The weak chemical interactions that drive protein evolution: crowding, sticking and quinary structure in folding and function

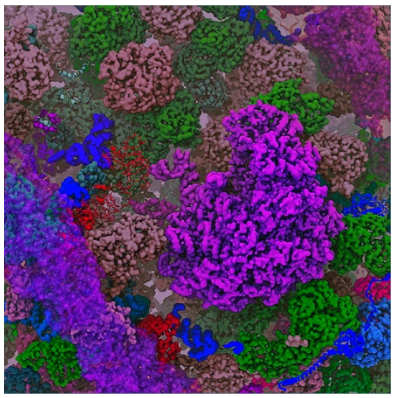
Drishti Guin[a] and Martin Gruebele[a,b,c,*]

[a]Department of Chemistry, University of Illinois, Urbana, IL 61801; [b]Department of Physics, University of Illinois, Urbana, IL 61801; [c]Center for Biophysics and Quantitative Biology, University of Illinois, Urbana, IL 61801.

AUTHOR INFORMATION
Corresponding author:
*mgruebel@illinois.edu

ORCID
Martin Gruebele: 0000-0001-9291-8123

Notes
The authors declare no competing financial interest.



In recent years, better instrumentation and greater computing power have enabled imaging of elusive biomolecule dynamics in cells, driving many advances in understanding the chemical organization of biological systems. The focus of this review is on interactions in the cell that affect both biomolecular stability and function, and modulate them. The same protein or nucleic acid can behave differently depending on time in the cell cycle, location in a specific compartment, or stresses acting on the cell. We describe in detail crowding, sticking and quinary structure in the cell and the current methods to quantify them both *in vitro* and *in vivo*. Finally, we discuss protein evolution in the cell in light of current biophysical evidence. We describe the factors that drive protein evolution and shape protein interaction networks. These interactions can significantly affect the free energy ΔG of marginally-stable and low-population proteins and due to epistasis direct the evolutionary pathways in an organism. We finally conclude by providing an outlook on experiments to come, and the possibility of collaborative evolutionary biology and biophysical efforts.

Table of Contents

1. Introduction

The cell was discovered in the 17th century by the Dutch scientist van Leeuwenhoek and observed in the simplest eukaryotic organism – a single-celled protozoan. Cells showcase many layers of biological organization, from membrane-bound organelles and phase-separated membrane-less 'liquid droplets' to the cytoskeleton. The structural organization of cells, from small solutes to organelles, is highly dynamic and far removed from the dilute buffers that are used in *in vitro* experiments.

The energetics of the cell cover a wide dynamic range, from interactions barely exceeding the thermal noise k_BT at room temperature, to strong chemical bonds. The protein and DNA backbones are made of stable bonds that are relatively hard to hydrolyze. Even the cellular cytoskeleton is held together by interactions on the order of 40 kJ/mol or 15 k_BT (the interaction free energy of tubulin dimers[1]). Although such specific and relatively strong interactions are easy to detect, weakly interacting (a few k_BT) and thus harder to detect multi-partner networks are more resilient towards malfunction (e.g. due to a mutation), and more flexible to evolve. While hubs in networks may be conserved, their many partners can evolve more independently.[2]

One of the hallmarks of complex networks is that many of their interactions are only on the order of a few k_BT. Weak interactions are numerous in cells, and can add up cooperatively, thus significantly affecting the spatial and temporal distribution of macromolecules in the cell. To fully understand these weak interactions, they are best studied directly in cells or even in living organisms, or at least in a solvation environment that mimics those aspects of the cell important for the interaction in question. We have only just discovered the tip of the iceberg of how biological organization emerges from and affects dynamics at the molecular level. Moving forward, quantitative in-cell and *in vivo* experiments will become more important to study weak interactions that are sensitive to their solvation environment.

In this review, we discuss three concepts that can be used to classify weak interactions inside the cell; 'crowding,' 'sticking' and 'quinary structure.' Crowding is due to the short-range repulsive wall of interaction potentials with macromolecules in the cell (Figure 1). The repulsive wall is mainly due to physical forces not subject to evolution, although the size distribution of macromolecules that exclude volume could evolve. Sticking is due to longer-range attractive forces among a combination of macromolecular surfaces and small solutes within the cell. The effective forces can be electrostatic or entropic in nature, and can evolve, for instance by changing the charge or hydrophobicity of an amino acid side chain on a protein surface. Sticking is not necessarily favorable to cell health, but when it evolves into a favorable interaction, the resulting transient structure is referred to as 'quinary structure,' as a continuation of the hierarchical levels of evolved protein organization: primary, secondary, tertiary and quaternary structure.

A number of literature reviews have discussed in detail the experimental work and the current state-of-the-art theoretical models that describe macromolecular crowding[3,4] and quinary structure.[5–8] We do not seek to recapitulate the comprehensive work of these published reviews.

Instead, the focus of this review is to inspect the available data on chemical forces from the point of view of evolutionary pressures that have shaped protein (and other biomolecular) interactomes. Evolution produces complex interaction networks robust to perturbation, but it has to use chemical interactions outlined in figure 1 as basic building blocks. Our main goal is to emphasize the importance of collaborative studies in evolution and biophysical chemistry to understand how structural and functional aspects of the cell have evolved. Finally, to conclude we provide an outlook on the future and the scope of further experiments in post-reductionist chemistry of the cell.

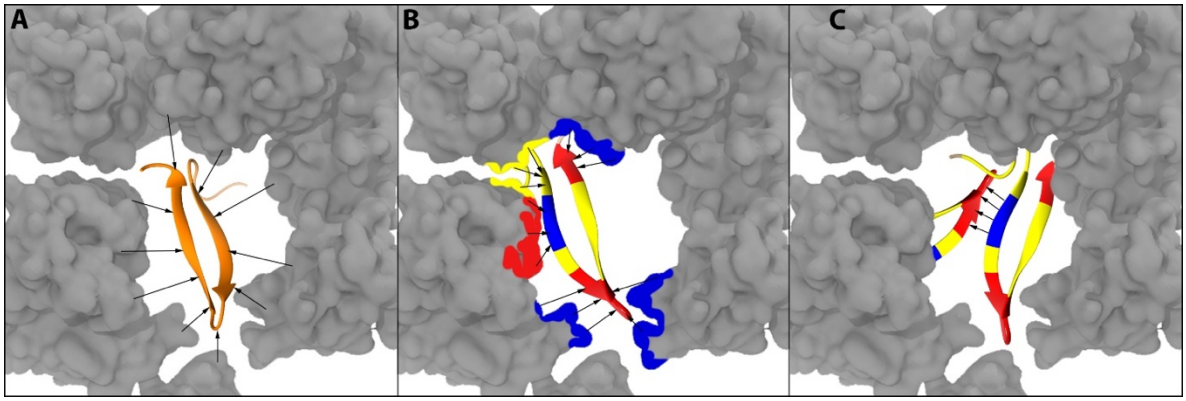


Figure 1: Cartoon representation of crowding, sticking and quinary interactions. The cytoplasm scaffold image was provided by Meredith Rickard. (A) Protein (orange) crowded by neighboring biomolecules (grey). (B) Ribbon structure of protein shows charged (negative – blue and positive – red) and hydrophobic patches (yellow). Patches on the surface of the protein interact with neighboring biomolecules via electrostatic and hydrophobic interactions. (C) Quinary functional interaction between two small proteins via surface electrostatic interaction.

2. Weak interactions in the cell – a brief history

In this section we discuss the three weak in-cell interactions (Figure 1) that contribute to organization near the k_BT level of thermal energy, namely crowding, sticking and quinary structure. All three of these can impact cell fitness; for example, crowding can enhance protein stability; sticking can lead to aggregation; and quinary structure of two enzymes can improve their substrate processivity. The terms macromolecular crowding and quinary structure were first introduced in the 1980s in the context of the cell. While the effects of crowding have been investigated rigorously *in vitro*, studies probing quinary structure have gained momentum only recently *in vitro* or *in vivo*. On the other hand, non-specific sticking has been known to exist since the earliest studies of protein aggregation.[9] However, its implications in cell have been observed only recently using in-cell NMR.[10]

2.1 Macromolecular Crowding

The bio-macromolecular concentration inside an *Escherichia coli* cell can reach over 300 mg/mL.[11] The crowding inside eukaryotic cells can be even higher due to the cytoskeletal

framework.[12] Macromolecular crowding was defined by Minton in 1983 as excluded volume effects exerted by the solutes in a solution.[13] In simple terms this means that solutes in a solution cannot overlap due to neighboring molecules' short-range repulsion, which is ultimately due to Pauli exclusion of atomic electrons on a length scale of ~ 0.1 nm.[14] About 30-40% of the cellular volume in eukaryotes is occupied by protein and nucleic acid molecules.[15] At these concentrations excluded volume effects can manifest as 1) macromolecular crowding: the volume excluded for one solute molecule by another and 2) macromolecular confinement: molecules confined to a smaller effective volume than the solution volume due to a high concentration of species. This situation can result in phenomena such as jamming, where macromolecular diffusion through interstitial spaces in the cytoplasm is severely hindered.[16]

Both crowding and confinement can have a significant effect on ubiquitous cellular functions such as association, activity, stability and conformation. For example, macromolecular crowding generally stabilizes proteins because the configurational entropy of the unfolded state is more severely reduced than that of the native state.[17–19] (It is worth noting that unfolded proteins have smaller molar volume than folded proteins *in vitro*, hence proteins denature at high pressure.) Similarly, crowding can also increase association constants by favoring lower entropy, lower energy complexes.[20–23] Although the many features of the cell such as pH, ionic strength, osmolarity and redox potential can be accounted for by using suitable buffers, the main feature of biological macromolecules missing from traditional buffered *in vitro* experiments is their size. The last decade has seen an explosion of experimental data characterizing the effects of macromolecular crowding by using large inert polymers to account for macromolecule size. However, cellular components are far from inert and interact with each other constantly. As a result, simple crowders are not necessarily good mimics of the in-cell environment.[24] In the next section we discuss these interactions as an effect of macromolecules that is complementary to crowding.

2.2 Sticking

In 2016, Pielak and coworkers found that the stability of the small protein SH3 remained unchanged in an *E. coli* cell using in-cell NMR.[25] This observation could not be explained by the simplest version of macromolecular crowding that stabilizes the folded state due to excluded volume effects. Moreover, excluded volume effects are purely entropic, but enthalpic contributions are often seen as a consequence of crowding.[4,26]

Macromolecular crowding and its effects have been characterized rigorously *in vitro* using inert synthetic polymers such as Ficoll, dextran or PEG. However, the interior of the cell is far from inert. Proteins and RNA are charged and interact via electrostatics (screened charge or polar interactions), hydrogen bonding, and hydrophobic interactions (which are partly entropic due to enhanced ordering of water molecules near aliphatic or aromatic amino acid side chains). These forces can together be grouped into the non-specific forces that lead to macromolecular sticking inside a cell. In 2017, Oliveberg and coworkers showed that two mammalian proteins that tumble freely in mammalian cells get stuck in the bacterial cytoplasm.[10] On the other hand native bacterial

proteins with the same fold tumble freely inside the bacterial cytosol. This is strong evidence that biomolecular surfaces are under evolutionary selection not just for residue-specific function, but also such that surface charge and hydrophobicity are optimized according to organism. For example, a freely tumbling and diffusing molecule can sample multiple potential binding partners in a short time. On the other hand, stickiness could evolve into a useful signaling interaction (see section 2.3). Either way it is clear that weak chemical forces in the cell can modulate stability,[27] activity[28] and diffusion.[29]

2.3 Quinary structure

The term 'quinary' was first used by Vaĭnshteĭn in 1973 to describe the fifth level of organization of proteins and nucleic acids in natural and synthetic aggregates such as those in viruses, chromosomes, molecular films etc.[30] In 1980, Edelstein also described quinary structure as being the fifth level of organization of protein subunits in helical lattices.[31] The term quinary structure as it is frequently understood today was introduced by McConkey, also as the fifth level of protein organization consisting of functional interactions that are weaker and more transient than quaternary protein structure.[32] Quinary structure can be hard to isolate or purify by harsh *in vitro* techniques. In McConkey's definition, quinary structure is a likely reason why most cellular proteins evolve slower than expected: protein surfaces are subject to maintaining functional interactions with a few (on the order of 1 to 20) partners, while at the same time avoiding interactions with thousands of other types of macromolecules in the cell.

McConkey also noted that even though hard to replicate *in vitro*, in his day there were already examples of quinary structure formation among cellular constituents. These included i) the ribosome that transiently interacts with many factors and ii) some of the cytoskeletal framework that was destroyed by the separation methods of the time. We now know of many more examples of transiently interacting complexes inside cells. This has allowed us to understand the functional role of quinary structure as well as the range of thermodynamic and kinetic parameters that make an interaction inherently transient. Quinary structure requires a balance of both thermodynamic stability and kinetics. If two states are of similar stability but have a high kinetic barrier, then the system will get trapped in each state for a long time. Similarly, if one state is significantly lower in free energy than the other, then the system will mostly end up in the lower free energy state. When both barriers and stability differences are on the order of a few $k_B T$ on the free energy scale, transient association results. Small free energy differences and barriers make quinary structure highly susceptible to modulation by the cellular environment, and changes in the cellular environment that occur during the cell cycle or stress.

3. Macromolecular Crowding

Over the years, the effects of macromolecular crowding on protein folding, assembly and other biological processes have been extensively characterized *in vitro* and *in vivo* by theoretical and experimental approaches.[33] Initial work by Zimmerman and Minton led to over two decades of

considerable research in the field. A number of reviews have comprehensively described the existing literature.[3,34–38] In this section, we briefly highlight the basic principles of crowding and confinement and discuss *in vitro* crowding agents in experiments referencing earlier reviews where more detail is needed. We discuss in a greater detail some recent developments in the field of theoretical and computational studies that have facilitated all-atom simulations of the cytoplasm. We finally conclude by detailing recent efforts in characterizing the state of crowding in the cell as well as the consequences of crowding in the pharmaceutical industry.

3.1 Excluded volume effect

Excluded volume effects, a direct consequence of crowding by biological macromolecules due to repulsive interactions at short range, are perhaps the best characterized interactions that occur in the cell on the thermal energy scale.[39] Crowding by macromolecules can be visualized in the simplest approximation as hard spheres, nearly close-packed in a fixed volume. The interstitial space between the spheres can be occupied by small solvent or solute molecules such as water and metal ions, but it cannot be occupied by other large spheres present in the packed structure. The same constraints prevent the addition of any more spheres into the volume. This interstitial volume is then said to be excluded for these spheres.

In the above example, the volume available to species of any shape or size is limited simply by the impenetrability of the hard spheres in the volume. For real macromolecules, short-range repulsion has a finite range, which adds further add to the excluded volume (Figure 2A). Hence, the introduction of a particle from a free into a crowded environment leads to a reduction in entropy which, when scaled by temperature, is equal to the work done to insert the particle. Therefore, systems are driven to minimize the excluded volume to increase entropy.

Rigorous descriptions of crowding due to hard particles in a fluid have been published.[40–43] The effects of crowding on folded biological macromolecules are well described by such simple models that treat macromolecules as hard particles with an effective coarse-grained size and shape. For example, unfolded proteins, can be treated as random chains which have a greater ability to thread through interstitial spaces as compared to folded proteins generally modeled as hard spheres. Models have been developed to mimic the effect that crowders have on biomolecule stability due to excluded volume.[44–46] Cheung and Thirumalai calculated that the change in melting temperature (T_m) for a protein with crowder volume fraction Φ_C can be estimated by:

$$\Delta T_m \approx 0.84 T_m(\text{Kelvin}) {\varphi_C}^{1.8} \quad [1]$$

The volume fraction Φ_C for spherical crowders with radii *R* comparable to the radius of gyration of the folded protein is

$$\varphi_C = \frac{4\pi}{3}(R)^3 \frac{M}{V}, \quad [2]$$

where M is the number of crowders and V is the volume of the smallest cubic box that can be drawn around the protein assuming a random coil denatured state.[44]

3.2 Crowding *vs.* confinement

Both crowding and confinement (Figure 2) are the consequences of macromolecular volume exclusion. Specifically crowding refers to the amount of free energy required to transfer a macromolecule from a dilute solution into a crowded environment. This is equivalent to the amount of energy expended to create a cavity large enough to accommodate the introduced macromolecule. Hall and Minton[36] showed that for a molecule "X" this free energy can be approximated by

$$\Delta F_X^{crowding} = -RTln\left(\frac{V_{availble,X}}{V_{total}}\right) = RTln(\gamma_X), \quad [3]$$

where $V_{available,X}$ is the volume available to a species "X" in a crowded environment, V_{total} is the total volume and $\gamma_X = \frac{V_{total}}{V_{availble,X}}$ is the activity coefficient of the species "X." Eq. [3] simply reflects the entropic cost of changing the available volume around a solute.

On the other hand, confinement refers to the free energy required to transfer a biomolecule from the set of configurations allowed in dilute solution to the set allowed in a bounded volume (such as a rigid cavity). Zhou and Minton[3] showed that for a molecule "X" this free energy change is given by

$$\Delta F_X^{confine} = -RTln(\frac{W_{allowed}}{W_{all}}), \quad [4]$$

Where, $W_{allowed}$ is the number of allowed configurations in a crowded solution and W_{all} is the number of all configurational states allowed in a dilute solution. Confinement is a somewhat more general concept than crowding, and reflects the entropic cost of reducing the number of configurations of a macromolecule, not just the cost of reducing the surrounding volume.

Crowding and confinement both limit the number of states macromolecules can occupy. Crowding favors more compact states of a molecule under a given set of solvation conditions. (However, the smaller molar volume of a random coil polypeptide relative to a native state will not generally lower its free energy sufficiently because of its reduced configurational entropy in a smaller free volume.) Confinement doesn't always prefer the most compact state. Rather confinement favors conformations that have shapes complementary to the shapes of the confining cavity. For example, spherical objects fit better in spherical pores whereas rod-like shapes fit better in a more cylindrical pore. The effect of confinement on both protein folding and association has been discussed in detail.[12,47]

3.3 *In vitro* crowding agents

The effects of crowding have been characterized extensively *in vitro* using chemical agents to mimic a crowded environment. This section briefly discusses popular crowding agents and their advantages and disadvantages. Readily available polysaccharides such as Ficoll and dextran, or the poly-ether poly(ethylene glycol) (PEG) are the most popular polymeric crowding agents.[37]

Each polymer is available in various molecular masses to mimic the size of different biomacromolecules. Both Ficoll and dextrans interact with proteins mostly via repulsive excluded volume effects. However, a number of studies have shown that for PEG this repulsive interaction is compensated to some extent by attractive interactions with hydrophobic and non-polar side chains on the protein surface.[48–52] For example, PEG seeks out mixed hydrophilic/hydrophobic patches on the protein surface and forms transient but recurring structure around them.[53] Thus even these simple crowders do not necessarily act like structureless spheres.

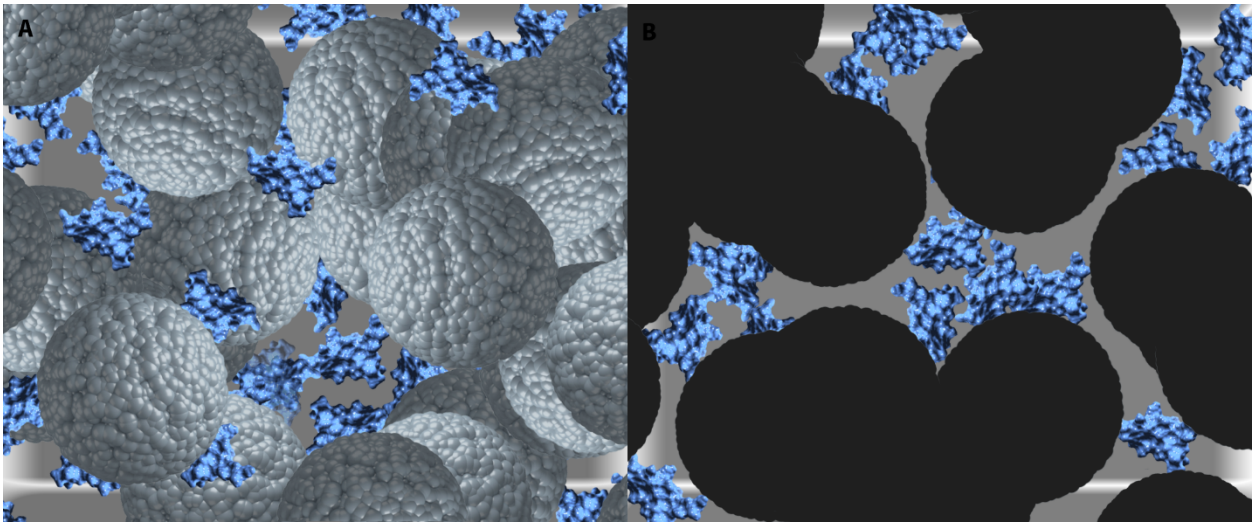


Figure 2: A qualitative representation of (A) crowding and (B) confinement. (A) Crowding agent represented by grey balls with a model protein in blue. Both crowder and protein can freely diffuse in solution. (B) Protein in blue is confined to pores by a matrix shown in black.

Crowding studies in general seek to mimic cellular congestion and the effect of congestion on other macromolecules. The intracellular space contains biomolecules such as proteins, DNA and RNA of various shapes and sizes. Most of the crowders described above are used because of the ease of availability and experimental manipulation. These non-biological agents are unlike the crowding biomolecules inside cells and might not, therefore, provide physiologically relevant information.[27,34,54] To probe the effect of biological macromolecular crowding, biomacromolecules such as proteins have also been used as crowders extensively.[20,22,23,28,54–56] Some of the commonly used proteins are bovine serum albumin (BSA),[20,23,55] SubL,[54] hemoglobin,[28] RNase A,[57] β-lactoglobulin[57] and lysozyme.[20] Numerous studies have shown that contrasting effects are observed for non-biological polymeric crowding agents and biological crowders on a wide variety of reactions such as enzyme activity, protein refolding and denaturation.[28,56,58–60] The difference between a polymeric crowder such as Ficoll and a protein crowder such as BSA is, in most cases, explained as a consequence of the difference between their shapes or sizes. For example, Derham and Harding concluded that in protein crowders, the formation of higher activity oligomeric states due to the excluded volume effect showed an initial increase in enzyme activity.[28] However, in polymer crowders a reduction of activity was seen across all concentrations of crowders. They postulated that this is because the reduction in molecular diffusion is higher for

polymers such as dextran or PEG as compared to globular protein crowders. This lower diffusion in polymeric crowders was thought to cancel out any positive contribution of oligomeric states to activity. Similar results were also observed for a small protein SubL on the unfolding of lambda repressor mutant 6-85,[56] and BSA on refolding of egg lysozyme[58,60] *vs.* non-biological crowders, the protein crowders being more effective.

