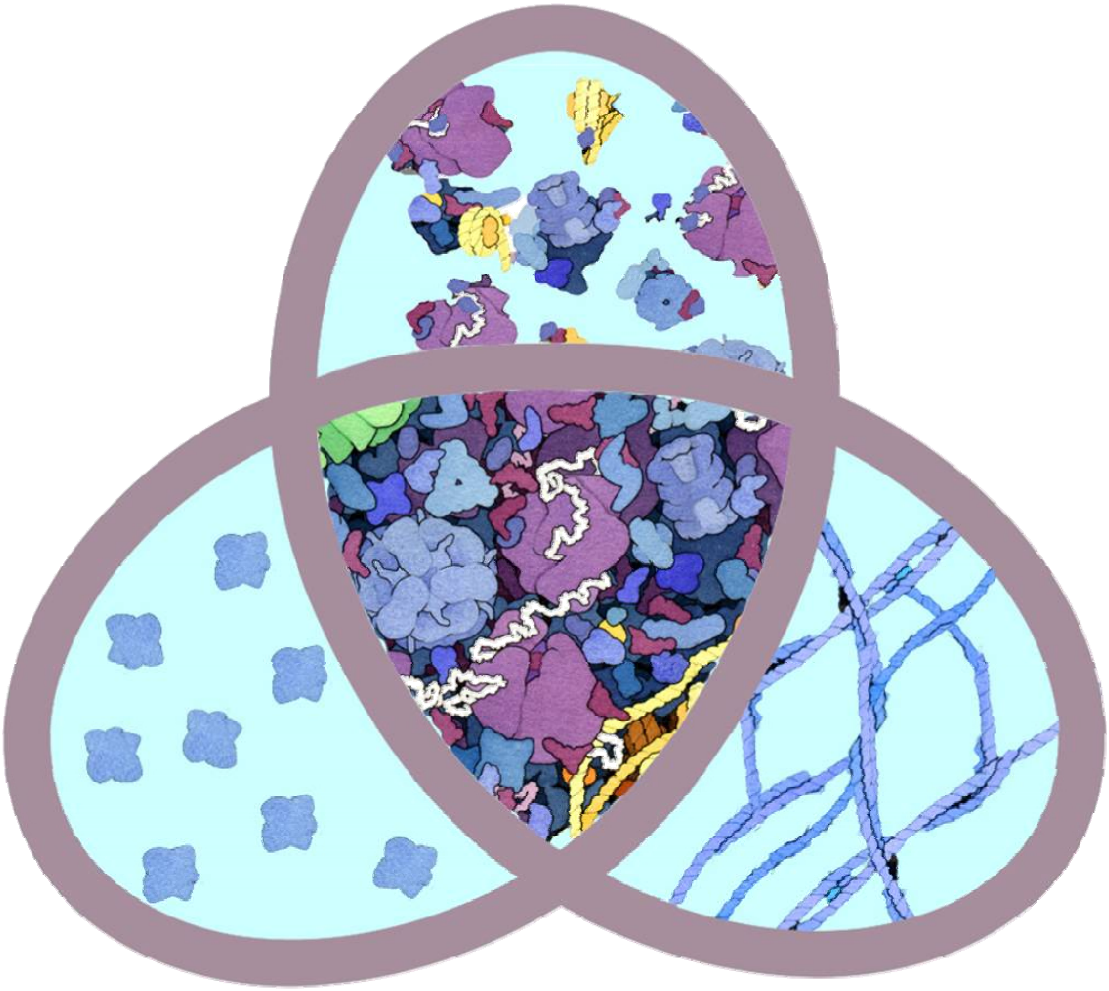
# Protein Interactions in Physiological Environments



**Ciara Kyne**

National University of Ireland, Galway

PhD, 2016


# Protein Interactions in Physiological Environments

PhD Thesis

This thesis was prepared at the School of Chemistry, National University of Ireland, Galway, from October 2011 to February 2016.

I declare that the work included in this thesis is my own work and has not been previously submitted for a degree to this or any other academic institution

---

Cover Art: Unravelling the Gordian knot of biochemistry
(based on the illustrations of David Goodsell, the Scripps Research Institute)

Ciara Kyne

Born in Galway 1989

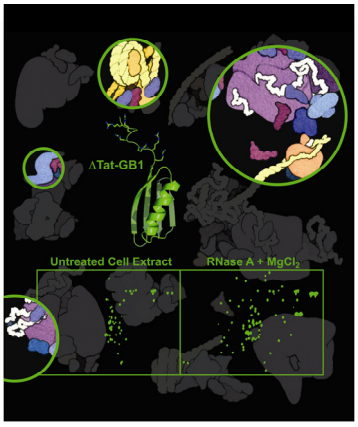
Supervisor: Dr Peter B. Crowley

External Examiner: Prof. Dr. Bert Poolman (University of Groningen)

Internal Examiner: Dr Andrew Flaus

Submitted: February 2016

**Summary of Contents**




The interactions of three proteins (ΔTat-GB1, cytochrome *c* and flavodoxin) were explored in *Escherichia coli* cells and extracts. The results indicate features of the physicochemical basis for protein assembly in the complex cytoplasm. Some of this work featured as a highlighted article in Protein Science.[1] Although key principles of molecular recognition have been gleaned from 'reductionist' studies performed in dilute solution, physiologically-relevant insights require protein characterisation under native or near-native conditions.[2] In particular, knowledge of the low-specificity 'quinary' interactions that dynamically organise the cytoplasm is scarce.

