_a measurements from nuclear magnetic resonance for the B1 and B2 immunoglobulin G-binding domains of protein G: Comparison with

calculated values for nuclear magnetic resonance and X-ray structures. *Biochemistry* **36**, 3580-3589 (1997).

197 Waudby, C. A. *et al.* In-cell NMR characterization of the secondary structure populations of a disordered conformation of ‚âà√≠¬¨¬±-synuclein within *E. coli* cells. *PloS ONE* **8**, e72286 (2013).

198 Orban, J., Alexander, P. & Bryan, P. Sequence-specfic ^{1}H NMR assignments and secondary structure of the streptococcal protein G B1-domain. *Biochemistry* **31**, 3604-3611 (1992).

199 Cummings, J. H. & Macfarlane, G. T. The control and consequences of bacterial fermentation in the human colon. *J. Appl. Bacteriol.* **70**, 443-459 (1991).

200 Li, C. *et al.* Differential dynamical effects of macromolecular crowding on an intrinsically disordered protein and a globular protein: Implications for in-cell NMR spectroscopy. *J Am Chem Soc* **130**, 6310-6311 (2008).

201 Schlesinger, A. P., Wang, Y., Tadeo, X., Millet, O. & Pielak, G. J. Macromolecular crowding fails to fold a globular protein in cells. *J. Am. Chem. Soc.* **133**, 8082-8085 (2011).

202 Rule, G. S. & Hitchens, T. K. *Fundamentals of protein NMR spectroscopy*. (Springer, 2007).

203 Pielak, G. J. *et al.* Protein nuclear magnetic resonance under physiological conditions. *Biochemistry* **2009**, 226-234 (2009).

204 Minhaz Ud-Dean, S. M. Nanoreactors for pH controlled sequential activity switching in multistep enzymatic processes. *IET Nanobiotechnol* **3**, 65-70 (2009).

205 Li, L., Sham, Y. Y., Bikadi, Z. & Elmquist, W. F. pH-dependent transport of pemetrexed by breast cancer resistance protein. *Drug Metab Dispos* **39**, 1478-1485 (2011).

206 Glasoe, P. K. & Long, F. A. Use of glass electrodes to measure acidities in deuterium oxide *J. Phys. Chem.* **64**, 188-190 (1960).

207 Delaglio F *et al.* Nmrpipe: A multidimensional spectral processing system based on unix pipes. *J. Biomol. NMR* **6**, 277-293 (1995).

208 Johnson BA & Blevins RA. NMR view: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603-614 (1994).

209 Va∆í,â†nshte∆í,â†n, B. K. Three-dimensional electron microscopy of biological macromolecules. *Physics-Uspekhi* **16**, 185-206 (1973).

210 Edelstein, S. J. Patterns in the quinary structures of proteins. Plasticity and inequivalence of individual molecules in helical arrays of sickle cell hemoglobin and tubulin. *Biophys J* **32**, 347-360 (1980).

211 McConkey, E. H. Molecular evolution, intracellular organization, and the quinary structure of proteins. *Proc. Natl. Acad. Sci. USA* **79**, 3236-3240 (1982).

212 D'Souza, S. F. & Srere, P. A. Binding of citrate synthase to mitochondrial inner membranes. *J. Biol. Chem.* **258**, 4706-4709 (1983).

213 French, J. B. *et al.* Spatial colocalization and functional link of purinosomes with mitochondria. *Science* **351**, 733-737 (2016).

214 Ogston, A. G. & Phelps, C. F. The partition of solutes between buffer solutions and solutions containing hyaluronic acid. *Biochem. J.* **78**, 827-833 (1961).

215 Laurent, T. C. The interaction between polysaccharides and other molecules: The solubility of proteins in the presence of dextran. *Biochem. J.* **89**, 253-257 (1963).

216 Zhou, H.-X., Rivas, G. à. n. & Minton, A. P. Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **37**, 353-373 (2008).

217 Senske, M. *et al.* Protein stabilization by macromolecular crowding through enthalpy rather than entropy. *J. Am. Chem. Soc.* **136**, 9036-9041 (2014).

218 Harada, R., Tochio, N., Kigawa, T., Sugita, Y. & Feig, M. Reduced native state stability in crowded cellular environment due to protein,äö√ñ√¨protein interactions. *J. Am. Chem. Soc.* **135**, 3696-3701 (2013).

219 Gallagher, T., Alexander, P., Bryan, P. & Gilliland, G. L. Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry* **33**, 4721-4729 (1994).

220 Alexander, P., Fahnestock, S., Lee, T., Orban, J. & Bryan, P. Thermodynamic analysis of the folding of the streptococcal protein G IgG-binding domains B1 and B2: Why small proteins tend to have high denaturation temperatures. *Biochemistry* **31**, 3597-3603 (1992).

221 Alexander, P., Orban, J. & Bryan, P. Kinetic analysis of folding and unfolding the 56 amino acid IgG-binding domain of streptococcal protein G. *Biochemistry* **31**, 7243-7248 (1992).

222 Linderstrom-Lang, K. Deuterium exchange between peptides and water. *Chemistry Society (London) Special Publication* **2**, 1-20 (1955).

223 Englander, S. W. & Kallenbach, N. R. Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.* **16**, 521-655 (1983).

224 Ghaemmaghami, S., Fitzgerald, M. C. & Oas, T. G. A quantitative, high-throughput screen for protein stability. *Proc. Natl. Acad. Sci. USA* **97**, 8296-8301 (2000).