In highly heterogenous environments, diffusion can be complex and is sometimes described as "anomalous diffusion". This occurs when the mean squared displacement, $\langle r^2(t)\rangle$, follows a power law of the form

$$\langle r^2(t)\rangle = 6\mathrm{D}t^{\alpha}, \qquad [5]$$

where α is degree of deviation from normal diffusion given by $\alpha = 1$ and *D* is a constant that does not depend on time. In recent years, several diffusion studies of crowding (and sticking) effects have been conducted *in vitro* by tracking the movement of a fluorescent tracer molecule using Fluorescence Correlation Spectroscopy (FCS).[61–63] These tracer molecules are either labeled with small fluorescent probe molecules such as Alexa488 and fluorescein or with a fluorescent protein such as EGFP. The main thrust of this research area is to characterize the diffusion behavior of macromolecules in crowded solutions. However, the results from these studies are contradictory, with some groups arguing for anomalous diffusion[63] and others for normal diffusion.[62] A recent study using hole-burning[64,65] of fluorescence intensity coupled with whole-cell imaging showed that anomalous diffusion best describes the overall flow of the proteins GFP and FRET-labeled PGK in mammalian cells, although a good semi-quantitative description of the diffusion of folded protein could be obtained by a normal diffusion model.

Even with these limitations, the diffusion behavior in heterogenous, crowded environments can shed light on the behavior of macromolecules inside the cell and the deviations from ideality reported in many crowding studies *in vitro*.

3.4 Towards theoretical tools to characterize crowding

Many theoretical models have predicted the effect of crowding based on statistical mechanics. Eq. [1] is an example. These statistical models, in most cases, assume that crowders are inert and do not interact, other than through purely repulsive forces, with the macromolecules of interest. These theories have been extensively described and reviewed .[3,34,66]

The Zhou and Minton groups have used simple crowding models where both protein and crowder can be modeled by effective hard spheres or rods. The Zhou group used experiments as well theoretical calculations[67–69] based on scaled particle theory (SPT)[40] and Widom's particle insertion method[43] to probe the effect of crowding on protein stability[70] and binding[71] to another protein as well as on membrane proteins.[72] Using the SPT for hard spheres, Zhou predicted the free energy change due to crowding of a polymer chain as [69]

$$\frac{\Delta F_U^{crowd}}{RT} = -ln(1-\varphi) + 3\varphi Y^2\left(1+\frac{1}{Y\sqrt{2\pi}}\right). \qquad [6]$$

The first term is similar to eq. [1], written in terms of the excluded volume fraction φ. Y is the ratio of the radius of gyration of the unfolded chain in dilute solution to the radius of the crowder. For large biomolecules or small crowders, the correction to simple excluded volume is thus bilinear in both excluded volume fraction and ratio of biomolecule to crowder size, and increases the crowding free energy further. The Minton group has also used statistical mechanics-based calculations relying on SPT to predict the effect of crowding.[36,45,46] Recently, Pielak and coworkers used SPT to predict the magnitude of the stabilizing effect of hard-core repulsion on two distinct shapes.[73] Crowders such as sucrose, BSA and lysozyme were modeled as hard sphere crowders using SPT and the magnitude of stabilization on an elongated dumbbell-shaped *vs.* a more compact dimer was calculated. In all their calculations the more compact dimer showed a higher degree of stabilization. Moreover, crowding effects were slightly destabilizing for highly elongated dimers, such as where the monomers are only touching. This reiterates the observation that the crowded cellular milieu prefers a more compact shape than an elongated shape and could be part of the reason why a globular protein like PGK is stabilized in the cytoplasm whereas the more elongated VlsE is destabilized.[24,27]

Moving away from these simple analytical models,[74] the Cheung and Thirumalai groups used molecular simulations to fold a small, fast-folding, model protein WW-domain in a crowded environment.[44] The Cheung group also conducted similar molecular simulations studies in collaboration with experimental groups of Wittung-Stafshede[74,75] and Waxham.[76,77] In 2012, Cheung and Wang used the all-atom cytoplasmic model of Elcock and McGuffee and developed an algorithm that coarse-grains slices of the cytoplasm.[78] They studied the thermodynamic properties of apoazurin in this coarse-grained cytoplasm and compared it to two models where each macromolecule in the cytoplasm is replaced by either (1) hard spheres of 55 Å, the size of Ficoll70, or (2) hard spheres of the same volume as the macromolecule. While the hard sphere model allows heterogeneity in only the size, the coarse-grained cytoplasm has heterogeneity in both size and shape of crowding macromolecules. As was expected, the size of the crowding agent affects the predicted thermodynamic parameters; the stability of apoazurin was ~5 °C lower in Ficoll70 model than in either the variable-size hard sphere model or the coarse-grained cytoplasm. However, the variable-size hard sphere model could not adequately reflect the extent of structural fluctuations observed in the coarse-grained cytoplasm even though the average values of the thermodynamic parameters are similar in both. This concludes that not just crowder size but also crowder shape contribute to crowding effects.

These methods, while computationally more expensive, can in principle provide more accurate fitting models for experimental observations of kinetics and conformational changes induced by the environment.

3.5 Characterizing crowding inside cells

Fluorescence is a tool used extensively to measure observables inside living cells, partly due to the ease of tagging cellular components with fluorescent labels, and its relatively non-perturbative nature in cells.[79] Fluorescence techniques such as FCS, Fluorescence Recovery After

Photobleaching (FRAP) and Förster Resonance Energy Transfer (FRET) have been used to characterize crowding in cells. FCS has been used widely to measure diffusion to indirectly quantify crowding inside cells, for example via changes in diffusion.[29,80] Weiss and Guigas used FCS to measure the viscoelastic properties of Alexa488 labeled gold colloidal particles (5 nm diameter) in a variety of cell lines from various organisms and at different states of health. They then used viscoelasticity as a measure of crowding and surprisingly, found a lower degree of crowding in the nucleus than the cytoplasm despite the high DNA content. They also found, not surprisingly, that diffusion was size-dependent. For example, the colloidal particles or ~6 nm diameter BSA both show anomalous diffusion whereas much smaller species such as GFP diffused normally.

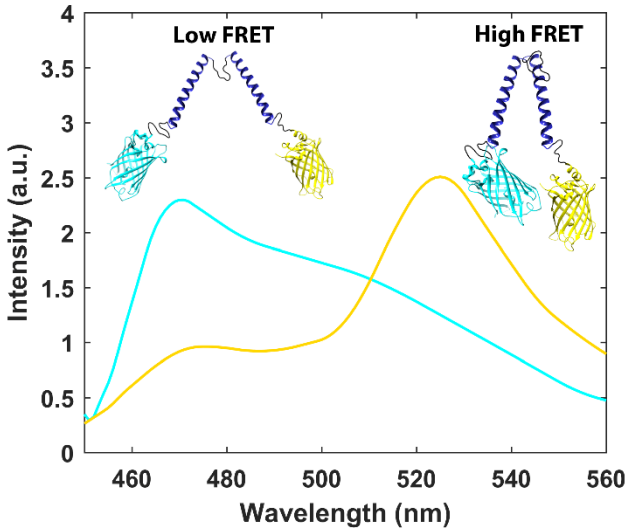


Figure 3: An in-cell crowding sensor designed by Poolman and Boersma highlighting the low- and high-FRET states. Data was taken from reference 79. The sensor consists of two helices connected via a flexible linker (inset) labeled with mCerulean at the N-terminus and mCitrine at the C-terminus. In dilute solutions the sensor adopts a low-FRET state where the two fluorescent tags are further apart (left inset), and fluorescence is dominated by mCerulean fluorescence (blue trace). In a crowded environment the two helices are pushed closer together adopting the high-FRET state (right inset). Fluorescence is dominated by mCitrine fluorescence due to mCerulean-mCitrine FRET (yellow trace).

A newer area of research in the field is the development of molecular crowding sensors that enable direct measurement of crowding inside cells.[81,82] Poolman and Boersma designed FRET-based crowding sensors (Figure 3) that report on the crowding-induced conformational changes in the protein.[81,83] The sensors contain an α-helical peptide labeled with the FRET pair mCerulean3 and mCitrine at the N- and C-terminus, respectively. The protein adopts more compact high-FRET conformations in crowded environment and can be genetically encoded in various cell types. Using this sensor, the authors were able to calibrate the crowding in *E. coli* cells and found it to be ~180 mg/g with a volume fraction of 0.13. Poolman, Boersma and Liu later designed a set of sensors that are useful where in-cell calibration of the sensor is prohibited, for

example during time-lapse measurements.[83] A set of nine probes was designed with the same FRET pair but varying linker sizes. Six of the nine probes showed efficacy for in-cell measurements. Other sensors, including PEG-based ones developed by Ebbinghaus and coworkers, use similar FRET-based approaches to measure crowding.[39,82,84]

The capability of some sensors to be genetically encoded enabled the measurement of free volume change inside cells due osmotic pressure.[85] These sensors could potentially be used to measure the heterogeneity of crowding in subcellular compartments via localization tags as well as on phase-separated membrane-less organelles. Such studies could yield key information about crowding inside the cell and its effects on modulation of biological processes.

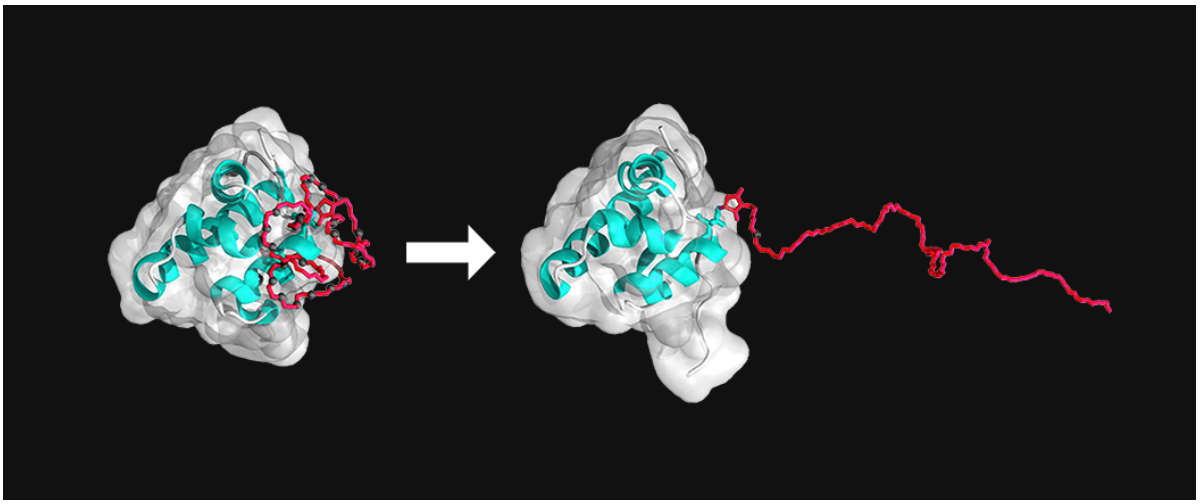


Figure 4: The interaction between PEG and $\lambda_{6\text{-}85}$. Calculations for this figure are described in reference 51. Figure shows PEG interacting with protein surface on the left and extended into solution on the right.

3.6 Crowding outside the cell

The apparent stabilization of proteins in crowded environments has been harnessed by the pharmaceutical industry in protein- and peptide-based drug therapies. Even though protein drugs hold great promise, they are easily degraded by proteolytic enzymes, cleared by the kidneys, have a short circulating half-life and generate neutralizing antibodies. One of the most widely used methodologies to circumvent these problems is PEGylation, or the addition of PEG chains to proteins. The first evidence of improved drug delivery through PEGylation was shown by Frank Davis and his colleagues in 1970.[86] A detailed review of the use of PEG to improve drug delivery was recently published by Chess and Harris.[87] PEG interacts with proteins via transient interactions at the protein surface that can be specific in nature.[51,53] Gruebele and Chao showed that when hooked to the protein surface, PEG preferentially interacted with specific patches on the surface of $\lambda_{6\text{-}85}$ irrespective of the attachment site thereby stabilizing particular secondary structures. Furthermore, PEG becomes structured near lysine residues that are surrounded by hydrophobic amino acids (Figure 4). This happens as PEG oxygen can hydrogen bond with lysine

sidechains while the methylene groups interact with the hydrophobic amino acids. This strategy could be potentially useful in designing drug targets that require stabilization at particular sites or to improve efficacy of PEG to stabilize particular proteins.

Another area of research is to study proteins in gel matrices. Entrapment of biological molecules in gels have been used in many processes such as drug delivery,[88] sensors,[89,90] separation methods,[91] and microfluidics,[92] *etc*. Bulk measurements in gel matrices show stabilization for apomyoglobin[93,94] and lysozyme,[93] among others. However, here too the picture of inert crowding may not be as simple as with PEG: it was recently found that the behavior of a model protein, PGK, at the gel surface was not the same as in the bulk of the matrix.[95] It was also found that the interaction of the gel with PGK, and not degree of confinement, influenced protein properties in gels. Likewise, zwitterionic polymers, originally thought not to interact significantly with proteins, have been shown to affect protein stability and unfolded state compactness via interactions with the protein surface.[96] Gels, undeniably, have potential benefits in many applications relying on enhanced protein stability. However, their interaction with proteins need to be explored beyond simple crowding and confinement effects.

4. Sticking

In the cytoplasm, the packing of macromolecules causes what we simply refer to as crowding; we already alluded in the previous section to limitations of the simple steric picture. Crowding has been characterized *in vitro* using inert macromolecules and polymers. However, crowding macromolecules in the cell include proteins and nuclei acids that contain charged, polar, and nonpolar patches. Proteins and RNA surfaces hence are not inert repulsive walls, and interact via electrostatics (charged and polar interactions), hydrogen bonding and hydrophobic interactions. These somewhat longer-range forces, sometimes attractive and sometimes repulsive, are the cause of stickiness of the cytoplasm. In this review we refer to these interactions collectively as "sticking" in the cell.

4.1 The thermodynamic consequences of sticking

The change in free energy ΔG^0 of a protein has both entropic and enthalpic contributions according to the Gibbs-Helmholtz equation (N = native, U = unfolded)

$$\Delta G^0_{N\to U} = \Delta H^0_{N\to U} - T\Delta S^0_{N\to U}, \qquad [7]$$

and these can be separated by the temperature dependence of the free energy. Crowding manifests as excluded volume effects in the cell. As described above the contribution of excluded volume effects is purely entropic. However, enthalpic effects are often seen in presence of crowders. These enthalpic effects can arise from non-specific forces that include electrostatic attraction and repulsion.

In-cell NMR is a non-perturbative technique to characterize these forces inside the native environment of the cell.[97] One such case was reported for the SH3 domain of the *Drosophila* signal transduction protein using in-cell NMR. Pielak and coworkers studied the stability of SH3 by fluorine labeling of the tryptophan residue in *E. coli* cells and compared it to the stability in polymeric crowding agents.[25] As expected, SH3 was stabilized in Ficoll, dextran and PEG but inside cells no such stabilization was observed. On the contrary both T_m and ΔG^0_u either decreased or were unchanged as compared to buffer. Similar results were obtained in cell lysate where no stabilization was observed. These results point to a destabilizing effect that counteracts the entropic stabilization from crowding. Indeed, when polycationic protein crowders such as BSA and lysozyme were used with the polyanionic SH3, a destabilizing effect was also observed. The degree of destabilization reduced when NaCl was used to screen the charges or the overall positive charges on the protein crowders was decreased by altering the pH. These results indicate that weak non-specific attractive interactions can effectively counteract stabilizing entropic effects due to crowding.

Similar results were also observed for wild-type GB1 protein, where attractive interaction due to mutating a surface aspartic acid to lysine destabilized GB1 in *E. coli* cells.[98,99] Moreover, while this mutation was innocuous in buffers, the average effect seen in the cell is 10-fold larger. GB1 and *E. coli* proteins (on average) are slightly negatively charged, hence mutating a negatively charged aspartic acid to a positive lysine changes the overall charge of the protein by +2 and therefore increases attractive interaction with other cellular proteins. Such charge reversal mutations have also proved useful to differentiate non-specific from functional interactions using in-cell NMR for other proteins.[100]

The stickiness of the cytoplasm can therefore negatively affect protein stability and function. These effects abound in the cell where the bio-macromolecular concentration is very high and where these macromolecules are not inert. Attractive interactions with other biomolecules can destabilize proteins because the number of favorable interactions can increase as the protein unfolds. As we will discuss later, stickiness in the cell caused by a random mutation could however evolve into new favorable quinary structure.

4.2 Evidence of stickiness *in vitro*

Several *in vitro* studies point towards the existence of these weak hydrophobic and electrostatic interactions and hydrogen bonding that cause sticking even outside the cell. The effects of sticking have been characterized by studying protein stability in buffers with globular proteins as crowders such as BSA and lysozyme.[97,101,102] In this scenario sticking is due to either attractive or repulsive interactions between charged biomolecules. Attractive interactions are destabilizing as they favor exposure of more surface which leads to unfolding. For example, negatively charged $SOD1^{I35A}$ (-0.5 e) and GB1 (-4 e) are destabilized in positively charged lysozyme (+8.5 e). Repulsive interactions stabilize the native state by reinforcing hard-core repulsion due to excluded volume. For example a negatively charged SOD1 dimer (-5 e) is slightly stabilized in BSA (-8.5 e) due to non-specific electrostatic repulsion.[102] However, in many cases the stabilization due to the net

charge on the interacting proteins is easily overcome by non-specific localized attractive interactions between protein surfaces. Lysozyme (+8.5 e), BSA (-8.5 e) and an anionic lysate destabilize chymotrypsin inhibitor 2 (CI2).[97,103,104] Similar effects have also been seen for ubiquitin in BSA and lysozyme where the stability is either decreased or unchanged and for $SOD1^{I35A}$ in BSA where the stability is largely unaffected.[105]

Protein self-association due to sticking has been a major problem in the biopharmaceutical industry.[106,107] Aggregated proteins in pharmaceutical applications can have far-reaching consequences on patient health. Aggregated protein in the blood stream can persist and cause an immune response leading to the patient becoming immune to the drug or in worst-case scenarios acquiring an autoimmune disease.[108–110] The formation of weak transient non-specific complexes due to sticking precedes the formation of more stable protein aggregates. These protein self-associated transient complexes are difficult to isolate as stable intermediates and must be studied using a method that is sensitive to low-population states undergoing very fast assembly-disassembly kinetics *in situ*.

Paramagnetic relaxation enhancement (PRE) is one such technique. For example, the human growth hormone (hGH) is involved in many regulatory processes. GH deficiency (GHD) causes slow muscular development and stunted growth and is generally treated using hormone replacement therapy.[111] One of the main challenges in hGH therapy is the propensity of hGH to form soluble dimers, trimers and higher oligomers as well as insoluble aggregates.[112] Consequently the aggregation of hGH has been extensively studied.[112–114] Using PRE, Led and coworkers showed that hGH forms transient weakly associated complexes that give way to longer-lived aggregates as the concentration increases.[115,116] In order for aggregation to occur, multiple sites on the protein surface interact via weak non-specific interactions with other hGH molecules (K_d = 0.9 mM). Similar ultra-weak self-association ($K_d \geq 15$ mM) was also observed for an *E. coli* protein histidine-containing phosphocarrier protein (HPr).[117] These metastable oligomeric species could act as nucleation events for the formation of higher order aggregates such as amyloid fibrils or viral capsids. Transient oligomeric species are in rapid equilibrium with the monomer and can either be assimilated into a higher order aggregate or dissolve back into the solution. PRE measurements are a useful tool to characterize these early metastable intermediates that form due to sticking.

4.3 *In vivo* methods to characterize stickiness

In-cell spectroscopy has gained considerable momentum in the last decade driven by the need to study biomolecules in their native environment. While *in vitro* experiments have been indispensable to our understanding of biomolecule structure and function, a large part of the interactome relies on weak interactions that are disrupted *in vitro* unless a native-like solvation environment can be painstakingly reproduced. In this section we discuss the two main methods to characterize sticking *in vivo*: 1) Fluorescence microscopy and 2) In-cell NMR.

Fluorescence microscopy is a time-honored technique to study biomolecule dynamics *in vivo*. The ease of availability of fluorescent labels as well as robust and flexible tagging methods have made fluorescence microscopy the method of choice for studying many different processes *in vivo*. These include but are not limited to diffusion,[29,80,118] binding[119,120] and stability[27,85,121,122] of biomolecules. Moreover, the availability of different fluorescence techniques such as Förster Resonance Energy Transfer (FRET), fluorescence lifetime imaging (FLIM),[123] fluorescence correlation spectroscopy (FCS),[61,124] or fluorescence recovery after photobleaching (FRAP)[125] can identify different aspects of the tagged biomolecule dynamics. An important outcome of the stickiness of the cytoplasm is the slowdown of biomolecules as they diffuse through the cytoplasm. This can be easily visualized by fluorescently labeling biomolecules and then tracking them as they diffuse through the cytoplasm.[126] Gruebele and Guo measured the diffusion of a GFP-labeled protein, phosphoglycerate kinase (PGK), in both the folded and unfolded state inside cells and saw anomalous diffusion.[29] They also observed that unfolded PGK diffused slower than folded PGK in the cell. It was concluded that this slow-down could not be explained by the larger hydrodynamic radius of the unfolded protein chain alone, but required sticking of exposed hydrophobic patches to other cellular constituents.

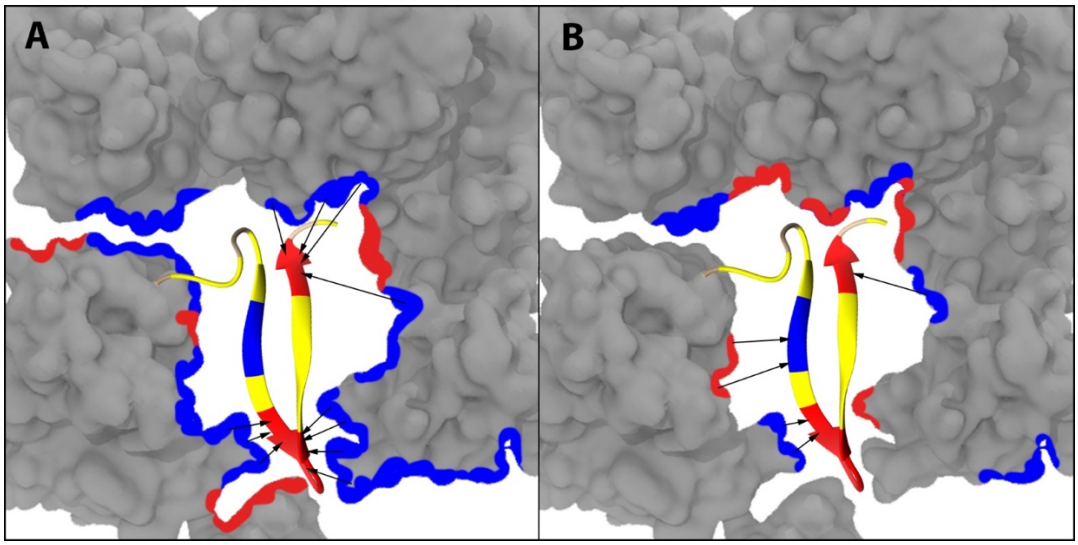


Figure 5: Cartoon representation of sticking in a (A) bacterial *vs.* (B) mammalian cytoplasm (grey). Arrow lengths signify interaction strength, longer arrows show weaker interactions and vice versa. Cytoplasm scaffold image was provided by Meredith Rickard. (A) A mammalian protein in a bacterial cytoplasm. Many bacterial proteins are negatively charged and proteins from other organisms such as mammals may get stuck and show slowed diffusion due to attractive electrostatic interactions. (B) A mammalian protein in a mammalian cytoplasm. A mammalian cytoplasm has a distribution of charges that is evolved to be compatible with native mammalian proteins. The cartoon shows a smaller number of negatively charged patches that reduces sticking.