In-cell NMR indicated that all three proteins interact pervasively with the cytoplasm, despite their different sizes and surface properties. NMR, size exclusion chromatography (SEC) and/or native gel electrophoresis was used to probe the physicochemical basis for the interactions of each test protein in *E. coli* extracts.[1,3] The charge of cytochrome *c* and four of its mutants was determined by membrane confined electrophoresis. These findings indicate the effects of decreased charge on cytochrome *c* interactions in extracts. The weak, low-specificity cytochrome *c*-flavodoxin complex was studied in volume-occupied solutions to dissect the effects of the cytoplasmic architectures on this 'quinary-like' interaction.[3] The intricate interplay of charge, hydrophobic interactions and preferential hydration in governing cytoplasmic interactions is emphasised. These findings provide fresh evidence in support of a physicochemical model for cytoplasmic structuring.

A novel strategy for $^{19}$F tryptophan incorporation in *E. coli* was also developed which involved the addition of 30-60 mg/L of the tryptophan precursor 5-fluoroindole to the growth medium.[4] $^{19}$F tryptophan-labelled flavodoxin and GB1 were subsequently studied by $^{19}$F NMR in *E. coli*[4] or extracts.[3]

**References** (1) Kyne *et al. Protein Sci.* **2015**, 24, 310, (2) Kyne & Crowley, *FEBS J* **2016**, *In press*, (3) Kyne *et al.* (Manuscript in preparation) **2016**, (4) Crowley *et al. Chem. Commun.* **2012**, 48, 10681.

# Contents

# Introduction

## Composition of the *Escherichia coli* Cytoplasm

This thesis focuses on assessing protein interactions in physiological environments such as the *E. coli* cytoplasm. The composition of *E. coli* cells is relatively well-defined, making it a useful model for molecular biology, genetics, biotechnology and cellular physicochemistry.[1] The cytoplasm, which accommodates the majority of biochemical activity, is replete with small molecules (including ions and metabolites) and macromolecules (including proteins and nucleic acids). How does this complex mixture of interacting components capacitate the living cell? In this section the main components of the cell are introduced briefly. The cited papers and reviews provide more details for the interested reader.

The total inorganic ion concentration of the *E. coli* cytoplasm is 300 mM.[2] At a concentration of 100-250 mM, $K^+$ is the most abundant inorganic ion.[3,4] Considerable differences in the total intracellular $Mg^{2+}$ concentration have been reported (*e.g.* 20-100 mM).[5,6] Importantly, most intracellular $Mg^{2+}$ is associated with nucleic acids and, thus, the concentration of 'free' $Mg^{2+}$ is ~1 mM.[7] Other ions, such as $Cl^-$, $Na^+$, $Ca^{2+}$, and phosphates ($H_2PO_4^-/HPO_4^{2-}/PO_4^{3-}$) are present in low mM quantities. The total concentration of several transition metals (such as $Fe^{2+}$, $Zn^{2+}$, $Cu^+$ and $Mn^{2+}$) have been estimated at sub μM in *E. coli*.[8,9] Notably, the concentration of free transition metals is low due to their high affinities for metal-binding proteins and their poor water solubility at neutral pH.[5] The concentrations of over 100 metabolites from glucose-fed, exponentially growing *E. coli* have been determined.[3] The *E. coli* metabolome is dominated by a limited number of compound classes such as nucleotides, amino acids, and central carbon intermediates (*e.g.* glutamate, shikimate and citrate). The total metabolite concentration reaches up to 300 mM in the *E. coli* cytoplasm with approximately a third of this comprising glutamate.[3] Putrescine, glutathione, fructose 1,6 bisphosphate and adenosine triphosphate (ATP) are the next most abundant metabolites at intracellular concentrations of 30,[10] 17, 15 and 10 mM, respectively.[3] Importantly, ion and metabolite concentrations in the *E. coli* cytoplasm are governed by the extracellular environment and consequently tend to fluctuate.[4,11] For instance, the $K^+$ and glutamate concentration in *E. coli* cells increased from 0.23 to 0.93 M and from 0.03 to 0.26 M, respectively when the osmolality of the growth medium was increased from 0.1 to 1.1 molal.[4] Similarly, the pH of the *E. coli* cytoplasm is governed by the

extracellular environment and can range from 6.0-8.0.[12] Physiologically-relevant changes in the ionic strength and pH of the cytoplasm modulate its dielectric constant (and therefore the strength of charge-charge interactions) and conductivity.[5,13]