225 Huyghues-Despointes, B. M. P., Scholtz, J. M. & Pace, C. N. Protein conformational stabilities can be determined from hydrogen exchange rates. *Nat. Struct. Biol.* **6**, 910-912 (1999).

226 Hickey, E. W. & Hirshfield, I. N. Low-pH-induced effects on patterns of protein synthesis and on internal pH in *Escherichia coli* and *Salmonella typhimurium*. *Appl. Environ. Microbiol.* **56**, 1038-1045 (1990).

227 Richards, F. M. Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151-176 (1977).

228 Fleming, P. J. & Rose, G. D. Do all backbone polar groups in proteins form hydrogen bonds? *Protein Sci.* **14**, 1911-1917 (2005).

229 Zimmerman, S. B. & Trach, S. O. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J. Mol. Biol.* **222**, 599-620 (1991).

230 McGuffee, S. R. & Elcock, A. H. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput. Biol.* **6**, e1000694 (2010).

231 Wang, Y., Sarkar, M., Smith, A. E., Krois, A. S. & Pielak, G. J. Macromolecular crowding and protein stability. *J. Am. Chem. Soc.* **134**, 16614-16618 (2012).

232 Dill, K. A. Dominant forces in protein folding. *Biochemistry* **29**, 7133-7155 (1990).

233 Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface. *Biochemistry* **38**, 16419-16423 (1999).

234 Tzul, F. O., Schweiker, K. L. & Makhatadze, G. I. Modulation of folding energy landscape by charge-charge interactions: Linking experiments with computational modeling. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E259-266 (2015).

235 Smith, C. K., Withka, J. M. & Regan, L. A thermodynamic scale for the β-sheet forming tendencies of the amino acids. *Biochemistry* **33**, 5510-5517 (1994).

236 Becktel, W. J. & Schellman, J. A. Protein stability curves. *Biopolymers* **26**, 1859-1877 (1987).

237 Chick, H. & Martin, C. J. On the "heat coagulation" of proteins. *J. Physiol.* **40**, 404-430 (1910).

238 Fisher, E. Einfluss der configuration auf die wirkung den enzyme. *Ber. Dtsch. Chem. Ges.* **27**, 2985-2993 (1984).

239 Anson, M. L. & Mirsky, A. E. On some general properties of proteins. *J. Gen. Physiol.* **9**, 169-179 (1925).

240 Edsall, J. T. Hsien Wu and the first theory of protein denaturation (1931). *Adv. Protein Chem.* **46**, 1-5 (1995).

241 Mirsky, A. E. & Pauling, L. On the structure of native, denaturated and coagulated proteins. *Proc. Natl. Acad. Sci. U. S. A.* **22**, 439-447 (1936).

242 Ignatova, Z. & Gierasch, L. M. Effects of osmolytes on protein folding and aggregation in cells. *Methods Enzymol.* **428**, 355-372 (2007).

243 Bai, Y., Milne, J.S., L. Mayne and S.W. Englander. Protein stability parameters measured by hydrogen exchange. *Proteins: Struct., Funct., Genet.* **20**, 4-14 (1994).

244 Muñoz, V., Blanco, F. J. & Serrano, L. The hydrophobic-staple motif and a role for loop-residues in α-helix stability and protein folding. *Nat. Struct. Biol.* **2**, 380-385 (1995).

245 Blanco, F. J., Ortiz, A. R. & Serrano, L. Role of a nonnative interaction in the folding of the protein G B1 domain as inferred from the conformational analysis of the α-helix fragment. *Fold Des.* **2**, 123-133 (1997).

246 Pakula, A. A. & Sauer, R. T. Reverse hydrophobic effects relieved by amino-acid substitutions at a protein surface. *Nature* **344**, 363-364 (1990).

247 Harada, R., Tochio, N., Kigawa, T., Sugita, Y. & Feig, M. Reduced native state stability in crowded cellular environment due to protein-protein interactions. *J. Am. Chem. Soc.* **135**, 3696-3701 (2013).

248 Hecht, M. H., Sturtevant, J. M. & Sauer, R. T. Effect of single amino acid replacements on the thermal stability of the nh_2-terminal domain of phage λ repressor *Proc. Natl. Acad. Sci. U. S. A.* **81**, 5685-5689 (1984).

249 Green, S. M., Meeker, A. K. & Shortle, D. Contributions of the polar, uncharged amino acids to the stability of staphyloccal nuclease: Evidence for mutational effects on the free energy of the denatured state. *Biochemistry* **31**, 5717-5728 (1992).

250 Krowarsch, D. & Otlewski, J. Amino-acid substitutions at the fully exposed p_1 site of bovine pancreatic trypsin inhibitor affect its stability. *Protein Sci.* **10**, 715-724 (2001).

251 Wrabl, J. O. & Shortle, D. Perturbations of the denatured state ensemble: Modeling their effects on protein stability and folding kinetics. *Protein Sci.* **5**, 2343-2352 (1996).

252 Betz, S. F. & Pielak, G. J. Introduction of a disulfide bond into cytochrome c stabilizes a compact denatured state. *Biochemistry* **31**, 12337-12344 (1992).

253 Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333-366 (2006).

254 Liu, C., Sawaya, M. R. & Eisenberg, D. B2-microglobulin forms three-dimensional domain-swapped amyloid fibrils with disulfide linkages. *Nat. Struct. Mol. Biol.* **18**, 49-55 (2011).

255 Ruschak, A. M., Religa, T. L., Breuer, S., Witt, S. & Kay, L. E. The proteasome antechamber maintains substrates in an unfolded state. *Nature* **467**, 868-871 (2010).

256 Sanger, F., Nicklen, S. & Coulson, R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463-5467 (1977).