While fluorescence techniques have been extensively diversified for in-cell applications, they are somewhat limited when conformational dynamics of biomolecules needs to be monitored. Most of these techniques also involve using fluorescent proteins that are bulky and could perturb the

tagged biomolecule.[127] In the last decade, in-cell NMR has been developed to complement fluorescence microscopy.[100,128–130] Isotopically labeled proteins for in-cell NMR can be both over-expressed or microinjected directly in the cell. Generally, NMR line broadening is a signature of interaction between biomolecules and can be used as such in-cells to characterize both functional and non-specific interactions. Non-specific interactions are easy to visualize when exclusively eukaryotic proteins are studied in prokaryotic organisms such as *E. coli*. In this case most functional partners for the protein of interest are absent and only non-specific interactions occur.[100] In-cell NMR and mutagenesis have been used to predict the role of specific surface charges to non-specific electrostatic interactions as explained in section 4.1.[97–99,101] Gierasch, Wang and Zhuravleva showed that these non-specific interactions are not modulated by a single variable, but are a function of several factors like overall charge, distribution of hydrophobic patches and conformational flexibility.[99]

In most cases, sticking has a negative consequence for protein stability or function. For example, ProtL,[131] SOD1,[102] and GB1-D40K[98] are destabilized in-cells as compared to buffer. This is evidence that entropic stabilization from crowding can be overcome by destabilization due to sticking. Furthermore, there is also evidence that sticking can lead to disruption of functional interactions. Oliveberg and coworkers showed that two mammalian proteins HAH1 and SOD1 diffused freely in mammalian cells but seemed to get stuck in the bacterial cytoplasm (Figure 5).[10] A freely diffusing protein can form functional interactions with its partners. A slowdown such as the one observed for SOD1 and HAH1 in *E. coli* can reduce the sampling of such interactions.

In-cell measurements are just starting to scratch the surface of the widespread non-specific interactions that are abundant in the cell. The usability of in-cell NMR is somewhat limited to studying eukaryote-specific proteins in prokaryotes. Eukaryotic proteins that are conserved in prokaryotes pose a challenge because very strong functional interactions can broaden lines beyond detection. However, in-cell NMR has been successful in eukaryotic systems by mutating out residues in the binding pocket to reduce broadening due to strong interactions. For example, Shirakawa and coworkers were able to record in-cell NMR spectra of a ubiquitin derivative (Ub-3A), whose affinity to cytosolic proteins was reduced by making three point mutations in the binding interface.[132] While wild-type ubiquitin showed peak broadening due to interactions with endogenous proteins *in vivo*, Ub-3A showed well-resolved peaks. Hydrogen exchange in combination with NMR also showed that binding to proteins in the cell destabilize wild-type ubiquitin *in vivo* compared to *in vitro*. Interestingly, even though Ub-3A has lower binding affinity, it is still destabilized *in vivo* due to sticking.

4.4 Towards creating a sticky cytoplasm using theoretical models

A new area of theoretical advancement is the explicit coarse-grained or even atomistic treatment of protein-protein interactions in crowded environments. In 1996, Field and Bicout published the first research study on modeling the cytoplasm of an *E. coli* cell.[133] The Field version of the cytoplasm consisted of three particles, ribosomes, proteins and tRNA, modeled as spheres. The interactions between these particles consisted of short-range Lennard-Jones and long-range

electrostatic terms. In 2008, Ellison and coworkers developed another model of the bacterial cytoplasm which also represented proteins as spheres.[134] This model was an important step forward from the model of Field and Bicout in that the model included >100 proteins at physiologically relevant concentrations. However, due to the limitations of this model it could not reproduce the *in vivo* diffusion rate of GFP accurately. Contrary to the 10-fold decrease in diffusion *in vivo*[135] this model only produced a 2-fold reduction in diffusion rate. The authors hypothesized that since the model's prediction was based solely on steric repulsion due to excluded volume between macromolecules, there are perhaps other effects inside cells that need to be accounted for to get more accurate results. As we will see later, sticking due to attractive potentials is one such effect.

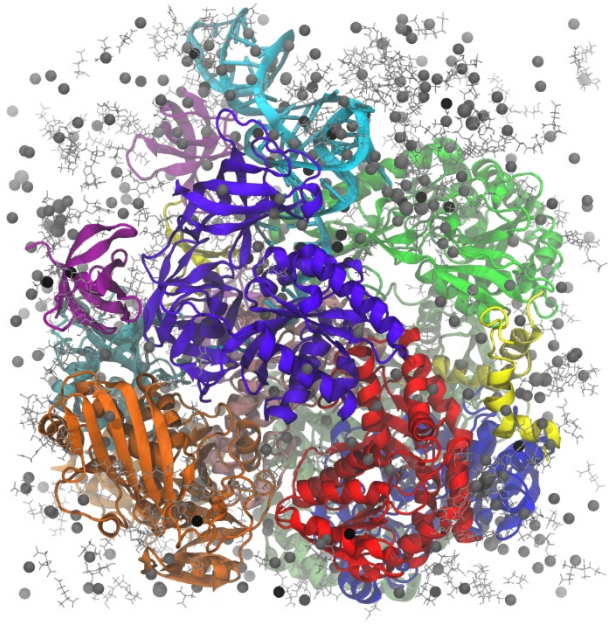


Figure 6: A snapshot showing proteins packed into a theoretical model of the bacterial cytoplasm. Image was provided by Dr. Taras Pogorelov and Meredith Rickard. The model consists of the most abundant cytosolic proteins, metabolites, ions and water molecules in an *E. coli* cell. Such a model allows for both crowding and sticking via protein surface charges.

In recent years, other more realistic models for crowding and sticking have been developed. In 2010, Elcock and McGuffee developed an all-atom model of the *E. coli* cytoplasm (Figure 6). The model consists of the 50 most abundant macromolecules of the *E. coli* cytoplasm, 45 of which are proteins. Moreover, this model considers two of the most common types of interactions in the cell *i.e.* hydrophobic and electrostatic. With these additional considerations Elcock and McGuffee were able to provide a quantitative rationalization of the destabilization of CRABP *in vivo* and $\lambda_{6\text{-}85}$'s unchanged stability *in vivo* (*vs. in vitro*). In these cases, sticking of the unfolded protein counteracts the crowding-enhanced stability of the native state. However, this all-atom cytoplasmic model is very computationally intensive and still limited in its uses for simulating cellular phenomena.

By 2015, Feig and coworkers presented another atomistic model of a cytoplasm based on *M. genitalium*.[136] This model of the cytoplasm is essentially complete and consists of all the components necessary for protein translation, folding and degradation in addition to the metabolic core functions. Using this methodology Feig and coworkers were able to construct complete metabolic pathways in a cytoplasmic subsection. With the increase in computational power in leaps and bounds over the past decade, the drawbacks of these all-atom simulations owing to their time-intensive nature could be resolved in the near future and allow for the simulation of biological reactions in a highly detailed model of the cell.

5. Quinary Structure

The term 'quinary' was used by McConkey to describe transient functional assemblies present in the cell.[32] Proteins show four main levels of organization: primary (amide bond formation, $-\Delta G$ = 8-16 kJ/mol in solvent), secondary (hydrogen bonds during α-helix and β-sheet formation, $-\Delta G <$ 1 kJ/mol in solvent), tertiary (disulfide bond formation, hydrophobic core, salt bridges etc., $-\Delta G$ = 0-0.1 kJ/mol/residue in solvent) and quaternary (hydrogen bonding, electrostatics and hydrophobic interactions in protein oligomerization, $-\Delta G$ = 40-60 kJ/mol). Quinary structure is the 5th level of protein structural organization, where proteins interact weakly ($K_d > 1$ μM) and form short-lived functional complexes in the cell.[85] Quinary structure is characterized by low thermodynamic stability and a low kinetic barrier, but unlike sticking, it has useful functional consequences that improve cell health. The electrostatic, hydrophobic and other interactions that underlie quinary structure formation are referred to here simply as 'quinary interactions.'

In this section we briefly describe the advances in quinary structure determination and our current knowledge of the existence of quinary structure formation in biological systems. We briefly describe the concept of the metabolon. The metabolon is one the most well-characterized instances of quinary structure formation in cells. In the last decade technical advancements have led to the discovery of many more examples of quinary structure. We describe in detail two such cases where quinary structure may play key roles in regulating biological function: 1) phase separation in living systems and 2) formation of encounter complexes.

5.1 What is and isn't quinary structure?

For an interaction to be quinary it must satisfy three conditions: 1) low stability of the complex, 2) rapid kinetics of dissociation/association and 3) interaction should confer some functionality. For a very high kinetic barrier the interaction wouldn't be transient, *i.e.* the complex may get trapped in one state for a long time. Similarly, for a highly stable state, the system will more often end up in that state with the interaction being more long-lived than transient. Both above-mentioned scenarios lead to tightly bound stable complexes instead of transient quinary structure. And if there is no significant function, the interaction is merely 'sticking'. These features make quinary structure highly susceptible to disruption by *in vitro* biochemical separation methods.

In evolutionary terms, it is possible that sticking evolves into functional quinary structure, which then evolves into more specific, stronger interactions. The difference between sticking and quinary structure is not the energy scale, but that the former is frequently a destabilizing force without any functional contribution, whereas the latter endows functionality. One possible reason that most quinary structure has not evolved greater specificity is that proteins are involved in networks with multiple binding partners. A protein's surface is finite in extent, and can only accommodate so many binding partners: either a few strong ones (larger surface area occupied by each binding interaction) or more weaker ones (smaller surface area occupied by each of the binding partners). In terms of information theory, the surface of a protein can only encode a certain number of bits of information, and these must be divided up among multiple binding interactions in a zero-sum game. If we take an amino acid surface area $A_A \approx 10$ Å^3 and a protein surface area A_P, and assume that either +, -, polar or non-polar can be encoded on each amino acid patch (four states), then the total information is $\sim(A_P/A_A)^4$, or in terms of an information entropy (proportional to number of bits because $\ln[x] = \ln_2[x]/\ln[2]$),

$$S_{quinary} \sim 4R ln[\frac{A_P}{A_A}]. \quad [8]$$

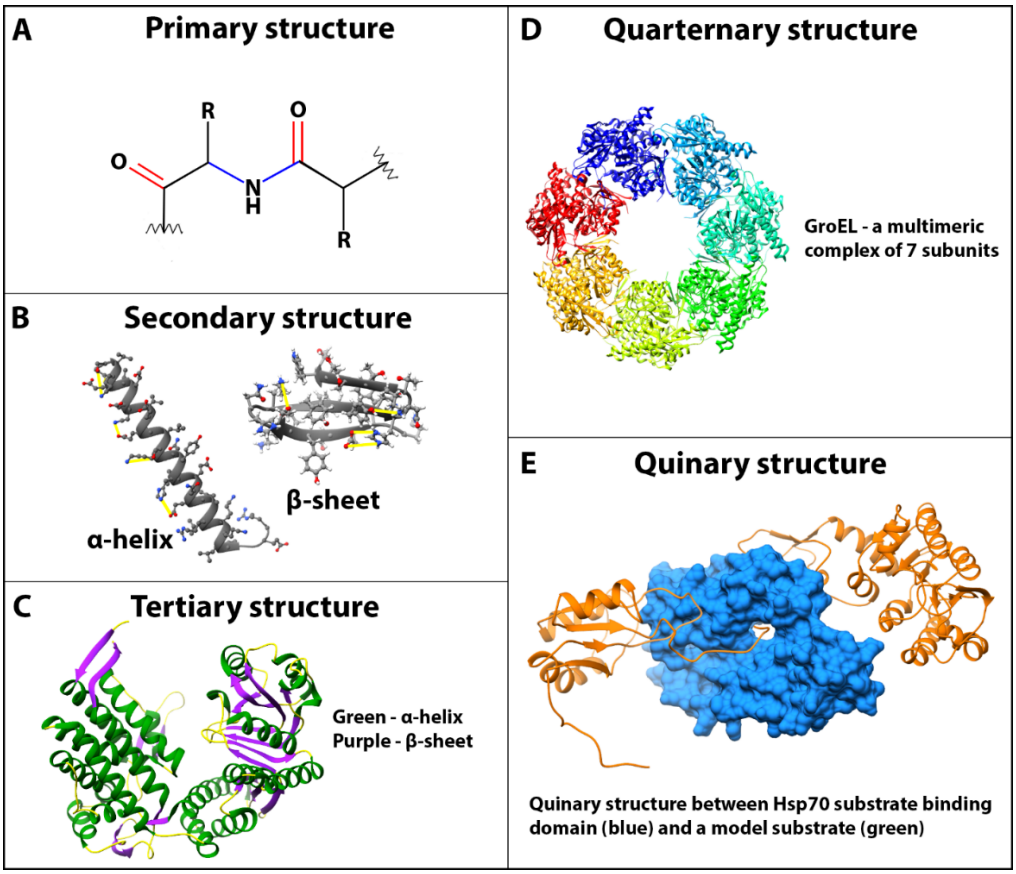


Figure 7: The different levels of organization of in protein folding. (A) Primary structure consists of backbone amide bonds. (B) Secondary structure shows two example folds, α-helices and β-sheets held together by hydrogen bonds shown in yellow. (C) Tertiary structure shows a fully folded GroEL monomer (PDB ID: 1SS8) consisting of both α-helices and β-sheets. (D) GroEL (PDB

ID: 1SS8) – a multimeric complex made of 7 monomers, an example of quaternary structure. (E) Quinary structure (weak binding) between Hsp70 substrate binding domain (blue, PDB ID: 2KHO) and a model substrate phosphoglycerate kinase (orange, PDB ID: 3PGK).

Quinary structure, due to its low stability, is highly susceptible to the environment. Cellular properties, such as whole proteome pI values, are highly organism-specific. For example, anionic proteins are more abundant in *E. coli* at physiological pH.[137–139] Since quinary structure is highly sensitive to the local environment, it is possible that they are also organism-specific. In-cell NMR studies have proved useful in probing the energy scale of weak interactions in cells.[140] However, many of these studies rely on the examination of eukaryotic proteins in prokaryotes to reduce peak broadening due to strong interactions that may occur in the native eukaryotic cytosol (Figure 5).[100] Due to its organism specificity, this strategy could disrupt quinary structure associated with the protein of interest that are present in the eukaryotic cytosol. For this reason, in-cell NMR studies of eukaryotic proteins in *E. coli* have been classified as sticking and described in Section 4.3.

It is important to differentiate between sticking and quinary structure since both have similar energetics. We also do not refer to strong, long-lived interactions, such as those underlying the hemoglobin tetramer or GroEL formation, as quinary, but rather as quaternary (Figure 7).

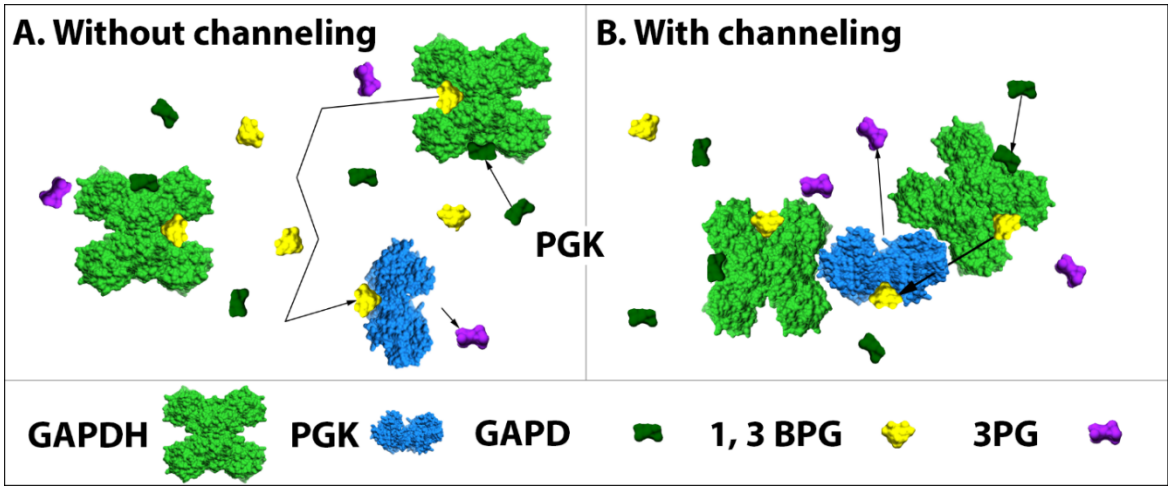


Figure 8: Cartoon showing substrate channeling between GAPDH and PGK during glycolysis. GAPDH (green) and PGK (light blue) were accessed by PDB IDs 1IHY and 3PGK. GAPD (dark green), 1,3 BPG (yellow) and 3PG (purple) were accessed by PubChem numbers 729, 683 and 439183. All structures were rendered using UCSF Chimera. (A) Without channeling the substrate 1,3 BPG must diffuse after being produced by GAPDH before being bound to PGK. (B) During channeling the substrate 1,3 BPG is channeled between both enzymes.

5.2 The metabolon

Perhaps one of the best representative examples of quinary structure formation is the metabolon. The word metabolon was coined by Paul A. Srere in 1985.[141] Quoting the 1985 communication by Srere, a metabolon is a 'supramolecular complex of sequential metabolic enzymes and cellular

structural elements'. This supramolecular complex is formed of many enzymes associating into quinary structure and increases reaction efficiency through substrate-channeling (Figure 8).[142–144] Metabolic pathways generate many intermediates, a majority of which have no specific function other than to be fed into the next reaction in the sequence. During substrate-channeling, substrates are prevented from escaping into the bulk cytoplasm by efficiently channeling them to the next processing enzyme in the supramolecular complex. This dramatically accelerates reaction rates by 1) avoiding the time delay for the enzyme or substrate to diffuse in the cytoplasm to encounter one another, and 2) allowing the enzyme to compete for relatively low copy number substrate molecules before they decay into side products.

Metabolons exist in many metabolic pathways, including fatty acid oxidation,[142] amino acid metabolism,[145] glycolysis,[146] lipid biosynthesis[142] and the tricarboxylic acid (TCA) cycle.[144,147] The TCA cycle, also known as the Krebs cycle, is an essential metabolic pathway that generates the energy rich molecule ATP through aerobic respiration in cells.[148] The Krebs metabolon has been studied in great detail and consists of eight enzymes forming a supramolecular complex via quinary interactions.[143,144,147,149] Theoretical modeling showed that on association the charge patters on the enzyme surfaces rearrange to create continuous positively charged zones.[143] This allows substrate-channeling of negatively charged substrates from one enzyme to another across the positively charged surface. An analytical equation was developed to characterize the effect of substrate-channeling on the kinetics of a bi-enzyme complex. Assume a simple coupled reaction scheme with enzyme E_1 and E_2:

$$S \xrightarrow{E_1,\ V_0} I \xrightarrow{E_2,\ K_m,\ V_{max}} P \qquad [9]$$

where, the substrate S is converted to intermediate I by enzyme 1 that operates at the constant velocity V_0 which is then converted to product P by enzyme 2 with Michaelis constant K_m and maximum attainable velocity V_{max}. For this bi-enzyme system following pseudo first-order kinetics the time required for the intermediate I to build up to sufficient levels so as to maintain steady state flux is called the transient time, τ. The variable τ can be used to characterize the effect of efficient channeling of substrate by the metabolon as substrate-channeling leads to dramatic decrease in the transient time. By building on older models by Easterby[150] and Ovádi,[151] Elcock and coworkers formulated, using an analytical approach, the following general equation for the calculation of τ:[152]

$$\tau = \frac{K_m(1-p_c p_r)}{V_{max}} \qquad [10]$$

where, p_c is the channeling probability that describes whether the intermediate I is successfully transferred to the next enzyme and p_r is the probability that complex formation successfully leads to formation of product P rather than dissociating to reform the intermediate. Using this formula, the transient time for two representative enzymes, citrate synthase (CS) and mitochondrial malate dehydrogenase (mMDH), in the Krebs metabolon was determined to be 0.03 secs and 2.5 secs with and without channeling, a nearly two orders of magnitude difference in efficiency.[143] In the above example the importance of substrate-channeling is evident in the ~100-fold reduction in τ.

This is particularly important in the case of CS/mMDH because the forward mMDH reaction has an unfavorable equilibrium constant and their intermediate, oxaloacetate, occurs at concentrations that are too low to sustain experimental reaction rates.[153,154]

Surprisingly, the quinary structures involved in substrate channeling have rarely been imaged directly in cells. One recently developed technique uses cell volume modulation in response to osmotic pressure to modulate enzyme association.[85] Using osmotic pressure Gruebele and coworkers were able to characterize the K_d of two weakly associating enzymes, $GAPDH_4$ and PGK, both members of the glycolysis metabolon. The K_d for $GAPDH_4$-PGK oligomerization was calculated to be ~14 μM in U-2 OS cells. Moreover, the authors showed that two GAPDH tetramers associated with one PGK. It is known that GAPDH copy numbers in-cells is at least twice as high as PGK.[155] Therefore, the formation of the $GAPDH_4$-PGK-$GAPDH_4$ ternary complex may help GAPDH compete for cellular levels of PGK and in the efficient transfer of bis-phosphoglycerate substrate from GAPDH tetramer to PGK during glycolysis (Figure 8). Other studies have also shown substrate channeling in the glycolysis metabolon. Molecular dynamics simulations by Barton, Minteer, Sigman and coworkers demonstrated substrate channeling between another enzyme hexokinase (HK) and G6PDH (glucose-6-phosphate dehydrogenase) in the glycolytic pathway.[156] The interaction between G6PDH and HK precede the ATP generation step involving PGK and GAPDH. Taken together, these results potentially emphasize the channeling of multiple substrates between enzymes during glycolysis.

The benefits of substrate channeling are clearly visible in many other biological processes. For example, during protein synthesis, multi-synthetase complexes provide amino acid substrates to the ribosome in the form of aminoacyl-tRNAs. These aminoacyl-tRNAs are channeled directly from aminoacyl-tRNA synthetases to elongation factor to the ribosome without diffusing into the bulk cytoplasm.[157,158] Such quinary structure has resisted characterization as it is disrupted by *in vitro* purification methods.[158] The many regulatory functions of the aminoacyl-tRNA synthetase complexes have been discussed in detail in published reviews.[158]

Shakhnovich and coworkers showed that there is evolutionary pressure to develop interaction networks that support substrate channeling and select against non-functional interactions.[159] The authors investigated in detail the reasons behind the gene dosage toxicity (GDT). In simple terms GDT is a phenomenon where overexpression of certain genes has toxic effects. Using the enzyme DHFR (dihydrofolate reductase) the authors showed that imbalance of protein-protein interactions due to overexpression is the result of toxicity during DHFR overexpression. One of the key findings revealed by these results was that evolutionary selection shapes protein-protein interactions to facilitate the formation of metabolons that support efficient substrate channeling. These interaction networks are highly organism specific such that when *E. coli* DHFR was replaced by a foreign DHFR, promiscuous mis-interactions increased indicating that there is selection pressure against such promiscuity.

Quinary structure is thus an important biological organizing principle ensuring that reactions in the cell proceed efficiently. Weak interactions allow supramolecular complexes to be assembled

and disassembled quickly in response to cellular signals and act as effective biological switches. With the advent of better *in vivo* imaging techniques, the observation of metabolon formation in many other biological compartments may surface in the future.