The *E. coli* cytoplasm is extremely crowded with macromolecules.[14,15] For instance, a single *E. coli* cell contains 2.4 million protein molecules which correspond to > 4,000 different proteins and a concentration of 200-320 g/L (volume of cell ~2 fL).[16,17] *E. coli* contain a single, circular 4.6 Mbp chromosome, which condenses upon binding to nucleoid associated proteins (NAPs) and by supercoiling to form the nucleoid architecture.[18] The concentration of DNA and RNA in *E. coli* is 11-18 and 75-120 g/L, respectively.[5,16] The high degree of volume occupancy (20-40%)[14,19] means that the average distance between macromolecules in the cytoplasm is on the same order of magnitude as their molecular diameters. Macromolecular crowding has at least two consequences. Firstly, the space occupied by macromolecules in the crowded cell interior cannot be accessed by other macromolecules.[14] Thus, macromolecular crowding imposes physical barriers to molecular diffusion,[20,21] leading to non-uniform macromolecular partitioning throughout the cytoplasm and the formation of microcompartments or microenvironments.[22] Moreover, steric repulsion increases the thermodynamic activity of biomolecules,[15] which is defined as the number of molecules of the solute per unit available volume. However, macromolecular crowding effects do not originate solely from repulsive forces, because the cell interior is chemically diverse and biomolecular surfaces are interactive. The abundant, heterogeneous surface area and high collision frequency ($10^9$-$10^{10}$ collisions/second)[23] of macromolecules *in vivo* promotes myriad, cooperative weak interactions. Moreover, crowding decreases the protein concentration required to induce phase separation.[24,25] Liquid-liquid phase separation has been demonstrated *in vivo* and the resulting dynamically responsive, membraneless bacterial microcompartments characterised (see below).[26] Thus, biomolecular interactions dynamically organise the cytoplasm to form local microenvironments which likely have unique viscosities, pH and dielectric properties.

## Molecular Organisation in Synthetic Cells: Biological Implications

Molecular self-organisation is a hallmark of living systems. Although the basic molecular composition of the *E. coli* cytoplasm is known, knowledge of the physicochemical mechanisms governing molecular organisation *in celluo* is scant. Studies of the physicochemical properties of the cell interior are often precluded by cellular complexity. Bottom-up approaches, which involve the *de novo* construction of synthetic cells from a few components,[27] offers a simplified route toward the systematic exploration of intracellular molecular organisation. Minimally, cells are defined as semi-permeable compartments. Compartmentalisation therefore always features as a fundamental element of synthetic cells.[28,29] In nature, phospholipids self-assemble to form bilayer membranes and, accordingly, liposomes (lipid bilayer membrane-enclosed vesicles) have been used extensively in the construction of synthetic cells. For instance, Miyata and Hotani encapsulated G-actin into 1:1 (wt:wt) dimyristoyl phosphatidyicholine/cardiolipin liposomes at 0 °C and initiated actin polymerisation by increasing the temperature to 30 °C.[30] Actin bundle formation conferred the otherwise flaccid liposomes with mechanical stiffness and novel morphologies (*e.g.* dumbbell or disk shapes).[30] Similar lipid vesicles have been used extensively for studies of biochemical function (*e.g.* exocytosis,[31] protein synthesis[32,33] and origin of life studies[29,34]). Comparatively fewer investigations have explored the fundamental physicochemical principles that govern molecular organisation in lipid vesicles. In 2005, Long *et al.* demonstrated the utility of bottom-up cell constructions for studies of molecular self-assembly by investigating the partitioning of polyethylene glycol (PEG) and dextran in synthetic cells.[35] The resulting solution phase-separated within narrow temperature ranges to form a dextran-rich phase (or microenvironment) in the centre of the liposome which was surrounded by the immiscible PEG-rich phase. This partitioning influenced the distribution of added proteins (*e.g.* streptavidin) or DNA depending on their affinities for dextran or labelled-PEG (*e.g.* biotinylated-PEG). Although partitioning of the neutral polymers occurred in the absence of lipid bilayer confinement over a timescale of hours to days, it took minutes in the lipid vesicle.[35] Moreover, in these synthetic cells, phase separation was induced by increasing the external osmotic pressure (with 0.25 M sucrose). The role of semi-permeable lipid bilayer encapsulation in orchestrating and accelerating intracellular chemistry was

demonstrated. Moreover, this study indicated that reversible phase separation could enable dynamic microcompartmentation *in cellulo*.[35]