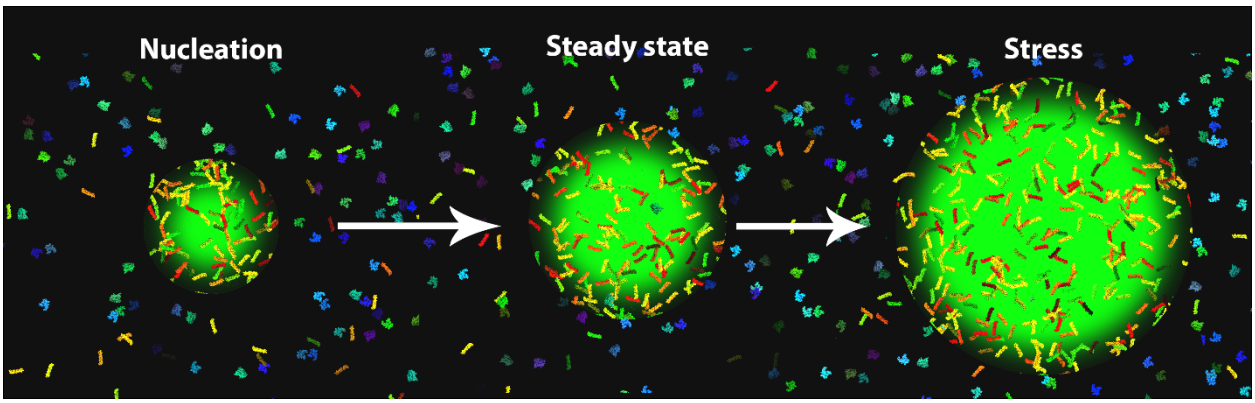


Figure 9: Cartoon representation showing the three steps in the formation of phase separated droplets in the cell. During nucleation a set of constituent proteins are recruited and form a small droplet. The droplet grows and reaches steady state. However, if stress is applied the droplet growth increases as a stress response mechanism to protect or shield its constituents from unfolding/misfolding.

5.3 The role of quinary structure in cellular organization

A single quinary interaction may not confer a lot of additional functionality or stability to the cell, but in large numbers they can add up and result in robust networks of interactions. Weaker interactions allow a protein to interact with more partners by using fewer bits of the recognizable information stored on a protein's surface. Such weakly interacting networks are more likely to be robust against deletion of any one component or interaction.[2]

In extreme cases, the highly crowded cellular environment can even lead to phase separation of biomolecules due to quinary interactions, similar to that observed in saturated solutions (Figure 9). This creates microenvironments, also referred to as membraneless organelles or 'liquid droplets,' with defined functionality and specific composition: examples include Cajal bodies, stress granules, nucleolus, P-bodies and paraspeckles.[160] The composition of these membraneless organelles typically ranges from a few to several hundred protein species or RNA molecules.[160]

These microenvironments are considered quinary structure due to two main reasons: 1) The interacting molecules are highly dynamic, showing liquid-like properties where molecules dynamically exchange with the surrounding environment and many components are recruited only transiently in response to certain stimuli and 2) they are functional and improve cell health.[161–163] At the center of this phenomenon are multivalent interactions between groups of binding partners often involving highly charged disordered proteins (IDPs) and RNA molecules.[164,165] Many of these quinary interactions occur in the low-complexity regions (LCRs) in IDPs that consequently have been shown to be modulators of phase separation in cells.[166] Quinary

interactions has also been implicated during stress-induced phase separation[166,167] and reversible aggregation of endogenous proteins.[168]

The biochemical environment in these phase-separated regions has important functional implications. Phase separation can affect reaction kinetics and specificity by substantially increasing the local concentration of the reactants. For example the rate of mRNA processing is significantly reduced when key components fail to concentrate within the histone locus body or Cajal bodies in zebrafish.[169–171] Phase separation could also inhibit activity by sequestering molecules in these regions or act as an on/off switch where functions can be rapidly regulated through the formation and dissolution of the condensed phase.[160] For example key components of the protein synthesis machinery are sequestered inside stress granules during heat shock in *Saccharomyces cerevisiae*.[167] Finally, by either releasing molecules from the condensate or recruiting molecules into the condensate, phase separation can help to maintain stable levels of molecules in the bulk phase despite fluctuations in expression.[160]

The eukaryotic cytoskeleton also exhibits many instances of quinary structure formation with K_d in the ~μM regime. In yeast, *Saccharomyces cerevisiae,* evidence suggests that actin and microtubule cytoskeletons may function together during mitosis and in mating cells.[172] Barnes and coworkers found that a component of the actin cytoskeleton, coronin 1p (Crn1p), provides a functional link between the actin and microtubular cytoskeleton in yeast.[173] Crn1p interacts only weakly with microtubules with a K_d of 15-20 μM. However, in the presence of actin, the K_d for Crn1p and microtubule association increases 10-fold. This indicates that quinary structure formation (K_d – 2 μM) via Crn1p crosslinks actin filaments and microtubules in mitotic and mating cells. Another example of quinary structure is evident for the cofactor, dynactin, of the cytoplasmic dynein-1 motor that transports cargos along the microtubular cytoskeleton.[174] The microtubule binding domain of a dynactin subunit interacts with microtubules with a K_d of 10 μM.[175] These cytoskeletal quinary structures abound in the eukaryotic cytoplasm and provide structure and shape to the cell.

5.4 Quinary structure and encounter complexes

Phase separation is not the only phenomenon in biological systems that is driven by quinary structure. Biological systems need regulatory switches for a large number of processes. These switches must be highly sensitive and specific and signaling must be performed with a high fidelity to ensure proper function. Most signaling pathways consists of the formation and dissolution of multicomponent complexes consisting of proteins, nucleic acids and other small molecules.[176] These complex must not only associate with high specificity but also dissociate when the signal is turned off. A tightly bound complex that does not dissociate cannot be turned off.

The cell solves this problem using encounter complexes (Figure 10).[177] In simple terms, biomolecules diffuse freely and form weak transient encounter complexes via collisions with a high k_{on} and k_{off} rate. These weak complexes can then reorient, reposition or undergo conformational changes to give rise to stronger and more specific interactions with a low K_d. These weak transient complexes are functional and signaling pathways that lead to their formation are possibly formed through evolutionary selection under which non-functional interactions evolve

into productive interactions. In addition to evolved functionality, the K_d for these interactions are in the range of 10s of μM[178] and hence we group them under quinary structure formation for the purpose of this review. In effect, quinary structure can be a short-lived precursor for stronger binding.

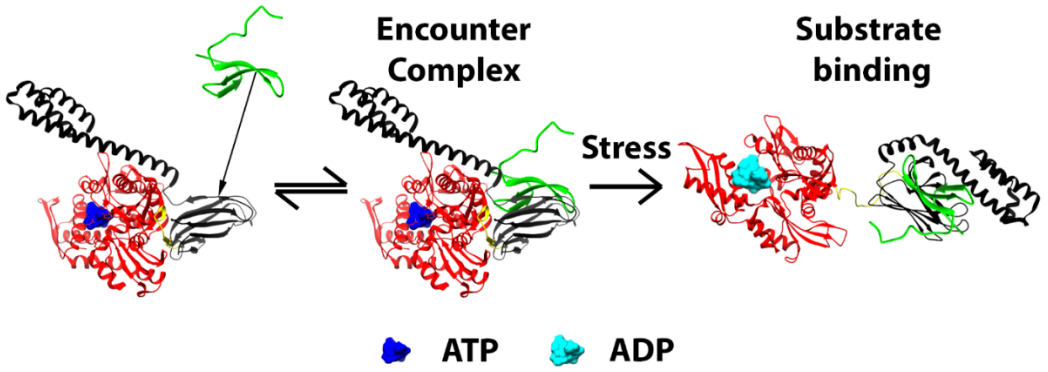


Figure 10: Formation of encounter complexes leading to binding in the Hsp70 system. Red and black show the N-terminal and C-terminal nucleotide and substrate binding domain respectively. In the ATP bound state Hsp70 forms transient encounter complexes with potential substrate proteins. During stress ATP is replaced by ADP which facilitates conversion of the encounter complexes to productive stably bound substrate-chaperone complex. Hsp70 molecules were rendered using UCSF Chimera and accessed using PDB IDs (ATP bound state – 4B9Q and ADP bound state – 2KHO).

Such a mechanism is utilized by the molecular chaperone 70 kDa heat shock protein (Hsp70). Hsp70 is maintains cellular proteostasis by binding unfolded, misfolded or nascent chain peptides and preventing further unfolding (Figure 10).[179,180] In a normal cell, Hsp70 is in the ATP bound state where it binds substrates with a high k_{on} and k_{off} rate.[181] Substrate binding then induces ATP hydrolysis which in turn leads to a conformational change in the C-terminal domain of Hsp70. This conformational change results in a more tightly bound Hsp70-substrate complex.

Paramagnetic relaxation enhancement (PRE) is a useful method to detect low-population (<10%) encounter complexes. To do this, paramagnetic labels are generally introduced outside the specific interaction site. Clore and coworkers showed that the encounter complexes formed due to quinary interactions are important intermediates that increase the rate of formation of specific functional interactions by reducing the dimensionality of the search process.[178,182] In the bacterial signaling system that catalyzes phosphorylation reactions (phosphotransferase system) the authors were able to show evidence of the formation of these encounter complexes using PRE. They observed these rare, fast-exchanging complexes for the N-terminal domain of enzyme I (EIN), $IIA^{Mannitol}$ and $IIA^{Mannose}$ with the phosphocarrier protein (HPr) with $K_d \approx$ 10-50 μM.[178] These encounter complexes are formed due to longer-range electrostatic and shorter-range van der Waals attraction and have been discussed in detail in published reviews.[183,184] In the case of the bacterial phosphotransferase system, the negatively charged residues on EIN, $IIA^{Mannitol}$ and $IIA^{Mannose}$ interact with the positively charged surfaces of HPr. In the case of EIN-HPr, a small

population of a ternary encounter complex (HPr-EIN-HPr) was also observed.[185] This second type of encounter complex occurs predominantly when the active site of Enzyme I is occupied, and possibly helps in efficiently reloading the enzyme active site the moment it frees up, as well as competing for the cellular pool of HPr.

A similar ternary complex was also observed for protein-DNA binding.[182] In the nucleus, DNA is present at mM base pair concentration. Under nucleus-like experimental conditions involving sub-millimolar free DNA, Clore and Iwahara showed that the association of protein and DNA occurs via a ternary encounter complex where free DNA associates with a DNA-protein bound complex rather than the association of free protein with free DNA. The formation of the ternary complex can accelerate target recognition rate in protein-DNA interactions resulting in translocation rates that are up to three orders of magnitude faster than the *in vitro* protein-DNA dissociation rate for a transcription factor, HOXD9 homeodomain. This phenomenon simultaneously explains the highly dynamic nature of protein-DNA interactions observed *in vivo*, as well as the long half-life of the complex measured by traditional *in vitro* biochemical methods.[186,187]

Therefore, quinary interactions underlying encounter complex formation plays important biological roles in living systems. The two roles discussed in detail in this section relate to the efficient formation of a specific enzyme complex and the efficient reloading of the substrate at the active site of this complex. Both these functions are important for improving enzymatic turnover *in vivo*, however, this scratches only the surface of quinary structure formation and its biological implications. There is also evidence that formation of these weak complexes is involved in enhancing electron transfer for plastocyanin[188] and cytochrome c.[189,190] Future studies are required to fully appreciate the functional diversity and landscape of these short-lived functional encounter complexes.

5.5 Characterizing quinary structure formation *in vivo* and *in vitro*

Because quinary structure is transient and weakly bound, on the order of a few k_BT in the energy scale, routine purification methods can disrupt it and therefore, it must be characterized *in vivo* where possible.

This is not to say that quinary structure formation cannot be observed *in vitro*. For example, using NMR Shekhtman and coworkers showed that ribosome-protein quinary structure plays a key role in enzymatic activity of thymidylate synthase (TS).[191] Addition of ribosomes *in vitro* enhances enzymatic activity 20-fold, in good agreement with the ~10-fold increase observed *in vivo*.[192] Interestingly, the authors also found that ribosome-specific quinary interactions can decrease the activity of another enzyme, dihydrofolate reductase (DHFR). By modulating the enzyme activity via quinary structure formation, the ribosome plays an important role in metabolism by acting as the hub where enzymes and metabolites are concentrated.[142] Ribsome-mediated quinary structure formation with the 30S subunit of the ribosome has also been shown with mRNA, and for aminoacyl-tRNA synthetases such as with LysRS as described in section 5.2.[193] These interactions are significantly weakened when antibiotics that inhibit the 30S subunit are added.

In-cell NMR was originally established in bacteria, but eukaryotic systems have also recently been used, including yeast, human cell lines and *Xenopus laevis* oocytes.[100,128,132,194,195] NMR experiments have the potential to probe both sticking (see Section 4.3) and quinary structure which we describe in detail in this section. NMR experiments in yeast have been successfully used to demonstrate biomolecule interactions under physiological expression conditions in physiologically relevant cellular compartments. This is possible in yeast due to a wide variety of genetic tools facilitated by very well understood yeast genetics.[195] Moreover, a very stable yeast cell wall reduces leakage problems[196] and spectra can be obtained over long periods of time (~6 hours) with a high signal-to-noise ratio.

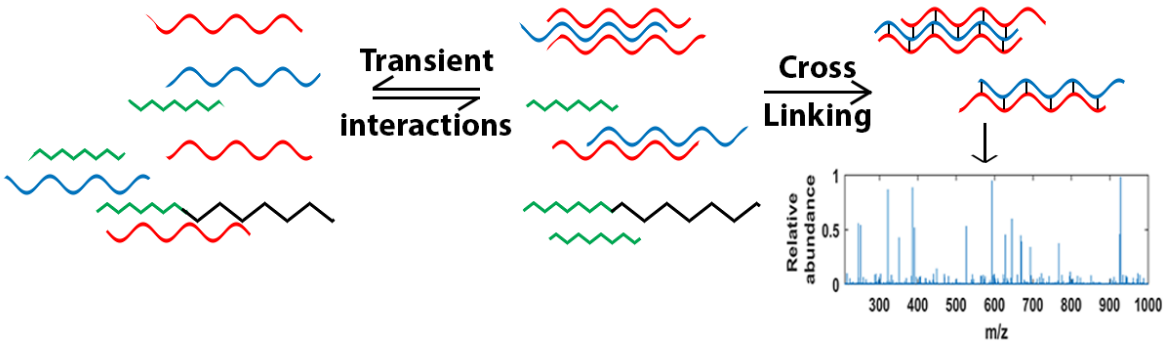


Figure 11: Chemical cross-linking mass spectrometry. Transiently interaction species are cross-linked to form more stable species that can be then isolated in vitro. These fragments can then be characterized using mass spectrometry giving rise to discrete peaks corresponding to complex molecular weight on the x-axis and relative abundance on the y-axis.

In addition to above-mentioned enzymes and mRNA, chemical cross linking and mass spectrometry revealed >800 proteins that can potentially bind to mRNA or ribosomes in eukaryotic cells (Figure 11).[191,197–199] Experiments in yeast using NMR by Shekhtman and coworkers showed that transient interactions with RNA play a key role in deciding biochemistry, like protein activity and localization, of ubiquitin and β-galcatosidase.[195,199] The resulting quinary structure is further modulated by growth medium and RNA levels in cells. For example, in yeast grown in methanol as the carbon source a well-resolved $^{1}H–^{15}N$ HSQC spectrum of ubiquitin is obtained. This is because the presence of a large amount of preprocessed mRNA and large ribosomal subunit leads to a reduction in RNA-protein quinary structure formation. In sharp contrast, quinary structure formation is enhanced, and the ubiquitin spectrum is broadened beyond detection in yeast grown in mixed dextrose-methanol medium. Moreover, in the mixed medium both proteins are sequestered in their inactive state in vesicles. Ubiquitin interferes with dextrose metabolism and its inactivation can allow the cell to explore more efficient metabolic pathways,[200] also known as catabolic inactivation.[201] RNA-protein quinary interactions can therefore act as important regulators of many in-cell processes.

Even with some disadvantages (e.g. high protein density required),[196] NMR in human cell lines[100,132] and *Xenopus laevis* oocytes[202,203] is a powerful techniques to probe quinary structure. Proteins are constantly interacting with other proteins in the cellular cytoplasm. These interactions

lead to peak broadening and can be easily visualized by NMR. Formation of quinary structure has been observed for both ubiquitin in HeLa cells and *Xenopus laevis* oocytes and profilin1 in human HEK293T cells.[100,132,203] The contribution of these interactions can be measured by introducing appropriate mutations and comparing their NMR spectra. For example, mutations in parts of the protein that interact strongly with other protein partners will lead to larger differences in the spectrum than those that interact weakly. This was successfully demonstrated for both ubiquitin and profilin1 by introducing mutations in the binding pocket and on the surface respectively, which then yielded well-resolved NMR spectra.[100,132]

Because the stabilization due to quinary structure formation is on the order of a few k_BT, they are highly susceptible to changes in the local environment. Gruebele and coworkers showed that different environments in the cell show different degrees of stabilization due to quinary structure formation.[121] A model protein phosphoglycerate kinase (PGK) is stabilized in mammalian U2OS cells compared to *in vitro*.[122] This is partially due to macromolecular crowding[18], however PGK stability in-cell is also modulated to some extent by its cellular localization.[121] PGK in the nucleus is more stable than that in the cytoplasm and endoplasmic reticulum (ER). The degree of crowding in the cytoplasm and the nucleus is similar according to diffusion measurements.[80] Since the local environment in the nucleus is very different from that in the cytoplasm, the nature of sticking or quinary interactions are possibly also different in both environments.[204]

Weak modulations of the cellular environment such as cell volume changes can be used to characterize the K_d for quinary structure formation. Such weak perturbations only modify quinary structure in the cell without destroying the cell, thus, allowing the detection of quinary structure in the native cellular environment. Cell volume changes by osmotic pressure modulation was used by Gruebele and coworkers to determine the K_d for $GAPDH_4$-PGK binding, described in detail in section 5.2 and for mCherry-AcGFP1 oligomerization to be ~2 μM, an order of magnitude lower than *in vitro*.[85] PRE-measurements described in detail in section 5.4 have also proved very useful to probe quinary structure and the formation of metastable encounter complexes.

6. Evolution

We now turn to the connection between physico-chemical interactions in the cell, such as crowding and productive (quinary) or disruptive (sticking) interactions, and evolution. Evolution of proteins has been investigated in great depth and many comprehensive reviews exist that critically analyze the developments in the field.

A recent article by Spitzer, Pielak and Poolman on the emergence of life sheds light on some interesting concepts that drive biological evolution.[205] The authors describe how, among other things, crowding plays an important role in driving evolution in biological systems. Evolution cannot occur in uncrowded systems where surfaces are far from each other and do not interact. In dilute conditions, non-covalent molecular forces cannot maintain cellular organization because thermal disordering effects overcome attractive ordering forces. Indeed evolution of biomolecular

surfaces under crowded or confined conditions has been shown to be a likely prerequisite for the transition from the inanimate to the living.[137]

Since the focus of this review is cellular forces at protein surfaces, we briefly touch upon how evolution has shaped surface interactions. Biomolecular surfaces are constantly evolving under constraints to improve fitness and function. The potential benefits of combining protein biophysical chemistry with evolutionary biology are many, and as such the inclusion of evolutionary biology in the study of protein biophysics already greatly benefits our understanding of protein function.[206,207] We discuss two main methods of analyzing evolutionary data: epistasis mapping and ancestral sequence reconstruction. Finally, we conclude the review by outlining what future collaborative efforts between evolutionary biology and biophysical chemistry could bring to the table in terms of our understanding of protein interactions in the cell.

6.1 How protein surfaces evolve – cytochrome *c* and heat shock proteins

Evolution means constant change in biology, but some proteins are peculiar because they have stayed effectively unchanged over millions of years. The structure of cytochrome *c* was solved in 1971 and was used to characterize the correlation between evolution and function.[208,209] Cytochrome *c* is a small protein (104 residues in vertebrates), present in the mitochondria of every eukaryotic organism. Dickerson estimated its evolutionary rate.[209] During random mutational drift, if m is the fraction of mutations in a polypeptide chain, then the fraction n of the polypeptide sequence that actually changes is:

$$n = (1 - e^{-m}). \qquad [11]$$

n takes into account that repeated mutations at the same locus do not reduce sequence identity. The calculated rate was then used to approximate the time required for a 1% change in sequence between two divergent lines of evolution in a million years (MY). For a small fibrinopeptide, whose function is to be excised out of fibrinogen and converted to fibrin in a blood clot, this time was 1.1 MY, a rapid rate of protein evolution. Presumably for fibrinogen, any change in the sequence is permissible if it still allows for successful excision of the peptide. The fibrinogen mutation rate is therefore close to the actual DNA mutation rate (Figure 12). Conversely, cytochrome *c* interacts with several large macromolecules and hence utilizes more of its overall surface for function. The time scale for cytochrome *c* is 20 MY and large portions of the cytochrome *c* surface have highly conserved charged and aromatic residues. There is also a general correlation between protein size and evolutionary rate.[209]

The evolutionary role and importance of the family of heat shock proteins has been discussed and probed in great detail.[210–212] They have many interaction partners and are evolutionary capacitors.[213] Not surprisingly, therefore, they are also highly conserved and well-known for their stress response function.[214,215] Among them, the 90 kDa heat shock protein (Hsp90) plays a fundamental role in the expression of genetic variation.[216–218] Its function is special because in addition to stress response, it can also expose or suppress genetic variation. Macro-scale changes in genetics through evolution are generally gradual, but all living species must survive

sudden environmental changes and maintain robust developmental systems that do not change abruptly during environmental stasis. This is generally done by storing a certain amount of unexpressed genetic variation that is typically hidden during developmental homeostasis, for example, in species hybrids.[219] Hsp90 can expose such variation by destabilizing and stabilizing transcription factors associated with promoter regions responsible for morphological remodeling of metamorphosis.[218] Likewise, in cell-cycle control Hsp90 supports both activators and inhibitors of the same function to control the process output downstream.[220] Hsp90, therefore acts as a capacitor, that through controlled exposure of cryptic alleles could account for the rapid morphological changes evident in fossil records. Another 70 kDa heat shock protein (Hsp70) that is also highly conserved maintains cellular stasis during stress and has been referred to as an evolution facilitator, for example by protecting less stable protein mutants and facilitating more sequence variation.[214,221]

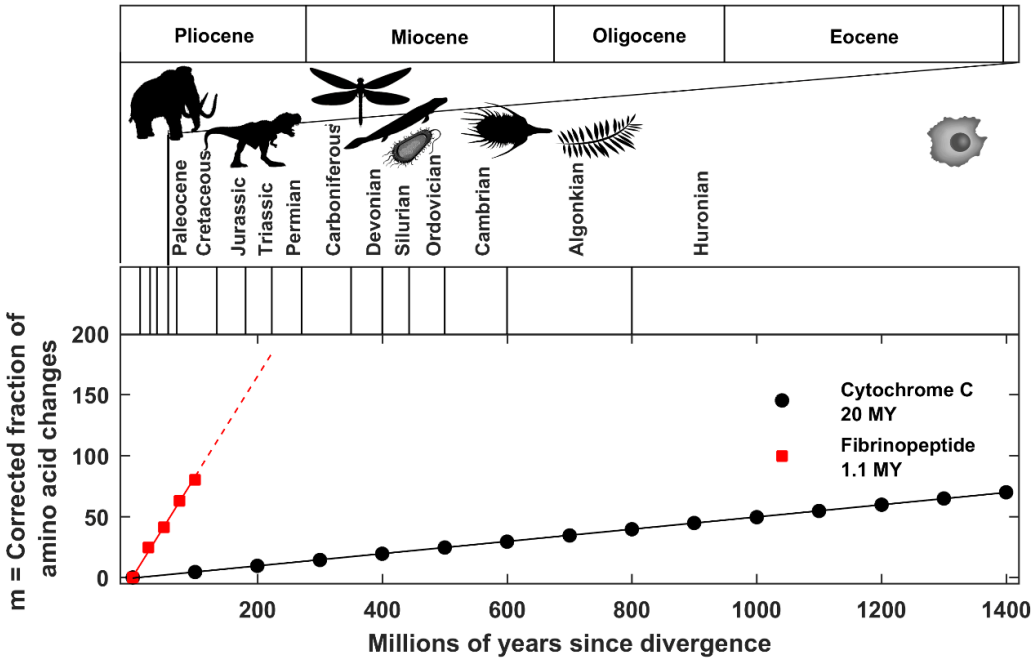


Figure 12: The evolution of Cytochrome *c* *vs.* fibrinopeptides. Data for this figure is adapted from reference 209. Cytochrome *c* evolves very slowly taking almost 20 MY for a 1% change in residues whereas fibrinopeptides evolve at the rate of occurrence of mutations, 1.1 MY. This shows that residues in Cytochrome *c* are highly conserved.