A 'cell structure mimic' was later presented by Marguet and colleagues who encapsulated poly(trimethylene carbonate)-b-poly(L-glutamic acid) polymersomes into a giant, poly(butadiene)-b-poly(ethylene oxide) polymersome.[36] Although exhibiting Brownian motion in the giant, 20 μM vesicle, the diffusion of the encapsulated nano-sized polymersomes was ~ 7 times slower when the giant vesicle was loaded with 300 g/L dextran.[36] Moreover, the polymersomes were immobile in giant vesicles containing 2% w/v alginate (p$I$ = 5.4). Together these findings demonstrate that crowding and confinement were efficiently mimicked in this synthetic cell.[36] So far, these systems have been employed as scaffolds for studies of enzymatic reactions with a view to developing functionalised synthetic cells (nanoreactors).[37] However, they may also serve as useful systems for studies of the effects of key organisational features on the self-assembly of added biomolecules. Indeed, biosupramolecular assembly has been explored in related systems recently. Peters and co-workers investigated the pH-sensitive interaction between an oligohistidine-tagged fluorescent protein (td-Tomato) and a Nickel(II) nitrilotriacetic acid (Ni-NTA)-presenting lipid on the inner membrane of a semi-permeable giant vesicle.[38] Encapsulated alcohol dehydrogenase (ADH) was used to alter the luminal pH, upon addition of NAD(H) and the conversion of small molecule substrates (*i.e.* ketones or alcohols are reduced or oxidized by ADH, respectively which, in turn, increases or decreases the pH).[38] The assembly of td-Tomato and Ni-NTA was inhibited at low pH (< 5.0) but promoted at pH 7.0 and the assembly/disassembly could be cycled several times within a single study indicating its dynamic nature.[38] This system therefore provides an ideal template on which studies of molecular recognition in more complex intracellular environments can be built.

Although synthetic polymers were employed in the above examples of microcompartmentalised cells, phase separation-induced microcompartmentation can also be mediated by proteins.[39] For instance, the disordered elastin-like proteins (ELPs) containing repeat pentapeptide units VPGXG (X = any nonproline amino acid) can undergo a reversible phase transition, resulting in the formation of aqueous two-phase systems. Several stimuli induce phase transitioning *e.g.* temperature, ionic strength and pH.[40] Phase transitioning is dependent on ELP chain length,

concentration and 'guest' residue composition.[41] In 2009, Ge *et al.* demonstrated that GFP-ELP fusion proteins (where valine, alanine and glycine occupy the guest position) induce the formation of liquid-liquid phase transitions in *E. coli* and tobacco cells to produce aqueous microcompartments that exclude ribosomes and nucleic acids.[42] This finding indicated that aqueous two-phase transitioning can occur in the cytoplasm.

The physiological-relevance of phase transitions in the formation of proteinaceous membraneless compartments (*e.g.* nucleolus,[43] Cajal bodies[44] and stress granules[45]) has been demonstrated in the past decade by several independent studies.[46] Specific solutes (*e.g.* RNA and proteins) preferentially partition within the phase separated, micron-scale 'droplets' to generate a unique microenvironment with specific function.[47] Typically, low complexity, unstructured protein domains mediate the assembly of the highly dynamic droplets.[39,48] Due to their structural plasticity and multispecificity, intrinsically disordered regions (IDRs) are amenable to multifarious interactions. This suggests that various mechanisms may govern droplet formation. The role of electrostatics in droplet assembly has been identified,[24,49] which coincides with the fact that IDRs are enriched in charged residues.[50] Recently, Nott and co-workers studied human Ddx4-mediated droplet formation *in vitro* and in HeLa cells.[49] The droplets were readily reconstituted *in vitro* with internal dynamics and a thermal stability closely similar to that observed in live cells.[49] Moreover, a truncated Ddx4 variant comprising the unstructured N-terminus ($Ddx4^{N1}$) displayed similar phase behaviour to the wild type.[49] The transition temperature of $Ddx4^{N1}$ decreased with increasing ionic strength (*i.e.* by 25 ˚C when the ionic strength was increased by 100 mM).[49] This finding suggests that charge-charge interactions are important in mediating droplet formation. $Ddx4^{N1}$ is composed of 'clustered charge motifs' enriched in either cationic or anionic residues. Upon mutating $Ddx4^{N1}$ such that it retained all of the same residues (and a similar p*I*) but not its charge-dense clustering, the protein could no longer form droplets.[49] The significance of charge-clustering in $Ddx4^{N1}$-based droplet assembly *in vivo* is therefore apparent and is consistent with the role of compositionally biased charge regions in many regulatory proteins.[51] The prevalent role of attractive, multivalent charge-charge interactions between biopolyelectrolytes in droplet formation makes them reminiscent of complex coacervates, *i.e.* charge-mediated liquid-liquid phase separation.[25,52,53]

Lessons from *in vitro* studies of complex coacervates, which are relatively well-studied in polymer chemistry,[25,54] should therefore guide our exploration of protein-based liquid droplets *in vivo*. For instance, Semenov and co-workers noted that oppositely charged (polyglutamate/polylysine) polypeptide coacervates disassembled in the presence of trypsin.[55] This finding prompted Aumiller and Keating to explore other modes of enzymatic action, namely phosphorylation/dephosphorylation, on phase-forming biopolymer systems.[56] The RNA/arginine based peptide system did not phase separate after phosphorylation of a peptide serine but droplet formation was reobserved upon dephosphorylation.[56] These findings suggested a role for phosphatase and kinase activity in directing the reversible formation of liquid droplets. Moreover, several factors influenced the composition of the coacervates including: peptide sequence and/or charge, degree of multivalent interactions, peptide phosphorylation state, and RNA sequence. The sensitivity of the coacervates likely supports the dynamic, responsive nature of physiological droplets, which may reduce the probability of rare nucleation events leading to aggregation in the protein-rich droplets.[24]

Protein interactions leading to phase-separated microcompartmentation will become important experimental systems for exploring spatiotemporal molecular self-organisation in live cells. These systems will offer unique insights into the physicochemical nature of the cytoplasm and distinct microenvironments. However, studies of droplet-forming proteins are still in their infancy. Similarly, investigations of the physicochemical basis of biomacromolecular assembly in intact cells are rare. Transmembrane protein machines, by comparison, are routinely studied in proteoliposomes,[57] which are crude cell mimics. These machines often use structural reorganisations to effect their function *in vivo*. Knowledge of the interaction changes of such proteoliposome systems in response to altered luminal compositions may therefore provide useful insights into intracellular physicochemistry. For instance, some transmembrane proteins, such as the bacterial ATP-dependent glycine betaine transporter OpuA, contain a cytoplasmic module that activates transport in response to elevated luminal ionic strengths (*e.g.* > 200 mM). OpuA activation is not observed at increased concentrations of electroneutral osmolytes, confirming the role of charge-charge interactions in governing OpuA activity.[58–60] Additionally, the activation threshold of OpuA (*i.e.* ionic strength conducive to OpuA activation)

increases with increasing concentration of anionic lipids in the liposome.[58] This result indicates that charge-charge interactions between the lipid head groups and the cystathionine β-synthase (CBS) module are implicated in OpuA activation. Indeed, charge-charge interactions appear to regulate the activity of three well-studied bacterial transporters.[61] This finding led Poolman, Spitzer and Wood to model transporter systems as "on-off electrostatic switches", in which the conformation of the cytoplasmic transmembrane protein module is governed by charge-charge interactions with the membrane.[61] In the case of OpuA, cationic regions on the CBS surface engage in attractive charge interactions with anionic lipid head groups at low ionic strengths (*e.g.* 100 mM) and lock the system in an inactive configuration. At elevated ionic strengths (*e.g.* 260 mM) counter ion interactions with the membrane and the CBS module decrease their surface charge densities, disrupting the membrane-protein interaction. The CBS module then favours a 'relaxed' conformation which is linked to system activation. Transporter modules that contain anionic surface charge densities are expected to have an opposite 'switching logic'.[61] In this case the charge repulsion between the anionic membrane and module favours a relaxed conformation of the module corresponding to transporter inactivation. Increased ionic strength reduces the anionic surface charge densities of the module and lipid head groups allowing cations to mediate module-lipid interactions. These interactions lead to a locked conformation which facilitates transporter activation.

Notably, the activation of OpuA takes place over a narrow, physiological ionic strength range *i.e.* from 100-260 mM where the Debye-length shifts from ~10 to ~6 Å, respectively.[61] However, the Debye-length is calculated from the linear Poisson-Boltzmann equation which becomes inconsistent at higher electrostatic potentials.[61] Therefore, the authors also used the Dissociative Electrical Double Layer (DEDL) theory,[62] to calculate a second parameter, the co-ion exclusion boundary, which is the average distance from a charged surface over which co-ions are excluded. The sum of the Debye-length and the co-ion exclusion boundary gives a more accurate value for the distance from the charged surface over which electrostatic forces apply. Poolman, Spitzer and Wood found that the co-ion exclusion boundary for two anionic species decreases with increasing ionic strength and equals zero when $I$ = 260 mM.[61] This finding indicates that an anionic surface no longer electrostatically repels a second anionic group at $I$ = 260 mM and the two

surfaces can therefore approach each other. By comparison, at $I$ = 100 mM, the co-ion exclusion boundary increases the Debye-length by ~5 Å.[61] These calculations quantitatively support the experimental findings for OpuA indicating that physiologically-relevant ionic strength changes can facilitate the charge-charge screening-based activation mechanism of bacterial transporters.