These two examples of protein evolution teach us something very important about the combination of evolutionary biology and biophysical chemistry. Cytochrome *c* shows us how important it is to consider evolution from the perspective of protein function. It also points to conclusions that could be made only by combining results derived separately from evolution and studies of protein function. Heat shock proteins highlight that evolutionary changes occur both on macro- and micro-scales. Interestingly, biological systems have evolved not only fitness with the help of genetic variation, but have also endowed the products of evolution itself, proteins, with the

latent power of evolutionary change. In the future heat shock proteins are prime candidates to tell us a more complete story of evolutionary change than what is visible only through genetic data.

Lastly, although evolution is the force for change, it is important to probe systems, such as heat shock proteins, that have resisted that change. By correlating systems that change frequently and those that remain static, we can begin to reconstruct the elusive protein evolutionary pathways that lead from now into the past.

6.2 What drives evolution of protein surfaces?

The cell is an environment where proteins must navigate crowded spaces so their surfaces can not only make functional interactions, but also avoid debilitating non-functional interactions with the majority of surfaces they encounter. As discussed in 5.3, only so many bits of information can be encoded on a protein's surface, and must be used up in a compromise between interaction strength (stronger = generally more surface), number of interactions encoded (more = less surface per interaction), and avoiding undesirable sticking (although this could evolve into quinary structure). Proteins evolve constantly and have been doing so since the Hadean eon (~4 billion years ago).[222,223] They have weathered extreme heat and extreme cold but have maintained function.[224] Interestingly, the principles that drive the complex evolutionary behavior in biological systems are fairly basic and follows two main rules: 1) maintain or improve organism fitness and 2) keep intact or improve function.

Perhaps the two driving forces in protein evolution that are most evident at first glance are folding and stability. Since misfolding and aggregation can have a considerable negative effect on organism viability,[225] there is selection pressure to evolve thermodynamically stable[226] and /or aggregation-resistant protein sequences.[227] Several different adaptations reduce aggregation propensity for proteins.[227] For example, protein sequences limit hydrophobicity and maintain an overall low net charge. Hydrophobic patches longer than 5 residues are represented significantly less than what would be predicted from a statistically independent distribution.[228] Moreover, disordered proteins have charge distributions to avoid aggregation, and globular protein fold to bury hydrophobic residues in the core and hence reduce the propensity of these hydrophobic surfaces from coming into contact and aggregating with other hydrophobic surfaces. Since aggregation is concentration-dependent, protein sequences have further evolved such that the aggregation propensity is just below their solubility limit *in vivo*.[229]

Early studies have also shown that different proteins from the same organism can evolve at vastly different rates.[230] Proteins with multiple partners evolve more slowly.[2,231] This is because a greater proportion of the protein is directly involved in function and the protein is only able to accept a reduced subset of mutations that are at either net neutral or positive for all the interacting partners. Any other evolutionary changes, for such a system must occur via coevolution where changes to one protein lead to an additional selection pressure for reciprocal changes in the interacting partners. This is in part explained by the neutral theory of evolution,[232] where the rate of protein sequence evolution is given by,

$$k = \mu p, \quad [12]$$

where k is the rate of protein sequence evolution, μ is the rate of mutation and p is approximated by the proportion of mutants that are neutral because beneficial mutations are considered too rare to affect the rate of evolution.

Another factor driving protein evolution is the pressure to reduce non-specific protein-protein interactions that compete with productive specific interactions. Interestingly, even though the cell is a sea of sticky surfaces where the need for specificity is paramount, protein interactions follow a scale-free network topology. In a scale-free topology, the minimum free energy gap (ΔE) between the weakest specific interaction and the most competitive non-specific interaction decreases in a power law fashion,[233,234] such that

$$\Delta E \sim N^{-\gamma} \quad [13]$$

where, N is the number of interfaces and γ is the scaling factor that determines how quickly the energy gap drops to 0 as the number of interfaces increase and correspondingly how quickly binding specificity is lost. Even though γ is small (0.13-0.19) the gap reduction is significant for typical proteome sizes, such that, for a typical yeast proteome a gap as small as ~2.5k_BT is already reached with ~1000 interfaces. Moreover, for a proteome with N protein types there are aN specific interactions and bN^2 non-specific ones (where a = 1 to 20 and b are constants), and hence there is significant evolutionary pressure to decrease deleterious non-specific interactions. Since increasing the proteome size radically increases non-specific interactions and decreases ΔE, this limits the number of proteins that can function effectively in the cell, and organisms must therefore survive with a limited number of proteins. Increasing protein size provides one way out by increasing the dynamic range of interactions (weak to strong) available but has its own problems: increased misfolding and aggregation requires increased chaperoning, i.e. yet more proteins that act as caretakers.

Indeed, even though the evolutionary distance is large, the number of proteins remains similar between simple multicellular organisms and humans.[235] Given these constraints and consistent with scale-free topologies, binding in biological systems is optimized by favoring networks where a few proteins interact with a large number of partners, while most proteins interact with just a few other partners.[233,236]

The pressure to reduce non-specific interactions has also been associated with protein abundance and surface hydrophobicity or hydrophilicity.[237] Protein abundance is negatively correlated with the number of its functional interaction partners as well as the number of non-specific interactions.[238] Functional interactions in a proteome generally involve hydrophobic interactions.[239,240] Consequently, hydrophobicity also decreases with protein abundance to decrease non-specific interactions. In agreement with this observation more abundant proteins in the *E. coli* cytoplasm are less hydrophobic.[241] Conversely, abundant hydrophilic IDPs (disordered proteins) make up to 30% of the eukaryotic proteome.[242] Thus, surface properties of a protein play a crucial role in determining protein evolutionary rates.[238,243]

Further reviews that discuss protein evolution in more detail have been published.[244–246] We discuss the biophysical aspects of protein interaction networks in more detail in the next section.

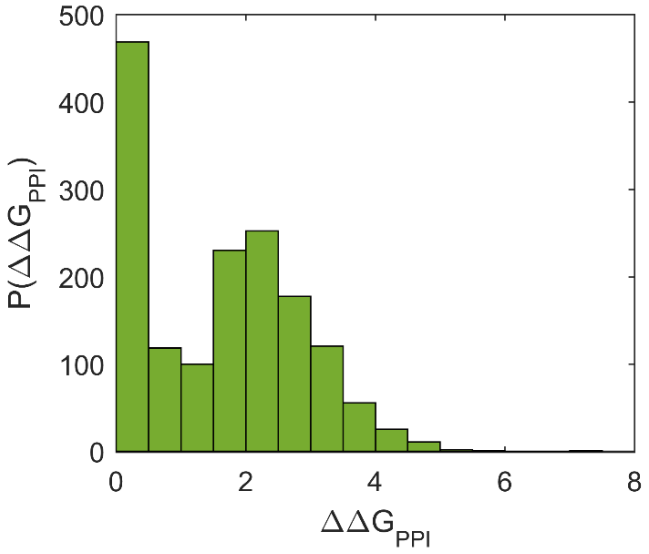


Figure 13: Bar plot showing relative abundance of proteins ($P(\Delta\Delta G_{PPI})$) as a function of stabilization due to PPIs ($\Delta\Delta G_{PPI}$). The data from this figure was taken from reference 213. The population at $\Delta\Delta G_{PPI}=0$ shows no stabilization due to PPIs.

6.3 Evolutionary advantages of protein interaction networks

The hydrophobic core of proteins is highly conserved and mutations in the core can quickly disrupt protein structure and stability.[247] Although surface mutations contribute only weakly to overall stability,[248] they also exhibit a surprising level of conservation.[249] This is due to protein-protein interactions (PPIs). In the previous section, we discussed how protein interaction networks follow scale-free topologies. These protein interaction networks (PINs) benefit the organism in many ways.

In spite of the obvious disadvantages due to non-specific interactions, highlighted in section 6.2, networks are of great importance to thriving biological systems. They facilitate signaling and ensure that the cell is robust to random failure of a few network components. Additionally, in higher organisms the number of interactions between proteins increases, and consequently larger multi-protein complexes are observed as compared to their ancestors.[250] Such multi-protein complexes are favored over larger size of individual proteins because larger proteins are more expensive to fold, as noted in 6.2.[251] Smaller individual proteins are easier to fold and are less prone to aggregation, thus improving overall fitness of the organism. Protein interactions can also give rise to allostery and cooperativity, which results in a more efficient on/off switch.

PPIs further reduce toxic aggregation by reducing the effective monomer concentration in the cytosol; monomeric proteins bound to partners are not available for aggregation and are therefore

removed from the aggregation-prone pool of monomers. This additional stabilization of the folded state due to suppression of aggregation was calculated[213] and contributes to the overall stability such that,

$$\Delta G = \Delta G_{folded} + \Delta\Delta G_{PPI} \quad [14]$$

$$\Delta\Delta G_{PPI} = k_B T ln\left(\frac{C_A}{(1+e^{\Delta G/k_B T})U_A}\right), \quad [15]$$

where C_A is the total concentration of any protein, U_A is the concentration of all unusable states of A, comprised of all unfolded peptides and insoluble oligomers and ΔG is the thermodynamic stability of the free monomeric state. In yeast, the stabilization due to PPIs was found to be on average about ~2k_BT and in some cases as high as 5-6k_BT (Figure 13).[213] Therefore, each protein in a PIN effectively stabilizes its interaction partners to some extent. On the other hand, 'foreign' proteins such as noted in the in-cell NMR experiments (section 4.3) experience sticking instead of effective stabilization.

Stabilization from PPIs can also offset small amounts of destabilizing variations in protein sequences, which often leads to a higher degree of functional diversification.[252,253] This phenomenon has been termed evolutionary capacitance.[213] We discuss the concept of capacitance further in section 6.5. Indeed, the contribution of $\Delta\Delta G_{PPI}$ becomes more important for proteins whose effective population is low (due to genetic drift) and for proteins with a low inherent stability (ΔG_{folded}), such as IDPs (disordered proteins).

The relationship between protein fitness and evolution has been investigated in detail. The Shakhnovich group studied the effect of mutations on the folding free energy and showed that protein abundance is negatively correlated with evolutionary rate using simulations.[254,255] Since destabilizing effects of deleterious mutations are multiplied by protein abundance, more abundant proteins lead to larger amounts of toxic misfolded structures as a result of lethal mutations. Protein abundance and fitness hence plan an important role in determining evolutionary rates: higher abundance leads to higher stabilities and slower evolutionary rates since average mutations are more deleterious. These relationships between the protein biophysical landscape and evolutionary rate have been discussed in detail in already published reviews.[256,257]

6.4 Current methods in protein evolutionary biology

In this section we briefly touch upon two methods that have been used to study evolution of biological systems, and that are related to protein evolution and cell health and can be applied to studies of protein surface evolution. Detailed description of the methods discussed below are beyond the scope of this review but have been discussed elsewhere.[258,259]

6.4.1 Epistasis modeling

The term 'epistatic' was first used in biology by Bateson in 1909 to describe masking of a gene by another gene.[260] From the point of view of proteins, epistasis refers to the modification of a

mutation's phenotype by another mutation.[261,262] For example, a deleterious mutation could be masked by a mutation at a remote site; or a mutation's enhanced activity could only arise in the context of a subsequent mutation at a remote site, so that what started out as genetic drift later becomes enhanced fitness (Figure 14). As a result, any mutation event in an evolutionary pathway is contingent on the context of many past events. Epistasis is responsible for evolutionary benefits that occur due to a potentiating mutation in an ancestor, and thus cannot be evolved easily by stochastic single point mutations.[263] Thus, epistasis is a symmetry-breaking event that gets frozen in as time progresses. The genetic definition used by Bateson is the same basic concept but defines these changes in the genotype of the organism instead of the phenotype.

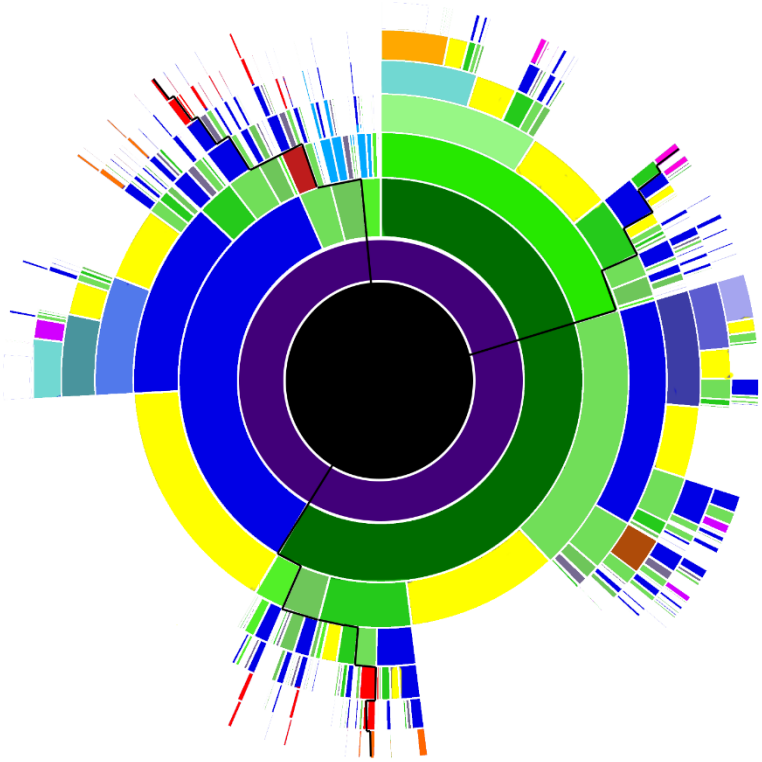


Figure 14: A sunburst plot showing epistasis. All proteins start from the black ancestor. Each corner represents a possible mutation and color differences show mutation accessibility from the previous mutation blue (easiest) to red (hardest). For example, it's easiest to go from blue to blue and hardest to go from blue to red. These affects add up, for example if starting at blue traveling along blue is the easiest. The three example black paths show three different evolutionary trajectories starting from the common black ancestor. Epistasis means that every mutation is to some extent determined by the path taken by the previous mutations and the landscape color of the previous mutation.

The recognition of pairwise epistasis goes back ~100 years,[260] however, the importance of higher-order epistasis has only recently been emphasized.[264–266] Pairwise epistasis is the difference in the effects of two mutations introduced together *vs.* separately. More generally higher-order epistasis refers to when a set of $n + 1$ mutations are introduced together *vs.* separately in *n-tuples*

(e.g. a triplet relative to the three pairs). Such higher-order epistasis is evident in many biological systems.[266]

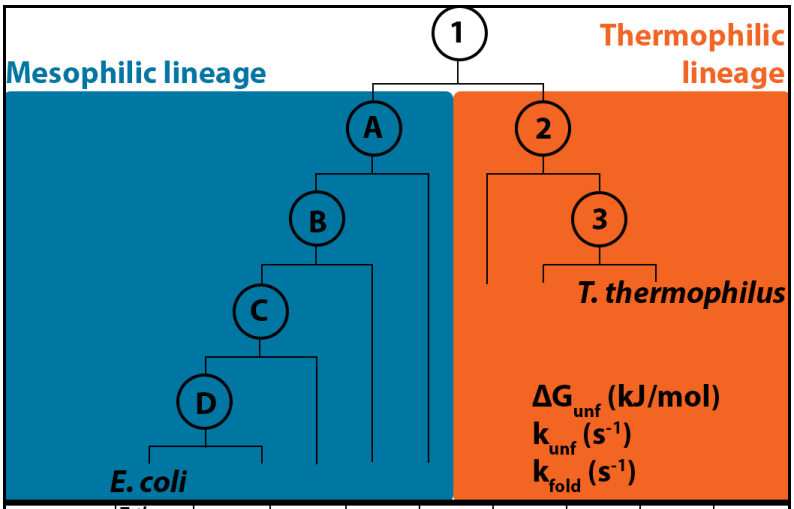
Historically epistasis has been studied either in simple model organism such as *E. coli, S. cerevisiae* or *C. elegans*, or by computationally modeling epistasis in model systems.[258] For example, Plotkin and coworkers modeled how epistasis affects accumulation of mutations over time.[262] They studied the stability of lysine-arginine-ornithine-binding periplasmic protein (*argT*) from *Salmonella typhimurium* as a function of mutations. They showed that at any given time, mutations that became successfully incorporated were contingent on previous mutations and typically would have been deleterious if introduced at an earlier time. Additionally, once a mutation is fixed, any reversal becomes increasingly detrimental to fitness. This is because a mutation, M, fixed at any given time interacts with many other mutations that occur at later times and deletion of M negatively affects all other mutations that were contingent on M. This phenomenon has been referred to as an "evolutionary Stokes shift,"[267] in analogy to light excitation and fluorescence: once an excited state is relaxing, blue fluorescence at the same wavelength as the absorption is increasingly unlikely to be emitted as relaxation progresses. Thornton and Harms, similarly, used an error prone polymerase to generate mutants of an ancestral protein of the vertebrate glucocorticoid receptor (GR).[268] They showed that the evolution of the GR protein from its ancestor occurred through a series of epistatic changes that are otherwise improbable and non-deterministic. Such an event is so rare, that if evolution were to reoccur from the start the GR protein would evolve either from a completely different ancestor or not at all; more generally the vertebrate endocrine system would be substantially different.

Evidence of epistasis has been also been shown experimentally in model organisms. For example, epistasis was shown to be responsible for evolving a citrate metabolism gene in *E. coli* over ~30000 generations.[263] A potentiating mutation that occurred at ~20000 generations led to significantly greater tendency for the evolution of the citrate gene in later clones. Similarly Kryazhimskiy, Desai and coworkers characterized epistasis, adaptability and fitness in *S. cerevisiae* over 500 generations.[269] They found that the adaptability of a genotype is negatively correlated with fitness; lower fitness leads to higher rate of adaptation, and differences in adaptability are almost entirely correlated with the overall fitness, rather than with individual genetic mutations underlying that fitness.[269]

Epistasis modeling can be a powerful tool to predict the evolutionary pathway undertaken by modern proteins. However, Harms and Sailer showed that even with current tools, evolution remains difficult to back-track due to epistasis.[265] Since proteins occur in an ensemble of conformations, any single mutation exerts its effect on each conformation in a slightly different way. Calculating this requires either the knowledge of how each conformation is affected by a mutation, which is currently impossible to measure, or the calculation of all higher-order epistases, which is currently computationally too difficult. Simply calculating the average ensemble effect of the mutation on the entire population, a mean field approach, leads to large uncertainties in prediction. For now, epistasis can predict phenotypes with relatively high accuracy and help connect the dots between how modern proteins came to be from ancestral ones.

6.4.2 Ancestral sequence reconstruction

With the advent of efficient sequencing techniques by Frederick Sanger in 1955,[270] Emile Zuckerkandl and Linus Pauling postulated that such sequences could be used to reconstruct the sequences of ancestral proteins. Ancestral sequence reconstruction was thus born. In simple terms, it is the extrapolation back in time of genetic sequences from current proteins to their common ancestors. It relies on sufficiently realistic models of evolution to predict ancestral states and immense progress has been made with the improvement in computational power and the development of more efficient algorithms. Details of some common methods and algorithms used in ancestral sequence reconstruction have been published.[259,271] Among the available prediction methods, Bayesian inference is believed to be able to estimate ancestral sequences with a high accuracy and has been widely used.[272]



	T. therm-ophilus	3	2	1	A	B	C	D	*E. coli*
ΔG_{unf}	53.6	52.3	43.9	42.7	39.3	36.0	38.9	38.9	40.6
k_{unf}	$4x10^{-6}$	$2.1x10^{-4}$	$1.1x10^{-3}$	$1.6x10^{-3}$	$2.4x10^{-4}$	$1.1x10^{-4}$	$1.7x10^{-4}$	$7.3x10^{-6}$	$1.7x10^{-5}$
k_{fold}	0.6	4.1	14.1	10.4	9.7	0.77	2.8	.13	.74

Figure 15: Thermodynamic and kinetic evolution of modern RNase H thermophilic and mesophilic homologs. Data for figure was adapted from ref. 279. Overall stability is reflected by ΔG_{unf}, folding rate by k_{fold} and unfolding rate by k_{unf}.

Once ancestral protein sequences have been predicted, they can be experimentally reconstructed in two ways. First, via step-by-step site-directed mutagenesis of specific residues in the modern protein, recapitulating in reverse the path from ancestral to modern protein.[273,274] This approach is limited to proteins for which structure-function relationships are well understood and only works under the assumption that the mutations themselves do not significantly affect the function or folding of the protein. Alternatively, the entire sequence of the ancestral sequence can be assembled in its entirety *de novo* provided that the ancestral sequence is known or can be inferred.[275] This method does not require prior knowledge of the structure-function relationship.

Of course, site-directed mutagenesis can also be used to reconstruct the full set of mutations of an ancestral protein at once.[276]

Benner and coworkers reconstructed the ancestor of the yeast protein that consumes and metabolizes ethanol. Since most organisms cannot metabolize alcohol, this lends yeast a significant survival advantage over other competing organisms.[277] The ancestor of this enzyme specialized not in alcohol consumption but in alcohol production. The alcohol production was a consequence of recycling NADH during anaerobic glycolysis. The alcohol is eventually lost to the environment. Similarly, sequence reconstruction also showed that ancient enzymes exhibit a relatively slow evolution of protein structure even as the amino acid sequence varies.[278] This suggests that the evolution of non-promiscuous activity of highly specialized enzymes and enzyme complexes may been completed in the era of the last universal common ancestor.