**Intracellular Organisation: The Abundance of Macromolecular Clusters**

Investigations of the physicochemical underpinnings of bacterial osmosensor activation provoked the construction of an electrochemical model for cytoplasmic structuring.[63–65] Although the experimental data inspiring this model pertains to a single, functional class of transmembrane proteins, the co-ion exclusion boundary calculations are generally applicable. Moreover, the case for the significance of charge-charge interactions in governing macromolecular assembly in the cytoplasm is implied from the strongly Coulombic nature of the cell interior.[66] For example, 70% of *E. coli* proteins are acidic[64,67,68] and, along with the anionic groups of nucleic acid and lipid bilayers, generate negatively charged microenvironments throughout the cytoplasm.[63,66] Positively charged proteins and counter-ions are expected to partially neutralise the negative charge to form local, supercrowded 'clusters' which are characterised by a net negative charge (Figure 1).[63] Anionic cluster surfaces are surrounded by ionic atmospheres enriched in counter-ions (cations). Screening decreases the distance over which electrostatic forces apply (*i.e.* the Debye-length). However, the average intermolecular distances are similar to the Debye-length at physiological ionic strengths (~10 Å) making screened electrostatic interactions significant. Thus, neighbouring clusters electrostatically repel each other,[69] preventing erratic, aggregation and non-specific, physiologically-detrimental interactions.

Several lines of experimental evidence support the existence of clusters. Elowitz *et al.* used fluorescence recovery after photobleaching (FRAP) to determine that the apparent diffusion co-efficient of GFP in living *E. coli* cells was ~11 times less than that measured in water.[70] This finding indicated that diffusion in the cytoplasm is hampered by macromolecular crowding.[70] Interestingly, His-tagged GFP diffusion was reduced by up to 40% in *E. coli* compared to the WT protein.[70] This result suggests that pervasive interactions involving the GFP His-tag and other

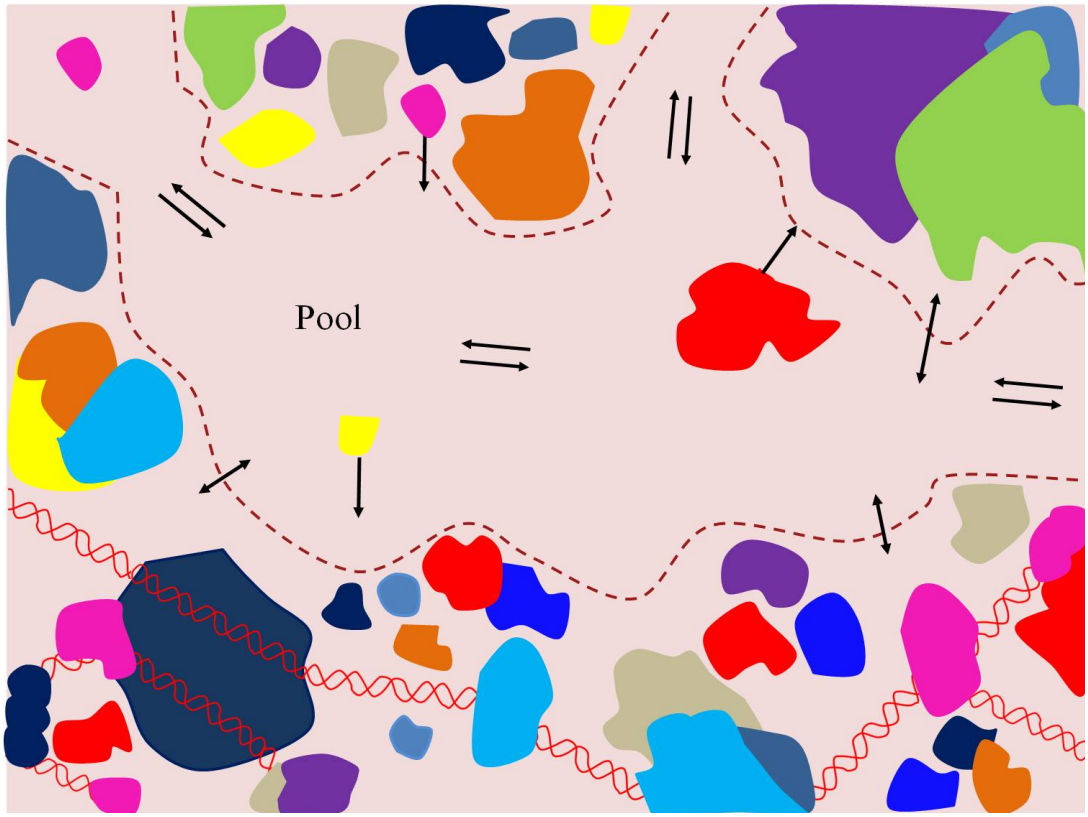



**Figure 1.** A millisecond snapshot of the cytoplasm at a region (~80 nm x 60 nm) peripheral to the nucleoid. Proteins, nucleic acids and their complexes are shown. The plasma membrane and cell wall are not depicted. Unequal, non-random crowding imposed by charge-charge interactions and steric repulsion leads to the formation of macromolecular clusters that interact to form functional regions *in vivo*. Macromolecular clusters structure the cytoplasm into a supercrowded cytogel and a dilute cytosol, the boundary between these zones is indicated by the dashed lines. Electrolytes and ions occupy the cytosolic 'pools' and pathways. Single-headed arrows show the diffusion of free macromolecules to/from the cytogels. Double-headed arrows show metabolite flows that arise due to enzymatic reactions. Ionic flows between pools are represented by equilibrium arrows. Figure adapted from Spitzer and Poolman's representation.[64]