More recently, ancestral sequence reconstruction was used to investigate evolution of the folding pathway of a model protein, RNaseH.[279,280] Marqusee and coworkers used sequence reconstruction to probe evolution of the thermodynamic and kinetic stability of RNase H (Figure 15) from a common ancestor into the mesophilic and thermophilic branches of modern day RNase H. They observed a 90 to 400-fold decrease in the protein unfolding rate (k_{unf}) as well as 10 to 20-fold decrease in protein folding rate (k_{fold}) for mesophiles and thermophiles over the ancestor. The large decrease in unfolding rate shows evolutionary pressure in favor of increasing kinetic stability. Increasing kinetic stability allows the protein to be more resistant to unfolding and subsequent aggregation or misfolding. Thus, reduction of unfolding rate is an evolved protein property. However, the smaller decrease in the folding rates of RNase H could indicate neutral drift. Therefore, evolution does not seek to evolve fast folding proteins beyond a certain limit. Good enough is often good enough.

While both thermophilic and mesophilic proteins are stabilized overall, thermophile RNase H stability is much higher. This is possible because RNase H is a three-state folder allowing for changes in the rate-limiting step, from native to the intermediate state, to only affect kinetic stability without much change to the overall thermodynamic stability.[281]

This can be easily visualized by the equation where for a two-state folder:

$$\Delta G_{unf} = -RT\ln\left(k_{unf}/k_{fold}\right) \quad [16]$$

whereas, for a three-state folder it is:

$$\Delta G_{unf} = \Delta G_{intermediate} + -RT\ln\left(k_{unf,rate-limiting}/k_{fold,rate-limiting}\right) \quad [17]$$

7. What is yet to come: looking at protein biophysics from an evolutionary perspective

Biological systems are highly evolved, robust and efficient molecular machines, not just structurally and functionally but also in terms of interactions. Promiscuous or generalist proteins

were more common in our ancestors. They provided a wider array of lesser catalytic activities, more useful in a rapidly changing fitness landscape such as the Paleoarchean era. Specificity evolved as enhancement of certain functional traits, better adopted to stable niches, improved fitness. However, not all proteins became specific. This is because of two reasons: 1) Specificity is difficult to maintain; two highly specific partners must co-evolve to keep function intact. 2) If all proteins became highly specialized, then the biomolecule load due to protein variety N in a cell would be enormous. As described before, non-specific interactions that scale as N^2 would then become highly detrimental to the organism's survival. Thus more proteins that are 'perfect ' is not necessarily better than fewer proteins that are 'good enough,' and so generalists still abound in modern day species.[282,283] Over a third of modern *E. coli* enzymes exhibit promiscuity.[283] However, promiscuity does remain a source for new specific functions in the future, should the environment change again.[284]

Quinary interactions likely evolved from non-functional sticking, very likely assisted by epistasis. A later mutation that made a pre-existing sticky interaction weakly functional would have been very beneficial to organism fitness. Moreover, for the few proteins that have many binding partners, like chaperones, promiscuity is an inherent trait. In *E. coli* the promiscuous interactions that have survived evolution have weakly-binding K_d in the μM regime,[282] a hallmark of quinary structure formation. Promiscuous interactions and quinary structure were therefore shaping the organization of even the earliest cells.

Quinary structure formation is strongly influenced by the crowded cellular environment and purification methods often disrupt such interactions. *In vitro* experiments in simple buffers, therefore, only have limited access to probing quinary structure and in-cell studies or carefully constructed more complex *in vitro* systems are needed. With technical advancements, instruments to image quinary structure inside cells are fortunately becoming more routine.

Much remains to be gained by collaborative efforts between evolutionary biology and biophysical chemistry. The integration of the knowledge of protein structure, function and evolution will provide a complete picture of how proteins interact inside cells.

Biographies

Martin Gruebele obtained his B.S. in 1984 and his Ph.D. in 1988 at UC Berkeley. He went on to investigate femtochemistry in the lab of Ahmed Zewail at Caltech, and then moved to the University of Illinois in 1992, where he is currently the James R. Eiszner Professor of Chemistry, Professor of Physics, and Professor of Biophysics and Computational Biology. He is a Fellow of the American Physical, Chemical, and Biophysical Societies, as well as a recipient of the Sacker International Prize in Biophysics, the ACS Nakanishi Prize, and the Wilhelm Bessel Prize. He is a member of the German National Academy of Science, of the American Academy of Arts and Sciences, and of the National Academy of Sciences (NAS). He served as Senior Editor of *J. Phys. Chem.* and Associate Editor of *JACS* and is currently Head of the Department of Chemistry. His research includes protein and RNA folding, biomolecular dynamics in live cells, vibrational

energy flow in molecules, quantum computing and quantum control, optically assisted scanning microscopy, glass dynamics, bacterial signaling, and vertebrate swimming behavior.

Drishti Guin received a B.Sc. and M.Sc in Chemistry from the Indian Institute of Technology, Kharagpur in 2013. Her master thesis research focused on molecular dynamics simulation of the Aβ peptide to study *in-silico* aggregation behavior. She is currently pursuing her PhD in Physical Chemistry at the University of Illinois Urbana-Champaign under the guidance of Prof. Martin Gruebele. She studies protein-protein binding inside cells to understand how the in-cell environment affects protein stability and function. Additionally she is also involved with the development and characterization of novel protein denaturants.

Acknowledgements

This work was supported by the National Science Foundation, grant MCB 1803786. We thank Dr. Taras Pogorelov and Ms. Meredith Rickard for creating the image for Figure 6.

REFERENCES

(1) Caplow, M.; Fee, L. Dissociation of the Tubulin Dimer Is Extremely Slow, Thermodynamically Very Unfavorable, and Reversible in the Absence of an Energy Source. *Mol. Biol. Cell* **2002**, *13*, 2120–2131.

(2) Lynch, M.; Hagner, K. Evolutionary Meandering of Intermolecular Interactions along the Drift Barrier. *Proc. Natl. Acad. Sci.* **2015**, *112*, E30–E38.

(3) Zhou, H.-X.; Rivas, G.; Minton, A. P. Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences. In *Annual Review of Biophysics* **2008**, *37*, 375–397.

(4) Sapir, L.; Harries, D. Is the Depletion Force Entropic? Molecular Crowding beyond Steric Interactions. *Curr. Opin. Colloid Interface Sci.* **2015**, *20*, 3–10.

(5) Cohen, R. D.; Pielak, G. J. A Cell Is More than the Sum of Its (Dilute) Parts: A Brief History of Quinary Structure. *Protein Sci.* **2017**, *26*, 403–413.

(6) Wirth, A. J.; Gruebele, M. Quinary Protein Structure and the Consequences of Crowding in Living Cells: Leaving the Test-Tube Behind. *BioEssays* **2013**, *35*, 984–993.

(7) Chien, P.; Gierasch, L. M. Challenges and Dreams: Physics of Weak Interactions Essential to Life. *Mol. Biol. Cell* **2014**, *25*, 3474–3477.

(8) Gierasch, L. M.; Gershenson, A. Post-Reductionist Protein Science, or Putting Humpty Dumpty Back Together Again. *Nat. Chem. Biol.* **2009**, *5*, 774–777.

(9) ALLISON, A. C. Properties of Sickle-Cell Haemoglobin. *Biochem. J.* **1957**, *65*, 212–219.

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

(11) Zimmerman, S. B.; Trach, S. O. Estimation of Macromolecule Concentrations and Excluded Volume Effects for the Cytoplasm of Escherichia Coli. *J. Mol. Biol.* **1991**, *222*, 599–620.

(12) Minton, A. P. Confinement as a Determinant of Macromolecular Structure and Reactivity. *Biophys. J.* **1992**, *63*, 1090–1100.

(13) Minton, A. P. Excluded Volume as a Determinant of Macromolecular Structure and Reactivity. *Biopolymers* **1981**, *20*, 2093–2120.

(14) Condon, E. U.; Mack, J. E. An Interpretation of Pauli's Exclusion Principle. *Phys. Rev.* **1930**, *35*, 579–582.

(15) Medalia, O.; Weber, I.; Frangakis, A. S.; Nicastro, D.; Gerisch, G.; Baumeister, W. Macromolecular Architecture in Eukaryotic Cells Visualized by Cryoelectron Tomography. *Science* **2002**, *298*, 1209–1213.

(16) Saxton, M. J. Anomalous Diffusion Due to Binding: A Monte Carlo Study. *Biophys. J.* **1996**, *70*, 1250–1262.

(17) Vöpel, T.; Makhatadze, G. I. Enzyme Activity in the Crowded Milieu. *PLoS One* **2012**, *7*, 3–8.

(18) Dhar, A.; Samiotakis, A.; Ebbinghaus, S.; Nienhaus, L.; Homouz, D.; Gruebele, M.; Cheung, M. S. Structure, Function, and Folding of Phosphoglycerate Kinase Are Strongly Perturbed by Macromolecular Crowding. *Proc. Natl. Acad. Sci.* **2010**, *107*, 17586–17591.

(19) Jiang, M.; Guo, Z. Effect Macromolecular Crowding on the Intrinsic Cat Eff of Enterobactin Isochorismate Synthase. *J. Am. Chem. Soc.* **2007**, *129*, 730–731.

(20) Munishkina, L.; Cooper, E.; Uversky, V.; Fink, A. The Effect of Macromolecular Crowding on Protein Aggregation and Amyloid Fibril Formation. *J. Mol. Recognit.* **2004**, *17*, 456–464.

(21) del Alamo, M.; Rivas, G.; Mateu, M. G. Effect of Macromolecular Crowding Agents on Human Immunodeficiency Virus Type 1 Capsid Protein Assembly in Vitro. *J. Virol.* **2005**, *79*, 14271–14281.

(22) Zorrilla, S.; Rivas, G.; Acuña, A. U.; Lillo, M. P. Protein Self-Association in Crowded Protein Solutions: A Time-Resolved Fluorescence Polarization Study. *Protein Sci.* **2008**, *13*, 2960–2969.

(23) Díaz-López, T.; Dávila-Fajardo, C.; Blaesing, F.; Lillo, M. P.; Giraldo, R. Early Events in the Binding of the PPS10 Replication Protein RepA to Single Iteron and Operator DNA Sequences. *J. Mol. Biol.* **2006**, *364*, 909–920.

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

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

(26) Benton, L. a; Smith, A. E.; Young, G. B.; Pielak, G. J. Unexpected Effects of Macromolecular Crowding on Protein Stability Unexpected Effects of Macromolecular Crowding on Protein Stability. *Biochemistry* **2012**, 9773–9775.

(27) Guzman, I.; Gelman, H.; Tai, J.; Gruebele, M. The Extracellular Protein VlsE Is Destabilized inside Cells. *J. Mol. Biol.* **2014**, *426*, 11–20.

(28) Derham, B. K.; Harding, J. J. The Effect of the Presence of Globular Proteins and Elongated Polymers on Enzyme Activity. *Biochim. Biophys. Acta - Proteins Proteomics* **2006**, *1764*, 1000–1006.

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

(30) Vaĭnshteĭn, B. K. Three-Dimensional Electron Microscopy of Biological Macromolecules. *Sov. Phys. Uspekhi* **1973**, *16*, 185–206.

(31) 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.* **1980**, *32*, 347–360.

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

(33) Rivas, G.; Minton, A. P. *Macromolecular Crowding In Vitro, In Vivo, and In Between*. Trends Biochem. Sci. **2016**, *41*, 970–981.

(34) Elcock, A. H. Models of Macromolecular Crowding Effects and the Need for Quantitative Comparisons with Experiment. *Curr. Opin. Struct. Biol.* **2010**, *20*, 196–206.

(35) Rivas, G.; Ferrone, F.; Herzfeld, J. Life in a Crowded World. In *EMBO Reports*; **2004**, *5*, 23–27.

(36) Hall, D.; Minton, A. P. Macromolecular Crowding: Qualitative and Semiquantitative Successes, Quantitative Challenges. *Biochimica et Biophysica Acta - Proteins and Proteomics*. **2003**, 1649, 127–139.

(37) Kuznetsova, I. M.; Turoverov, K. K.; Uversky, V. N. What Macromolecular Crowding Can Do to a Protein. *Int. J. Mol. Sci.* **2014**, *15*, 23090–23140.

(38) Christiansen, A.; Wang, Q.; Samiotakis, A.; Cheung, M. S.; Wittung-Stafshede, P. Factors Defining Effects of Macromolecular Crowding on Protein Stability: An In Vitro/In Silico

Case Study Using Cytochrome *C*. *Biochemistry* **2010**, *49*, 6519–6530.

(39) Gnutt, D.; Gao, M.; Brylski, O.; Heyden, M.; Ebbinghaus, S. Excluded-Volume Effects in Living Cells. *Angew. Chemie - Int. Ed.* **2015**, *54*, 2548–2551.

(40) Lebowitz, J. L.; Helfand, E.; Psaestgaakd, E. Scaled Particle Theory of Fluid Mixtures. *J. Chem. Phys.* **1965**, *43*, 774–779.

(41) Giddings, J. C.; Kucera, E.; Russell, C. P.; Myers, M. N. Statistical Theory for the Equilibrium Distribution of Rigid Molecules in Inert Porous Networks. Exclusion Chromatography. *J. Phys. Chem.* **1968**, *72*, 4397–4408.

(42) Zaccone, A.; Terentjev, E. M. Theory of Molecular Crowding in Brownian Hard-Sphere Liquids. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **2012**, *85*, 061202.

(43) Widom, B. Some Topics in the Theory of Fluids. *J. Chem. Phys.* **1963**, *39*, 2808–2812.

(44) Cheung, M. S.; Klimov, D.; Thirumalai, D. Molecular Crowding Enhances Native State Stability and Refolding Rates of Globular Proteins. *Proc. Natl. Acad. Sci.* **2005**, *102*, 4753–4758.

(45) Minton, A. P. Effect of a Concentrated “inert” Macromolecular Cosolute on the Stability of a Globular Protein with Respect to Denaturation by Heat and by Chaotropes: A Statistical-Thermodynamic Model. *Biophys. J.* **2000**, *78*, 101–109.

(46) Minton, A. P. Models for Excluded Volume Interaction between an Unfolded Protein and Rigid Macromolecular Cosolutes: Macromolecular Crowding and Protein Stability Revisited. *Biophys. J.* **2005**, *88*, 971–985.

(47) Hayer-Hartl, M.; Minton, A. P. A Simple Semiempirical Model for the Effect of Molecular Confinement upon the Rate of Protein Folding. *Biochemistry* **2006**, *45*, 13356–13360.

(48) Tubio, G.; Nerli, B.; Picó, G. Relationship between the Protein Surface Hydrophobicity and Its Partitioning Behaviour in Aqueous Two-Phase Systems of Polyethyleneglycol–Dextran. *J. Chromatogr. B* **2004**, *799*, 293–301.

(49) Winzor, D. J.; Wills, P. R. Molecular Crowding Effects of Linear Polymers in Protein Solutions. *Biophys. Chem.* **2006**, *119*, 186–195.

(50) Bloustine, J.; Virmani, T.; Thurston, G. M.; Fraden, S. Light Scattering and Phase Behavior of Lysozyme-Poly(Ethylene Glycol) Mixtures. *Phys. Rev. Lett.* **2006**, *96*, 087803.

(51) Chao, S. H.; Matthews, S. S.; Paxman, R.; Aksimentiev, A.; Gruebele, M.; Price, J. L. Two Structural Scenarios for Protein Stabilization by PEG. *J. Phys. Chem. B* **2014**, *118*, 8388–8395.

(52) Delgado, C.; Francis, G. E.; Fisher, D. The Uses and Properties of PEG-Linked Proteins. *Crit. Rev. Ther. Drug Carr. Syst.* **1992**, *9*, 249–304.

(53) Chao, S. H.; Schäfer, J.; Gruebele, M. The Surface of Protein Λ6-85Can Act as a Template for Recurring Poly(Ethylene Glycol) Structure. *Biochemistry* **2017**, *56*, 5671–5678.

(54) Balbo, J.; Mereghetti, P.; Herten, D. P.; Wade, R. C. The Shape of Protein Crowders Is a Major Determinant of Protein Diffusion. *Biophys. J.* **2013**, *104*, 1576–1584.

(55) Olsen, S. N. Applications of Isothermal Titration Calorimetry to Measure Enzyme Kinetics and Activity in Complex Solutions. *Thermochim. Acta* **2006**, *448*, 12–18.

(56) Denos, S.; Dhar, A.; Gruebele, M. Crowding Effects on the Small, Fast-Folding Protein Λ6–85. *Faraday Discuss.* **2012**, *157*, 451–500.

(57) Minton, a P.; Wilf, J. Effect of Macromolecular Crowding upon the Structure and Function of an Enzyme: Glyceraldehyde-3-Phosphate Dehydrogenase. *Biochemistry* **1981**, *20*, 4821–4826.

(58) Zhou, B. R.; Liang, Y.; Du, F.; Zhou, Z.; Chen, J. Mixed Macromolecular Crowding Accelerates the Oxidative Refolding of Reduced, Denatured Lysozyme: Implications for Protein Folding in Intracellular Environments. *J. Biol. Chem.* **2004**, *279*, 55109–55116.

(59) Bai, J.; Liu, M.; Pielak, G. J.; Li, C. Macromolecular and Small Molecular Crowding Have Similar Effects on α-Synuclein Structure. *ChemPhysChem* **2017**, *18*, 55–58.

(60) van den Berg, B.; Wain, R.; Dobson, C. M.; Ellis, R. J. Macromolecular Crowding Perturbs Protein Refolding Kinetics: Implications for Folding inside the Cell. *Embo J.* **2000**, *19*, 3870–3875.

(61) Biswas, S.; Kundu, J.; Mukherjee, S. K.; Chowdhury, P. K. Mixed Macromolecular Crowding: A Protein and Solvent Perspective. *ACS Omega* **2018**, *3*, 4316–4330.

(62) Goins, A. B.; Sanabria, H.; Waxham, M. N. Macromolecular Crowding and Size Effects on Probe Microviscosity. *Biophys. J.* **2008**, *95*, 5362–5373.

(63) Banks, D. S.; Fradin, C. Anomalous Diffusion of Proteins Due to Molecular Crowding. *Biophys. J.* **2005**, *89*, 2960–2971.

(64) Schermann, J. P. Hole-Burning Spectroscopy. In *Encyclopedia of Spectroscopy and Spectrometry*; (Academic Press, London) **2010**.

(65) Volker, S. Hole-Burning Spectroscopy. *Annu. Rev. Phys. Chem.* **1989**, *40*, 499–530.

(66) Zhou, H. X. Protein Folding in Confined and Crowded Environments. *Archives of Biochemistry and Biophysics*. **2008**, *469*, 76–82.

(67) Qin, S.; Zhou, H. X. Atomistic Modeling of Macromolecular Crowding Predicts Modest Increases in Protein Folding and Binding Stability. *Biophys. J.* **2009**, *97*, 12–19.

(68) Zhou, H. X. Effect of Mixed Macromolecular Crowding Agents on Protein Folding. *Proteins Struct. Funct. Genet.* **2008**, *72*, 1109–1113.

(69) Zhou, H. X. Protein Folding and Binding in Confined Spaces and in Crowded Solutions. *J. Mol. Recognit.* **2004**, *17*, 368–375.

(70) Batra, J.; Xu, K.; Zhou, H. X. Nonadditive Effects of Mixed Crowding on Protein Stability. *Proteins Struct. Funct. Bioinforma.* **2009**, *77*, 133–138.

(71) Batra, J.; Xu, K.; Qin, S.; Zhou, H. X. Effect of Macromolecular Crowding on Protein Binding Stability: Modest Stabilization and Significant Biological Consequences. *Biophys. J.* **2009**, *97*, 906–911.

(72) Zhou, H.-X. Crowding Effects of Membrane Proteins. *J. Phys. Chem. B* **2009**, *113*, 7995–8005.

(73) Guseman, A. J.; Perez Goncalves, G. M.; Speer, S. L.; Young, G. B.; Pielak, G. J. Protein Shape Modulates Crowding Effects. *Proc. Natl. Acad. Sci.* **2018**, *115*, 10965–10970.

(74) Chen, E.; Christiansen, A.; Wang, Q.; Cheung, M. S.; Kliger, D. S.; Wittung-Stafshede, P. Effects of Macromolecular Crowding on Burst Phase Kinetics of Cytochrome c Folding. *Biochemistry* **2012**, *51*, 9836–9845.

(75) Homouz, D.; Stagg, L.; Wittung-Stafshede, P.; Cheung, M. S. Macromolecular Crowding Modulates Folding Mechanism of α/β Protein Apoflavodoxin. *Biophys. J.* **2009**, *96*, 671–680.

(76) Tripathi, S.; Wang, Q.; Zhang, P.; Hoffman, L.; Waxham, M. N.; Cheung, M. S. Conformational Frustration in Calmodulin-Target Recognition. *J. Mol. Recognit.* **2015**, *28*, 74–86.

(77) Wang, Q.; Zhang, P.; Hoffman, L.; Tripathi, S.; Homouz, D.; Liu, Y.; Waxham, M. N.; Cheung, M. S. Protein Recognition and Selection through Conformational and Mutually Induced Fit. *Proc. Natl. Acad. Sci.* **2013**, *110*, 20545–20550.

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

(79) Specht, E. A.; Braselmann, E.; Palmer, A. E. A Critical and Comparative Review of Fluorescent Tools for Live-Cell Imaging. *Annu. Rev. Physiol.* **2017**, *79*, 93–117.

(80) Guigas, G.; Kalla, C.; Weiss, M. The Degree of Macromolecular Crowding in the Cytoplasm and Nucleoplasm of Mammalian Cells Is Conserved. *FEBS Lett.* **2007**, *581*, 5094–5098.

(81) Boersma, A. J.; Zuhorn, I. S.; Poolman, B. A Sensor for Quantification of Macromolecular Crowding in Living Cells. *Nat. Methods* **2015**, *12*, 227–229.

(82) Machiyama, H.; J. Morikawa, T.; Okamoto, K.; M. Watanabe, T.; Fujita, H. The Use of a Genetically Encoded Molecular Crowding Sensor in Various Biological Phenomena. *Biophys. Physicobiology* **2017**, *14*, 119–125.

(83) Liu, B.; Åberg, C.; van Eerden, F. J.; Marrink, S. J.; Poolman, B.; Boersma, A. J. Design

and Properties of Genetically Encoded Probes for Sensing Macromolecular Crowding. *Biophys. J.* **2017**, *112*, 1929–1939.

(84) Morikawa, T. J.; Fujita, H.; Kitamura, A.; Horio, T.; Yamamoto, J.; Kinjo, M.; Sasaki, A.; Machiyama, H.; Yoshizawa, K.; Ichimura, T.; et al. Dependence of Fluorescent Protein Brightness on Protein Concentration in Solution and Enhancement of It. *Sci. Rep.* **2016**, *6*, 22342.

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

(86) Davis, F. F.; Abuchowski, A.; van Es, T.; Palczuk, N. C.; Chen, R.; Savoca, K.; Wieder, K. Enzyme-Polyethylene Glycol Adducts: Modified Enzymes with Unique Properties. In *Enzyme Engineering, Broun G.B., Manecke G., Wingard L.B. (editors);* Springer, Berlin, 1978, pp 169–173.

(87) Milton Harris, J.; Chess, R. B. Effect of Pegylation on Pharmaceuticals. *Nature Reviews Drug Discovery*. **2003**, *2*, 214–221.