macromolecules and/or clusters hamper the diffusion of the variant.[70] Pulsed-FRAP and confocal microscopy experiments by van den Bogaart and co-workers showed that GFP, when evenly distributed throughout the *E. coli* cytoplasm, had an apparent diffusion co-efficient of 3.2 $\mu m^2\ s^{-1}$ which represents the average from a broad distribution of values (0.1 to 24 $\mu m^2\ s^{-1}$).[20] This distribution indicated that GFP diffusion is likely dependent on its local microenvironment, and decreases due to interactions with other macromolecules. Pulsed-FRAP measurements demonstrated that the diffusion co-efficient of GFP decreased by a factor of 10 when the NaCl concentration of the cell resuspension buffer was increased by 250 mM, corresponding to an osmotic upshift of 0.42 Osm. This finding indicates that physiologically-relevant changes in macromolecular crowding can have dramatic effects on macromolecular interactions (*e.g.* steric repulsion) and thus, diffusion.[20]

The promise of NMR spectroscopy for the atomic-level assessment of a single test protein in living cells[71] is hampered by the inability to detect most globular proteins by standard heteronuclear single quantum correlation (HSQC) measurements. Several groups have independently identified that extensive weak interactions between test proteins and high molecular weight assemblies in the cell interior slow the rate of test protein tumbling causing unfavourable relaxation properties and severe $^{1}H$ line broadening.[72–78] This limitation provides striking evidence for the abundance of large assemblies (*i.e.* clusters) in cells. For instance, in 2011 Wang and co-workers compared the in-cell $^{1}H$, $^{15}N$ HSQC spectra of several globular proteins of similar size but varying surface properties.[73] Although all three proteins produce high quality NMR spectra *in vitro*, only GB1 (B1 domain of protein G; 6.2 kDa; p*I* = 4.3) and NmerA (N-terminal metal-binding domain of mercuric ion reductase; 6.9 kDa; p*I* = 7.2) were detectable in *E. coli*.[73] Ubiquitin (8.7 kDa; p*I* = 5.7) was invisible by in-cell NMR. Relative to their sharp *in vitro* spectra, significant resonance broadening was observed for both GB1 and NmerA in live cells.[73] Interestingly, GB1 gave rise to a higher quality in cell spectrum than NmerA. A GB1 dimer (12.5 kDa) containing a single residue linker had a rotational correlation time twice that of GB1 *in vitro* but yielded a sharper spectrum than NmerA in *E. coli*. This indicates that the difference in NmerA and GB1 size was not responsible for the substantial difference in the quality of the in-cell spectra. A GB1-$L_{15}$-NmerA fusion protein, in which each domain fold is identical to the native protein, was also detectable in *E. coli*. Notably, the peak intensities corresponding to NmerA residues were lower than those of GB1. Additionally, the peaks corresponding to NmerA backbone amides were weaker than those of side chain amides. The relaxation of side chain amide nuclei are governed mainly by their high degrees of internal motion and therefore are not as dramatically affected by decreased rates of protein tumbling. These findings confirm that decreased rotational correlation times cause the increased line broadening effected in cells. GB1 is 'biologically inert' *i.e.* does not interact with other macromolecules in *E. coli*[74] and GB1 line width increases in *E. coli* are likely affected mainly by intracellular viscosity. Thus, by comparing GB1 $^{1}H^{N}$ line widths as a function of glycerol concentration (*i.e.* sample viscosity) with those measured for GB1 in *E. coli*, the apparent intracellular viscosity was determined to be 8-11 times that of water. Calculations of the per-residue NmerA $^{1}H^{N}$ line widths expected at the intracellular viscosities underestimated the degree of

line broadening observed in the in-cell NmerA spectrum. This finding indicated that factors other than viscosity, namely intermolecular interactions, influence NmerA tumbling. Advances in sensitivity-enhancing NMR pulse sequences are now facilitating important breakthroughs for in-cell NMR. For example, Majumder and co-workers recently used $^{1}H,^{15}N$ cross-correlated relaxation-enhanced polarization transfer (CRINEPT) heteronuclear multiple-quantum coherence (HMQC) transverse relaxation-optimised spectroscopy (TROSY) experiments to study the weak interactions of 4 'sticky' proteins (thioredoxin, FKBP, adenylate kinase, ubiquitin) in live cells.[79] They demonstrated that protein-RNA interactions comprise a major component of cytoplasmic assemblies and found that each protein, although < 30 kDa in size, had an apparent molecular weight of > 1 MDa *in vivo*.[79] These data support the idea that macromolecules assemble into functional clusters *in vivo*. Moreover, chemical shift perturbation mapping[80] indicated that the known, specific binding surfaces of the test proteins were involved in their weak, pervasive or 'quinary'[81] interactions.[79] This finding suggests that trends in molecular recognition, mainly indentified from crystal structure analysis,[82–84] are borne out within clusters.