(88) Zhang, Z.; Chen, X.; Chen, L.; Yu, S.; Cao, Y.; He, C.; Chen, X. Intracellular PH-Sensitive PEG- Block -Acetalated-Dextrans as Efficient Drug Delivery Platforms. *ACS Appl. Mater. Interfaces* **2013**, *5*, 10760–10766.

(89) Huang, J.; Wang, L.; Xiong, C.; Yuan, F. Elastic Hydrogel as a Sensor for Detection of Mechanical Stress Generated by Single Cells Grown in Three-Dimensional Environment. *Biomaterials* **2016**, *98*, 103–112.

(90) Jin, W.; Brennan, J. D. Properties and Applications of Proteins Encapsulated within Sol-Gel Derived Materials. *Analytica Chimica Acta*. **2002**, *461*, 1–36.

(91) Kisley, L.; Landes, C. F. Molecular Approaches to Chromatography Using Single Molecule Spectroscopy. *Analytical Chemistry*. **2015**, *9*, 83–98.

(92) Dong, R.; Jensen, T. W.; Engberg, K.; Nuzzo, R. G.; Leckband, D. E. Variably Elastic Hydrogel Patterned via Capillary Action in Microchannels. *Langmuir* **2007**, *23*, 1483–1488.

(93) Eggers, D. K.; Valentine, J. S. Molecular Confinement Influences Protein Structure and Enhances Thermal Protein Stability. *Protein Sci.* **2001**, *10*, 250–261.

(94) Peterson, E. S.; Leonard, E. F.; Foulke, J. A.; Oliff, M. C.; Salisbury, R. D.; Kim, D. Y. Folding Myoglobin within a Sol-Gel Glass: Protein Folding Constrained to a Small Volume. *Biophys. J.* **2008**, *95*, 322–332.

(95) Kisley, L.; Serrano, K. A.; Guin, D.; Kong, X.; Gruebele, M.; Leckband, D. E. Direct Imaging of Protein Stability and Folding Kinetics in Hydrogels. *ACS Appl. Mater. Interfaces* **2017**, *9*, 21606–21617.

(96) Kisley, L.; Serrano, K. A.; Davis, C. M.; Guin, D.; Murphy, E. A.; Gruebele, M.; Leckband, D. E. Soluble Zwitterionic Poly(Sulfobetaine) Destabilizes Proteins. *Biomacromolecules*

2018, *19*, 3894–3901.

(97) Sarkar, M.; Smith, A. E.; Pielak, G. J. Impact of Reconstituted Cytosol on Protein Stability. *Proc. Natl. Acad. Sci.* **2013**, *110*, 19342–19347.

(98) Monteith, W. B.; Cohen, R. D.; Smith, A. E.; Guzman-Cisneros, E.; Pielak, G. J. Quinary Structure Modulates Protein Stability in Cells. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 1739–1742.

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

(100) Barbieri, L.; Luchinat, E.; Banci, L. Protein Interaction Patterns in Different Cellular Environments Are Revealed by In-Cell NMR. *Sci. Rep.* **2015**, *5*, 14456.

(101) Monteith, W. B.; Pielak, G. J. Residue Level Quantification of Protein Stability in Living Cells. *Proc. Natl. Acad. Sci.* **2014**, *111*, 11335–11340.

(102) Danielsson, J.; Mu, X.; Lang, L.; Wang, H.; Binolfi, A.; Theillet, F.-X.; Bekei, B.; Logan, D. T.; Selenko, P.; Wennerström, H.; et al. Thermodynamics of Protein Destabilization in Live Cells. *Proc. Natl. Acad. Sci.* **2015**, *112*, 12402–12407.

(103) Miklos, A. C.; Sarkar, M.; Wang, Y.; Pielak, G. J. Protein Crowding Tunes Protein Stability. *J. Am. Chem. Soc.* **2011**, *133*, 7116–7120.

(104) Sarkar, M.; Lu, J.; Pielak, G. J. Protein Crowder Charge and Protein Stability. *Biochemistry* **2014**, *53*, 1601–1606.

(105) Wang, Y.; Sarkar, M.; Smith, A. E.; Krois, A. S.; Pielak, G. J. Macromolecular Crowding and Protein Stability. *J. Am. Chem. Soc.* **2012**, *134*, 16614–16618.

(106) Cleland, J. L.; Powell, M. F.; Shire, S. J. The Development of Stable Protein Formulations: A Close Look at Protein Aggregation, Deamidation, and Oxidation. *Crit. Rev. Ther. Drug Carrier Syst.* **1993**, *10*, 307–377.

(107) Carpenter, J. F.; Kendrick, B. S.; Chang, B. S.; Manning, M. C.; Randolph, T. W. Inhibition of Stress-Induced Aggregation of Protein Therapeutics. *Methods Enzymol.* **1999**, *309*, 236–255.

(108) Macdougall, I. C. Antibody-Mediated Pure Red Cell Aplasia (PRCA): Epidemiology, Immunogenicity and Risks. *Nephrol. Dial. Transplant.* **2005**, *20*, 9-15.

(109) Jiskoot, W.; Randolph, T. W.; Volkin, D. B.; Middaugh, C. R.; Schöneich, C.; Winter, G.; Friess, W.; Crommelin, D. J. A.; Carpenter, J. F. Protein Instability and Immunogenicity: Roadblocks to Clinical Application of Injectable Protein Delivery Systems for Sustained Release. *J. Pharm. Sci.* **2012**, *101*, 946–954.

(110) Moore, W. V.; Leppert, P. Role of Aggregated Human Growth Hormone (HGH) in Development of Antibodies to HGH. *J. Clin. Endocrinol. Metab.* **1980**, *51*, 691–697.

(111) Ranke, M. B.; Wit, J. M. Growth Hormone-Past, Present and Future. *Nat. Rev. Endocrinol.* **2018**, *14*, 285–300.

(112) St. John, R. J.; Carpenter, J. F.; Balny, C.; Randolph, T. W. High Pressure Refolding of Recombinant Human Growth Hormone from Insoluble Aggregates: Structural Transformations, Kinetic Barriers, and Energetics. *J. Biol. Chem.* **2001**, *276*, 46856–46863.

(113) Eckhardt, B. M.; Oeswein, J. Q.; Bewley, T. A. Effect of Freezing on Aggregation of Human Growth Hormone. *Pharmaceutical Research.* **1991**, *8*, 1360–1364.

(114) Gomez-Orellana, I.; Variano, B.; Miura-Fraboni, J.; Milstein, S.; Paton, D. R. Thermodynamic Characterization of an Intermediate State of Human Growth Hormone. *Protein Sci.* **1998**, *7*, 1352–1358.

(115) Johansson, H.; Jensen, M. R.; Gesmar, H.; Meier, S.; Vinther, J. M.; Keeler, C.; Hodsdon, M. E.; Led, J. J. Specific and Nonspecific Interactions in Ultraweak Protein-Protein Associations Revealed by Solvent Paramagnetic Relaxation Enhancements. *J. Am. Chem. Soc.* **2014**, *136*, 10277–10286.

(116) Jensen, M. R.; Kristensen, S. M.; Keeler, C.; Christensen, H. E. M.; Hodsdon, M. E.; Led, J. J. Weak Self-Association of Human Growth Hormone Investigated by Nitrogen-15 NMR Relaxation. *Proteins Struct. Funct. Genet.* **2008**, *73*, 161–172.

(117) Tang, C.; Ghirlando, R.; Clore, G. M. Visualization of Transient Ultra-Weak Protein Self-Association in Solution Using Paramagnetic Relaxation Enhancement. *J. Am. Chem. Soc.* **2008**, *130*, 4048–4056.

(118) Dhar, A.; Ebbinghaus, S.; Shen, Z.; Mishra, T.; Gruebele, M. The Diffusion Coefficient for PGK Folding in Eukaryotic Cells. *Biophys. J.* **2010**, *99*, L69–L71.

(119) Phillip, Y.; Kiss, V.; Schreiber, G. Protein-Binding Dynamics Imaged in a Living Cell. *Proc. Natl. Acad. Sci.* **2012**, *109*, 1461–1466.

(120) Margineanu, A.; Chan, J. J.; Kelly, D. J.; Warren, S. C.; Flatters, D.; Kumar, S.; Katan, M.; Dunsby, C. W.; French, P. M. W. Screening for Protein-Protein Interactions Using Förster Resonance Energy Transfer (FRET) and Fluorescence Lifetime Imaging Microscopy (FLIM). *Sci. Rep.* **2016**, *6*, 33621.

(121) Dhar, A.; Girdhar, K.; Singh, D.; Gelman, H.; Ebbinghaus, S.; Gruebele, M. Protein Stability and Folding Kinetics in the Nucleus and Endoplasmic Reticulum of Eucaryotic Cells. *Biophys. J.* **2011**, *101*, 421–430.

(122) Ebbinghaus, S.; Dhar, A.; McDonald, J. D.; Gruebele, M. Protein Folding Stability and Dynamics Imaged in a Living Cell. *Nat. Methods* **2010**, *7*, 319–323.

(123) Willem Borst, J.; Visser, A. J. W. G. Fluorescence Lifetime Imaging Microscopy in Life Sciences. *Measurement Science and Technology*. **2010**, *21*, 102002.

(124) Krieger, J. W.; Singh, A. P.; Bag, N.; Garbe, C. S.; Saunders, T. E.; Langowski, J.;

Wohland, T. Imaging Fluorescence (Cross-) Correlation Spectroscopy in Live Cells and Organisms. *Nat. Protoc.* **2015**, *10*, 1948–1974.

(125) Seksek, O.; Biwersi, J.; Verkman, A. S. Translational Diffusion of Macromolecule-Sized Solutes in Cytoplasm and Nucleus. *J. Cell Biol.* **1997**, *138*, 131–142.

(126) Etoc, F.; Balloul, E.; Vicario, C.; Normanno, D.; Liße, D.; Sittner, A.; Piehler, J.; Dahan, M.; Coppey, M. Non-Specific Interactions Govern Cytosolic Diffusion of Nanosized Objects in Mammalian Cells. *Nat. Mater.* **2018**, *17*, 740–746.

(127) Dave, K.; Gelman, H.; Thu, C. T. H.; Guin, D.; Gruebele, M. The Effect of Fluorescent Protein Tags on Phosphoglycerate Kinase Stability Is Nonadditive. *J. Phys. Chem. B* **2016**, *120*, 2878–2885.

(128) Serber, Z.; Selenko, P.; Hänsel, R.; Reckel, S.; Löhr, F.; Ferrell, J. E.; Wagner, G.; Dötsch, V. Investigating Macromolecules inside Cultured and Injected Cells by In-Cell NMR Spectroscopy. *Nat. Protoc.* **2007**, *1*, 2701–2709.

(129) Barbieri, L.; Luchinat, E.; Banci, L. Characterization of Proteins by In-Cell NMR Spectroscopy in Cultured Mammalian Cells. *Nat Protoc* **2016**, *11*, 1101–1111.

(130) Luchinat, E.; Banci, L. A Unique Tool for Cellular Structural Biology: In-Cell NMR. *Journal of Biological Chemistry*. **2016,** *8*, 3776–3784.

(131) 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.* **2011**, *133*, 8082–8085.

(132) Inomata, K.; Ohno, A.; Tochio, H.; Isogai, S.; Tenno, T.; Nakase, I.; Takeuchi, T.; Futaki, S.; Ito, Y.; Hiroaki, H.; et al. High-Resolution Multi-Dimensional NMR Spectroscopy of Proteins in Human Cells. *Nature* **2009**, *458*, 106–109.

(133) Bicout, D. J.; Field, M. J. Stochastic Dynamics Simulations of Macromolecular Diffusion in a Model of the Cytoplasm of Escherichia Coli . *J. Phys. Chem.* **2002**, *100*, 2489–2497.

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

(135) Konopka, M. C.; Shkel, I. A.; Cayley, S.; Record, M. T.; Weisshaar, J. C. Crowding and Confinement Effects on Protein Diffusion in Vivo. *J Bacteriol* **2006**, *188*, 6115–6123.

(136) Feig, M.; Harada, R.; Mori, T.; Yu, I.; Takahashi, K.; Sugita, Y. Complete Atomistic Model of a Bacterial Cytoplasm for Integrating Physics, Biochemistry, and Systems Biology. *J. Mol. Graph. Model.* **2015**, *58*, 1–9.

(137) 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.* **2009**, *73*, 371–388.

(138) Schwartz, R.; Ting, C. S.; King, J. Whole Proteome Pl Values Correlate with Subcellular Localizations of Proteins for Organisms within the Three Domains of Life. *Genome Res.* **2001**, *11*, 703–709.

(139) Kiraga, J.; Mackiewicz, P.; Mackiewicz, D.; Kowalczuk, M.; Biecek, P.; Polak, N.; Smolarczyk, K.; Dudek, M. R.; Cebrat, S. The Relationships between the Isoelectric Point and: Length of Proteins, Taxonomy and Ecology of Organisms. *BMC Genomics* **2007**, *8*, 163.

(140) Vaynberg, J.; Qin, J. Weak Protein-Protein Interactions as Probed by NMR Spectroscopy. *Trends in Biotechnology*. **2006**, *24*, 22–27.

(141) Srere, P. A. The Metabolon. *Trends in Biochemical Sciences*. **1985**, *10*, 109–110.

(142) Srere, P. A. Complexes of Sequential Metabolic Enzymes. *Annu. Rev. Biochem.* **1987**, *56*, 89–124.

(143) Wu, F.; Minteer, S. Krebs Cycle Metabolon: Structural Evidence of Substrate Channeling Revealed by Cross-Linking and Mass Spectrometry. *Angew. Chemie - Int. Ed.* **2015**, *54*, 1851–1854.

(144) 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.* **2016**, *11*, 2847–2853.

(145) Guo, M.; Ignatov, M.; Musier-Forsyth, K.; Schimmel, P.; Yang, X.-L. Crystal Structure of Tetrameric Form of Human Lysyl-TRNA Synthetase: Implications for Multisynthetase Complex Formation. *Proc. Natl. Acad. Sci.* **2008**, *105*, 2331–2336.

(146) Jang, S. R.; Nelson, J. C.; Bend, E. G.; Rodríguez-Laureano, L.; Tueros, F. G.; Cartagenova, L.; Underwood, K.; Jorgensen, E. M.; Colón-Ramos, D. A. Glycolytic Enzymes Localize to Synapses under Energy Stress to Support Synaptic Function. *Neuron* **2016**, *90*, 278–291.

(147) Zhang, Y.; Beard, K. F. M.; Swart, C.; Bergmann, S.; Krahnert, I.; Nikoloski, Z.; Graf, A.; George Ratcliffe, R.; Sweetlove, L. J.; Fernie, A. R.; et al. Protein-Protein Interactions and Metabolite Channelling in the Plant Tricarboxylic Acid Cycle. *Nat. Commun.* **2017**, *8*, 15212.

(148) Fernie, A. R.; Carrari, F.; Sweetlove, L. J. Respiratory Metabolism: Glycolysis, the TCA Cycle and Mitochondrial Electron Transport. *Current Opinion in Plant Biology*. **2004**, *7*, 254–261.

(149) 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* **1997**, *36*, 14271–14276.

(150) Easterby, J. S. A Generalized Theory of the Transition Time for Sequential Enzyme Reactions. *Biochem. J.* **2015**, *199*, 155–161.

(151) Ovádi, J.; Tompa, P.; Vértessy, B.; Orosz, F.; Keleti, T.; Welch, G. R. Transient-Time Analysis of Substrate-Channelling in Interacting Enzyme Systems. *Biochem. J.* **2015**, *257*, 187–190.

(152) Elcock, A. H.; Huber, G. A.; Andrew McCammon, J. Electrostatic Channeling of Substrates between Enzyme Active Sites: Comparison of Simulation and Experiment. *Biochemistry* **1997**, *36*, 16049-16058.

(153) Robinson, J. B.; Inman, L.; Sumegi, B.; Srere, P. A. Further Characterization of the Krebs Tricarboxylic Acid Cycle Metabolon. *J. Biol. Chem.* **1987**, *262*, 1786–1790.

(154) Beekmans, 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.* **1981**, *117*, 527–535.

(155) Beck, M.; Schmidt, A.; Malmstroem, J.; Claassen, M.; Ori, A.; Szymborska, A.; Herzog, F.; Rinner, O.; Ellenberg, J.; Aebersold, R. The Quantitative Proteome of a Human Cell Line. *Mol. Syst. Biol.* **2011**, *7*, 549.

(156) Liu, Y.; Hickey, D. P.; Guo, J. Y.; Earl, E.; Abdellaoui, S.; Milton, R. D.; Sigman, M. S.; Minteer, S. D.; Calabrese Barton, S. Substrate Channeling in an Artificial Metabolon: A Molecular Dynamics Blueprint for an Experimental Peptide Bridge. *ACS Catal.* **2017**, *7*, 2486–2493.

(157) Negrutskii, B. S.; Deutscher, M. P. Channeling of Aminoacyl-TRNA for Protein Synthesis in Vivo. *Proc. Natl. Acad. Sci.* **2006**, *88*, 4991–4995.

(158) Hausmann, C. D.; Ibba, M. Aminoacyl-TRNA Synthetase Complexes: Molecular Multitasking Revealed. *FEMS Microbiology Reviews*. **2008**, *32* 705–721.

(159) Bhattacharyya, S.; Bershtein, S.; Yan, J.; Argun, T.; Gilson, A. I.; Trauger, S. A.; Shakhnovich, E. I. Transient Protein-Protein Interactions Perturb E. Coli Metabolome and Cause Gene Dosage Toxicity. *Elife* **2016**, *5*, 1–22.

(160) Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. Biomolecular Condensates: Organizers of Cellular Biochemistry. *Nature Reviews Molecular Cell Biology*. **2017,** *18*, 285–298.

(161) Grousl, T.; Ivanov, P.; Frydlova, I.; Vasicova, P.; Janda, F.; Vojtova, J.; Malinska, K.; Malcova, I.; Novakova, L.; Janoskova, D.; et al. Robust Heat Shock Induces EIF2 - Phosphorylation-Independent Assembly of Stress Granules Containing EIF3 and 40S Ribosomal Subunits in Budding Yeast, Saccharomyces Cerevisiae. *J. Cell Sci.* **2009**, *122*, 2078–2088.

(162) Buchan, J. R.; Parker, R. Eukaryotic Stress Granules: The Ins and Outs of Translation. *Molecular Cell*. **2009**, *36*, 932–941.

(163) Hoyle, N. P.; Castelli, L. M.; Campbell, S. G.; Holmes, L. E. A.; Ashe, M. P. Stress-Dependent Relocalization of Translationally Primed MRNPs to Cytoplasmic Granules

That Are Kinetically and Spatially Distinct from P-Bodies. *J. Cell Biol.* **2007**, *179*, 65–74.

(164) Li, P.; Banjade, S.; Cheng, H. C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F.; et al. Phase Transitions in the Assembly of Multivalent Signalling Proteins. *Nature* **2012**, *483*, 336–340.

(165) Protter, D. S. W. W.; Rao, B. S.; Van Treeck, B.; Lin, Y.; Mizoue, L.; Rosen, M. K.; Parker, R. Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly. *Cell Rep.* **2018**, *22*, 1401–1412.

(166) Riback, J. A.; Katanski, C. D.; Kear-Scott, J. L.; Pilipenko, E. V.; Rojek, A. E.; Sosnick, T. R.; Drummond, D. A. Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* **2017**, *168*, 1028-1040.e19.

(167) Cherkasov, V.; Grousl, T.; Theer, P.; Vainshtein, Y.; Gläßer, C.; Mongis, C.; Kramer, G.; Stoecklin, G.; Knop, M.; Mogk, A.; et al. Systemic Control of Protein Synthesis through Sequestration of Translation and Ribosome Biogenesis Factors during Severe Heat Stress. *FEBS Lett.* **2015**, *589*, 3654–3664.

(168) Wallace, E. W. J.; Kear-Scott, J. L.; Pilipenko, E. V.; Schwartz, M. H.; Laskowski, P. R.; Rojek, A. E.; Katanski, C. D.; Riback, J. A.; Dion, M. F.; Franks, A. M.; et al. Reversible, Specific, Active Aggregates of Endogenous Proteins Assemble upon Heat Stress. *Cell* **2015**, *162*, 1286–1298.

(169) Tatomer, D. C.; Terzo, E.; Curry, K. P.; Salzler, H.; Sabath, I.; Zapotoczny, G.; McKay, D. J.; Dominski, Z.; Marzluff, W. F.; Duronio, R. J. Concentrating Pre-MRNA Processing Factors in the Histone Locus Body Facilitates Efficient Histone MRNA Biogenesis. *J. Cell Biol.* **2016**, *213*, 557–570.

(170) Strzelecka, M.; Trowitzsch, S.; Weber, G.; Lührmann, R.; Oates, A. C.; Neugebauer, K. M. Coilin-Dependent SnRNP Assembly Is Essential for Zebrafish Embryogenesis. *Nat. Struct. Mol. Biol.* **2010**, *17*, 403–409.

(171) Novotny, I.; Blazikova, M.; Stanek, D.; Herman, P.; Malinsky, J. In Vivo Kinetics of U4/U6{middle Dot}U5 Tri-SnRNP Formation in Cajal Bodies. *Mol. Biol. Cell* **2011**, *22*, 513–523.

(172) Read, E. B.; Okamura, H. H.; Drubin, D. G. Actin- and Tubulin-Dependent Functions during Saccharomyces Cerevisiae Mating Projection Formation. *Mol. Biol. Cell* **1992**, *3*, 429–444.

(173) Goode, B. L.; Wong, J. J.; Butty, A. C.; Peter, M.; McCormack, A. L.; Yates, J. R.; Drubin, D. G.; Barnes, G. Coronin Promotes the Rapid Assembly and Cross-Linking of Actin Filaments and May Link the Actin and Microtubule Cytoskeletons in Yeast. *J. Cell Biol.* **1999**, *144*, 83–98.

(174) Gill, S. R.; Schroer, T. A.; Szilak, I.; Steuer, E. R.; Sheetz, M. P.; Cleveland, D. W. Dynactin, a Conserved, Ubiquitously Expressed Component of an Activator of Vesicle Motility Mediated by Cytoplasmic Dynein. *J. Cell Biol.* **1991**, *115*, 1639–1650.

(175) Waterman-Storer, C. M.; Salmon, E. D. Actomyosin-Based Retrograde Flow of Microtubules in the Lamella of Migrating Epithelial Cells Influences Microtubule Dynamic Instability and Turnover and Is Associated with Microtubule Breakage and Treadmilling. *J. Cell Biol.* **1997**, *139*, 417–434.

(176) Gavin, A.-C.; Bösche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J. M.; Michon, A.-M.; Cruciat, C.-M.; et al. Functional Organization of the Yeast Proteome by Systematic Analysis of Protein Complexes. Nature **2002**, *415*, 141–147.

(177) Blundell, T. L.; Fernández-Recio, J. Cell Biology: Brief Encounters Bolster Contacts. *Nature*. **2006**, *444*, 279–280.

(178) Tang, C.; Iwahara, J.; Clore, G. M. Visualization of Transient Encounter Complexes in Protein-Protein Association. *Nature* **2006**, *444*, 383–386.

(179) Schlecht, R.; Erbse, A. H.; Bukau, B.; Mayer, M. P. Mechanics of Hsp70 Chaperones Enables Differential Interaction with Client Proteins. *Nat. Struct. Mol. Biol.* **2011**, *18*, 345–351.

(180) Mashaghi, A.; Bezrukavnikov, S.; Minde, D. P.; Wentink, A. S.; Kityk, R.; Zachmann-Brand, B.; Mayer, M. P.; Kramer, G.; Bukau, B.; Tans, S. J. Alternative Modes of Client Binding Enable Functional Plasticity of Hsp70. *Nature* **2016**, *539*, 448–451.

(181) Slepenkov, S. V.; Witt, S. N. Peptide-Induced Conformational Changes in the Molecular Chaperone DnaK. *Biochemistry* **1998**, *37*, 16749–16756.

(182) Iwahara, J.; Clore, G. M. Detecting Transient Intermediates in Macromolecular Binding by Paramagnetic NMR. *Nature* **2006**, *440*, 1227–1230.

(183) Ubbink, M. The Courtship of Proteins: Understanding the Encounter Complex. *FEBS Letters*. **2009**, *583*, 1060–1066.

(184) Schreiber, G.; Haran, G.; Zhou, H.-X. Fundamental Aspects of Protein-Protein Association Kinetics. *Chem. Rev.* **2009**, *109*, 839–860.

(185) Fawzi, N. L.; Doucleff, M.; Suh, J.-Y.; Clore, G. M. Mechanistic Details of a Protein-Protein Association Pathway Revealed by Paramagnetic Relaxation Enhancement Titration Measurements. *Proc. Natl. Acad. Sci.* **2010**, *107*, 1379–1384.

(186) Misteli, T. Protein Dynamics: For Nuclear Architecture and Gene Expression. *Science (80-.).* **2001**, *291*, 843–847.

(187) Iwahara, J.; Clore, G. M. Direct Observation of Enhanced Translocation of a Homeodomain between DNA Cognate Sites by NMR Exchange Spectroscopy. *J. Am. Chem. Soc.* **2006**, *128*, 404–405.

(188) Hansen, D. F.; Hass, M. A. S.; Christensen, H. M.; Ulstrup, J.; Led, J. J. Detection of Short-Lived Transient Protein-Protein Interactions by Intermolecular Nuclear Paramagnetic Relaxation: Plastocyanin from Anabaena Variabilis. *J. Am. Chem. Soc.* **2003**, *125*, 6858–6859.

(189) Volkov, A. N.; Worrall, J. A. R.; Holtzmann, E.; Ubbink, M. Solution Structure and Dynamics of the Complex between Cytochrome c and Cytochrome c Peroxidase Determined by Paramagnetic NMR. *Proc. Natl. Acad. Sci.* **2006**, *103*, 18945–18950.

(190) Bashir, Q.; Scanu, S.; Ubbink, M. Dynamics in Electron Transfer Protein Complexes. *FEBS J.* **2011**, *278*, 1391–1400.

(191) Demott, C. M.; Majumder, S.; Burz, D. S.; Reverdatto, S.; Shekhtman, A. Ribosome Mediated Quinary Interactions Modulate In-Cell Protein Activities. *Biochemistry* **2017**, *56*, 4117–4126.

(192) Davidi, D.; Noor, E.; Liebermeister, W.; Bar-Even, A.; Flamholz, A.; Tummler, K.; Barenholz, U.; Goldenfeld, M.; Shlomi, T.; Milo, R. Global Characterization of in Vivo Enzyme Catalytic Rates and Their Correspondence to in Vitro k Cat Measurements . *Proc. Natl. Acad. Sci.* **2016**, *113*, 3401–3406.

(193) Breindel, L.; DeMott, C.; Burz, D. S.; Shekhtman, A. Real-Time In-Cell Nuclear Magnetic Resonance: Ribosome-Targeted Antibiotics Modulate Quinary Protein Interactions. *Biochemistry* **2018**, *57*, 540–546.

(194) 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.* **2006**, *103*, 11904–11909.

(195) Bertrand, K.; Reverdatto, S.; Burz, D. S.; Zitomer, R.; Shekhtman, A. Structure of Proteins in Eukaryotic Compartments. *J. Am. Chem. Soc.* **2012**, *134*, 12798–12806.

(196) Sakai, T.; Tochio, H.; Inomata, K.; Sasaki, Y.; Tenno, T.; Tanaka, T.; Kokubo, T.; Hiroaki, H.; Shirakawa, M. Fluoroscopic Assessment of Protein Leakage during Xenopus Oocytes In-Cell NMR Experiment by Co-Injected EGFP. *Anal. Biochem.* **2007**, *371*, 247–249.

(197) Simsek, D.; Tiu, G. C.; Flynn, R. A.; Byeon, G. W.; Leppek, K.; Xu, A. F.; Chang, H. Y.; Barna, M. The Mammalian Ribo-Interactome Reveals Ribosome Functional Diversity and Heterogeneity. *Cell* **2017**, *169*, 1051-1065.e18.

(198) Castello, A.; Fischer, B.; Eichelbaum, K.; Horos, R.; Beckmann, B. M.; Strein, C.; Davey, N. E.; Humphreys, D. T.; Preiss, T.; Steinmetz, L. M.; et al. Insights into RNA Biology from an Atlas of Mammalian MRNA-Binding Proteins. *Cell* **2012**, *149*, 1393–1406.

(199) Majumder, S.; DeMott, C. M.; Reverdatto, S.; Burz, D. S.; Shekhtman, A. Total Cellular RNA Modulates Protein Activity. *Biochemistry* **2016**, *55*, 4568–4573.

(200) Narayanaswamy, R.; Levy, M.; Tsechansky, M.; Stovall, G. M.; O'Connell, J. D.; Mirrielees, J.; Ellington, A. D.; Marcotte, E. M. Widespread Reorganization of Metabolic Enzymes into Reversible Assemblies upon Nutrient Starvation. *Proc. Natl. Acad. Sci.* **2009**, *106*, 10147–10152.

(201) Gancedo, J. M. Yeast Carbon Catabolite Repression. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 334–361.

(202) Muntener, T.; Haussinger, D.; Selenko, P.; Theillet, F. X. In-Cell Protein Structures from 2D NMR Experiments. J. Phys. Chem. Lett. **2016**, *7*, 2821-2825.

(203) Sakai, T.; Tochio, H.; Tenno, T.; Ito, Y.; Kokubo, T.; Hiroaki, H.; Shirakawa, M. In-Cell NMR Spectroscopy of Proteins inside Xenopus Laevis Oocytes. *J. Biomol. NMR* **2006**, *36*, 179–188.

(204) Hancock, R. Internal Organisation of the Nucleus: Assembly of Compartments by Macromolecular Crowding and the Nuclear Matrix Model. *Biology of the Cell*. **2004,** *96*, 595–601.

(205) Spitzer, J.; Pielak, G. J.; Poolman, B. Emergence of Life: Physical Chemistry Changes the Paradigm. *Biology Direct*. **2015**, *10*, 33.

(206) Adkar, B. V.; Manhart, M.; Bhattacharyya, S.; Tian, J.; Musharbash, M.; Shakhnovich, E. I. Optimization of Lag Phase Shapes the Evolution of a Bacterial Enzyme. *Nat. Ecol. Evol.* **2017**, *1*, UNSP0149.

(207) Bershtein, S.; Mu, W.; Serohijos, A. W. R.; Zhou, J.; Shakhnovich, E. I. Protein Quality Control Acts on Folding Intermediates to Shape the Effects of Mutations on Organismal Fitness. *Mol. Cell* **2013**, *49*, 133–144.

(208) Dickerson, R. E.; Takano, T.; Eisenberg, D.; Kallai, O. B.; Samson, L.; Cooper, A.; Margoliash, E. Ferricytochrome c A General Features of the Horse and Bonito Proteins at 2.8 Å Resolution. *J. Biol. Chem.* **1971**, *246*, 1511–1535.

(209) Dickerson, R. E. The Structure of Cytochrome c and the Rates of Molecular Evolution. *J. Mol. Evol.* **1971**, *1*, 26–45.

(210) Çetinbaş, M.; Shakhnovich, E. I. Catalysis of Protein Folding by Chaperones Accelerates Evolutionary Dynamics in Adapting Cell Populations. *PLoS Comput. Biol.* **2013**, *9*, e1003269.

(211) Verma, A. K.; Diwan, D.; Raut, S.; Dobriyal, N.; Brown, R. E.; Gowda, V.; Hines, J. K.; Sahi, C. Evolutionary Conservation and Emerging Functional Diversity of the Cytosolic Hsp70:J Protein Chaperone Network of Arabidopsis Thaliana. *G3-Genes Genomes Genet.* **2017**, *7*, 1941–1954.

(212) Sørensen, J. G.; Kristensen, T. N.; Loeschcke, V. The Evolutionary and Ecological Role of Heat Shock Proteins. *Ecology Letters*. **2003**, *11*, 1025–1037.

(213) Dixit, P. D.; Maslov, S. Evolutionary Capacitance and Control of Protein Stability in Protein-Protein Interaction Networks. *PLoS Comput. Biol.* **2013**, *9*, e1003023.

(214) Schlesinger, M. J. Heat-Shock Proteins. *J. Biol. Chem.* **1990**, *265*, 12111–12114.

(215) Lindquist, S. The Heat-Shock Response. *Annu. Rev. Biochem.* **1986**, *55*, 1151–1191.

(216) Queitsch, C.; Sangstert, T. A.; Lindquist, S. Hsp90 as a Capacitor of Phenotypic Variation. *Nature* **2002**, *417*, 618–624.

(217) Rutherford, S. L.; Lindquist, S. Hsp90 as a Capacitor for Morphological Evolution. *Nature* **1998**, *396*, 336–342.

(218) Rutherford, S.; Hirate, Y.; Swalla, B. J. The Hsp90 Capacitor, Developmental Remodeling, and Evolution: The Robustness of Gene Networks and the Curious Evolvability of Metamorphosis. *Critical Reviews in Biochemistry and Molecular Biology*. **2007**, *42*, 355–372.

(219) Wade, M. J.; Johnson, N. A.; Jones, R.; Siguel, V.; McNaughton, M. Genetic Variation Segregating in Natural Populations of Tribolium Castaneum Affecting Traits Observed in Hybrids with T. Freemani. *Genetics* **1997**, *147*, 1235–1247.

(220) Lebesi, D. M.; Tzia, C. Effect of the Addition of Different Dietary Fiber and Edible Cereal Bran Sources on the Baking and Sensory Characteristics of Cupcakes. *Food Bioprocess Technol.* **2011**, *4*, 710–722.

(221) Chow, K.-C. Hsp70 (DnaK) – an Evolution Facilitator? *Trends in Genetics*. **2000**, *16*, 484–485.

(222) Gaucher, E. A.; Govindarajan, S.; Ganesh, O. K. Palaeotemperature Trend for Precambrian Life Inferred from Resurrected Proteins. *Nature* **2008**, *451*, 704–707.

(223) Gouy, M.; Chaussidon, M. Evolutionary Biology: Ancient Bacteria Liked It Hot. *Nature*. **2008**, *451*, 635–636.

(224) Sleep, N. H. The Hadean-Archaean Environment. *Cold Spring Harbor perspectives in biology*. **2010**, *2*, a002527.

(225) Drummond, D. A.; Bloom, J. D.; Adami, C.; Wilke, C. O.; Arnold, F. H. Why Highly Expressed Proteins Evolve Slowly. *Proc. Natl. Acad. Sci.* **2005**, *102*, 14338–14343.

(226) Zeldovich, K. B.; Chen, P.; Shakhnovich, E. I. Protein Stability Imposes Limits on Organism Complexity and Speed of Molecular Evolution. *Proc. Natl. Acad. Sci.* **2007**, *104*, 16152–16157.

(227) Monsellier, E.; Chiti, F. Prevention of Amyloid-like Aggregation as a Driving Force of Protein Evolution. *EMBO Reports*. **2007**, *8* 737–742.

(228) Patki, A. U.; Hausrath, A. C.; Cordes, M. H. J. High Polar Content of Long Buried Blocks of Sequence in Protein Domains Suggests Selection Against Amyloidogenic Non-Polar Sequences. *J. Mol. Biol.* **2006**, *362*, 800–809.

(229) Tartaglia, G. G.; Pechmann, S.; Dobson, C. M.; Vendruscolo, M. Life on the Edge: A Link between Gene Expression Levels and Aggregation Rates of Human Proteins. *Trends in Biochemical Sciences*. **2007**, *32*, 204–206.

(230) Troadec, J. P.; Gervois, A.; Oger, L. Statistics of Voronoi Cells of Slightly Perturbed Face-Centered Cubic and Hexagonal Close-Packed Lattices. *Europhys. Lett.* **1998**, *42*, 167–172.

(231) Fraser, H. B.; Hirsh, A. E.; Steinmetz, L. M.; Scharfe, C.; Feldman, M. W. Evolutionary Rate in the Protein Interaction Network. *Science* **2002**, *296*, 750–752.

(232) Givens, G. D.; Seidemann, M. F. *Middle Ear Measurements in a Difficult to Test Mentally Retarded Population. Mental Retardation* **1977**, *15*, 40-42.

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

(234) Zhang, J.; Maslov, S.; Shakhnovich, E. I. Constraints Imposed by Non-Functional Protein-Protein Interactions on Gene Expression and Proteome Size. *Mol. Syst. Biol.* **2008**, *4*, 210.

(235) Claverie, J. M. Gene Number: What If There Are Only 30,000 Human Genes? *Science*. **2001**, *291*, 1255–1257.

(236) Maslov, S.; Sneppen, K. Specificity and Stability in Topology of Protein Networks. *Science,* **2002**, *296*, 910–913.

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

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

(239) Keskin, O.; Gursoy, A.; Ma, B.; Nussinov, R. Principles of Protein–Protein Interactions: What Are the Preferred Ways For Proteins To Interact? *Chemical Reviews*. **2008,** 108 1225–1244.

(240) Janin, J.; Bahadur, R. P.; Chakrabarti, P. Protein-Protein Interaction and Quaternary Structure. *Quarterly Reviews of Biophysics*. **2008**, *41*, 133–180.

(241) Ishihama, Y.; Schmidt, T.; Rappsilber, J.; Mann, M.; Harlt, F. U.; Kerner, M. J.; Frishman, D. Protein Abundance Profiling of the Escherichia Coli Cytosol. *BMC Genomics* **2008**, *9*, 102.

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

(243) Yang, J.-R.; Liao, B.-Y.; Zhuang, S.-M.; Zhang, J. Protein Misinteraction Avoidance Causes Highly Expressed Proteins to Evolve Slowly. *Proc. Natl. Acad. Sci.* **2012**, *109*, E831–E840.

(244) Zhang, J.; Yang, J. R. Determinants of the Rate of Protein Sequence Evolution. *Nat. Rev. Genet.* **2015**, *16*, 409–420.

(245) Forsdyke, D. R. The Heat-shock Response and intracellular self/not-self discrimination. *Immunology* **2005**, *114*, 142-143.

(246) Harms, M. J.; Thornton, J. W. Evolutionary Biochemistry: Revealing the Historical and Physical Causes of Protein Properties. *Nature Reviews Genetics*. **2013**, *14*, 559–571.

(247) Mirny, L.; Shakhnovich, E. Evolutionary Conservation of the Folding Nucleus. *J. Mol. Biol.* **2001**, *308*, 123–129.

(248) Tokuriki, N.; Stricher, F.; Schymkowitz, J.; Serrano, L.; Tawfik, D. S. The Stability Effects of Protein Mutations Appear to Be Universally Distributed. *J. Mol. Biol.* **2007**, *369*, 1318–1332.

(249) Pazos, F.; Helmer-Citterich, M.; Ausiello, G.; Valencia, A. Correlated Mutations Contain Information about Protein - Protein Interaction. *J. Mol. Biol.* **1997**, *271*, 511–523.

(250) Ferńndez, A.; Lynch, M. Non-Adaptive Origins of Interactome Complexity. *Nature* **2011**, *474*, 502–506.

(251) Lynch, M. The Evolution of Multimeric Protein Assemblages. *Mol. Biol. Evol.* **2012**, *29*, 1353–1366.

(252) Khersonsky, O.; Roodveldt, C.; Tawfik, D. S. Enzyme Promiscuity: Evolutionary and Mechanistic Aspects. *Current Opinion in Chemical Biology*. **2006**, *10*, 498–508.

(253) Glasner, M. E.; Gerlt, J. a; Babbitt, P. C. Mechanisms of Protein Evolution and Their Application to Protein Engineering. *Adv Enzym. Relat Areas Mol Biol* **2007**, *360*, 228–250.

(254) Serohijos, A. W. R.; Rimas, Z.; Shakhnovich, E. I. Protein Biophysics Explains Why Highly Abundant Proteins Evolve Slowly. *Cell Rep.* **2012**, *2*, 249–256.

(255) Wylie, C. S.; Shakhnovich, E. I. A Biophysical Protein Folding Model Accounts for Most Mutational Fitness Effects in Viruses. *Proc. Natl. Acad. Sci.* **2011**, *108*, 9916–9921.

(256) Serohijos, A. W. R.; Shakhnovich, E. I. Merging Molecular Mechanism and Evolution: Theory and Computation at the Interface of Biophysics and Evolutionary Population Genetics. *Current Opinion in Structural Biology*. **2014**, *26*, 84–91.

(257) Bershtein, S.; Serohijos, A. W.; Shakhnovich, E. I. Bridging the Physical Scales in Evolutionary Biology: From Protein Sequence Space to Fitness of Organisms and Populations. *Current Opinion in Structural Biology*. **2017**, *42*, 31–40.

(258) Mackay, T. F. C. Epistasis and Quantitative Traits: Using Model Organisms to Study Gene-Gene Interactions. *Nat. Rev. Genet.* **2014**, *15*, 22–33.

(259) Joy, J. B.; Liang, R. H.; McCloskey, R. M.; Nguyen, T.; Poon, A. F. Y. Ancestral Reconstruction. *PLOS Comput. Biol.* **2016**, *12*, e1004763.

(260) Bateson, W.; Mendel, G. *Mendel's Principles of Heredity, by W. Bateson.*; 2011.

(261) Sailer, Z. R.; Harms, M. J. High-Order Epistasis Shapes Evolutionary Trajectories. *PLoS Comput. Biol.* **2017**, *13*, e1005541.

(262) Shah, P.; McCandlish, D. M.; Plotkin, J. B. Historical Contingency and Entrenchment in Protein Evolution under Purifying Selection. *Proc. Natl. Acad. Sci. USA.* **2014**, *112*, E3226-35.

(263) Blount, Z. D.; Borland, C. Z.; Lenski, R. E. Historical Contingency and the Evolution of a Key Innovation in an Experimental Population of Escherichia Coli. *Proc. Natl. Acad. Sci.* **2008**, *105*, 7899–7906.

(264) Weinreich, D. M.; Lan, Y.; Wylie, C. S.; Heckendorn, R. B. Should Evolutionary Geneticists Worry about Higher-Order Epistasis? *Current Opinion in Genetics and Development*. **2013**, *23*, 700–707.

(265) Sailer, Z. R.; Harms, M. J. Molecular Ensembles Make Evolution Unpredictable. *Proc. Natl. Acad. Sci.* **2017**, *114*, 11938–11943.

(266) Sailer, Z. R.; Harms, M. J. Detecting High-Order Epistasis in Nonlinear Genotype-Phenotype Maps. *Genetics* **2017**, *205*, 1079–1088.

(267) Pollock, D. D.; Thiltgen, G.; Goldstein, R. A. Amino Acid Coevolution Induces an Evolutionary Stokes Shift. *Proc. Natl. Acad. Sci.* **2012**, *109*, E1352–E1359.

(268) Harms, M. J.; Thornton, J. W. Historical Contingency and Its Biophysical Basis in Glucocorticoid Receptor Evolution. *Nature* **2014**, *512*, 203–207.

(269) Kryazhimskiy, S.; Rice, D. P.; Jerison, E. R.; Desai, M. M. Global Epistasis Makes Adaptation Predictable despite Sequence-Level Stochasticity. *Science* **2014**, *344*, 1519–1522.

(270) Sanger, F.; Thompson, E. O. P.; Kitai, R. The Amide Groups of Insulin. *Biochem. J.* **2015**, *59*, 509–518.

(271) Merkl, R.; Sterner, R. Ancestral Protein Reconstruction: Techniques and Applications. *Biological Chemistry*. **2016**, *397*, 1–21.

(272) Hall, B. G. Simple and Accurate Estimation of Ancestral Protein Sequences. *Proc. Natl. Acad. Sci.* **2006**, *103*, 5431–5436.

(273) Zhang, J.; Rosenberg, H. F. Complementary Advantageous Substitutions in the Evolution of an Antiviral RNase of Higher Primates. *Proc Natl Acad Sci U S A* **2002**, *99*, 5486–5491.

(274) Malcolm, B. A.; Wilson, K. P.; Matthews, B. W.; Kirsch, J. F.; Wilson, A. C. Ancestral Lysozymes Reconstructed, Neutrality Tested, and Thermostability Linked to Hydrocarbon Packing. *Nature* **1990**, *345*, 86–89.

(275) Chandrasekharan, U. M.; Sanker, S.; Glynias, M. J.; Karnik, S. S.; Husain, A. Angiotensin II-Forming Activity in a Reconstructed Ancestral Chymase. Science **1996**, *271*, 502–505.

(276) Thomson, J. M.; Gaucher, E. A.; Burgan, M. F.; De Kee, D. W.; Li, T.; Aris, J. P.; Benner, S. A. Resurrecting Ancestral Alcohol Dehydrogenases from Yeast. *Nat. Genet.* **2005**, *37*, 630–635.

(277) Pretorius, I. S. Tailoring Wine Yeast for the New Millennium: Novel Approaches to the Ancient Art of Winemaking. *Yeast*. **2000**, *16*, 675–729.

(278) Merkl, R.; Sterner, R. Reconstruction of Ancestral Enzymes. *Perspect. Sci.* **2016**, *9*, 17–23.

(279) Lim, S. A.; Hart, K. M.; Harms, M. J.; Marqusee, S. Evolutionary Trend toward Kinetic Stability in the Folding Trajectory of RNases H. *Proc. Natl. Acad. Sci.* **2016**, *113*, 13045–13050.

(280) Lim, S. A.; Bolin, E. R.; Marqusee, S. Tracing a Protein's Folding Pathway over Evolutionary Time Using Ancestral Sequence Reconstruction and Hydrogen Exchange. *Elife* **2018**, *7*, e38369.

(281) Lim, S. A.; Marqusee, S. The Burst-Phase Folding Intermediate of Ribonuclease H Changes Conformation over Evolutionary History. *Biopolymers* **2018**, *109*, e23086.

(282) Copley, S. D. An Evolutionary Biochemist's Perspective on Promiscuity. *Trends in Biochemical Sciences*. **2015**, *40*, 72–78.

(283) Nam, H.; Lewis, N. E.; Lerman, J. A.; Lee, D. H.; Chang, R. L.; Kim, D.; Palsson, B. O. Network Context and Selection in the Evolution to Enzyme Specificity. *Science* **2012**, *337*, 1101–1104.

(284) Almaas, E.; Kovács, B.; Vicsek, T.; Oltvai, Z. N.; Barabási, A. L. Global Organization of Metabolic Fluxes in the Bacterium Escherichia Coli. *Nature* **2004**, *427*, 839–843.