**The Bipartite Structure of the Cytoplasm**

Spitzer and Poolman suggest that macromolecular clustering leads to the formation of complementary dilute pools and their interconnecting channels (Figure 1). These uncrowded regions are reservoirs of electrolytes and metabolites such as $K^{+}$, ATP, ADP, glutamate, $H_2PO_4^{-}$, and $HPO_4^{2-}$. Freely diffusing protein and RNA molecules are likely found in pools and pathways, and thus, facilitate localisation by the 'diffusion and capture' mechanism (Figure 1).[85] Evidence for pool and pathway structures *in vivo* comes from pulsed-FRAP studies of fluorescently-labelled glucose in *E. coli* at different extracellular osmolalities.[21] The apparent diffusion co-efficient of glucose was 50 $\mu m^2\ s^{-1}$ under normal osmotic conditions and 4.3 $\mu m^2\ s^{-1}$ in cells osmotically upshifted with 2 M NaCl. Confocal microscopy confirmed that the fluorescently-labelled glucose explored the entire cytoplasmic space under osmotically stressed conditions.[21] By comparison, the diffusion co-efficient of GFP dropped to 0.001 $\mu m^2\ s^{-1}$ when the external osmolality was increased using 0.5 M NaCl and the distribution of GFP became uneven due to exclusion by the nucleoid region.[21] Thus, although macromolecular diffusion *in vivo* is hampered by the

presence of other biopolymers, especially under conditions of osmotic upshift, small molecules remain highly mobile. The topology of pools and pathways are defined by the charged surfaces of the macromolecular clusters and the plasma membrane (Figure 1). Importantly, each pool has a unique composition and electrochemical gradients therefore exist between pools, driving the flow of electrolytes via channels (Figure 1). The geometry and electrostatic potential of the cluster surfaces (*i.e.* pool and channel boundaries) also direct the transport of small molecules by means of steric and charge-charge repulsion or attraction.

The non-random structure of the cytoplasm into (i) cytogel (clusters) and (ii) cytosol (pools and pathways) phases facilitates the dynamic, responsive nature of intracellular chemistry which depends on variables such as growth stage and extracellular conditions such as temperature, water activity[86] ionic strength[21] and pH.[87] For instance, variations in cytoplasmic pH and ionic strengths will change the electrostatic potential of macromolecular surfaces leading to cluster rearrangements. Cohen *et al.* recently captured cluster rearrangements using in-cell NMR.[87] They found that a GB1 variant (GB1 K10H) could be detected by the $^{1}H$, $^{15}N$ HSQC experiment in *E. coli* cells when the intracellular pH was 7.5.[87] Upon decreasing the intracellular pH to 6.0 the HSQC spectrum of GB1 K10H deteriorated and the protein was undetectable in *E. coli* at pH 5.0.[87] Indeed, at lower pH values, the net charge of anionic proteins in *E. coli* (and GB1 K10H) decreases and charge-charge repulsions are likely diminished leading to more extensive macromolecular interactions. Cluster rearrangements are expected to alter pool and pathway dimensions and compositions. This intracellular reorganisation maintains cell viability in unfavourable environments *e.g.* pH extremes.

In conclusion, several lines of experimental evidence support Spitzer and Poolman's model of the bacterial cytoplasm. This comprehensive model therefore provides useful insights into the physicochemical basis for molecular assembly *in vivo*. The model also allows for the qualitative assessment of the physiological validity of data acquired under simplified cell-like conditions and acts as an important reference when designing experiments of increased intracellular-relevance.

## Scope of this Thesis

*In vivo*, molecular interactions beget function. Proteins, the cell's most abundant and diverse macromolecule, interact with a cohort of biomolecules to orchestrate cellular processes.[88] Knowledge of protein interactions and other physiological processes have been derived mostly from studies of isolated molecules in idealised solutions. However, fundamental protein properties *e.g.* charge, stability, diffusion and activity are influenced by the surrounding environment. Thus, the behaviour of proteins in complex, physiological milieus and simplified *in vitro* conditions can vary drastically. In the wake of post-reductionist biochemistry,[23,89] protein characterisations are increasingly performed under conditions that mimic important facets of the cell interior. This thesis explores protein interaction propensities under native and native-like conditions. The goal was to establish the fundamental physicochemical properties that govern protein interactions in conditions akin to the bacterial cytoplasm. This knowledge may advance the current understanding for cytoplasmic structuring and quinary interactions (Chapters 2, 3, 5, 6). A facile route for $^{19}$F-protein labeling was also developed by coaxing *E. coli* to incorporate a fluorinated amino acid precursor (Chapter 4). This method will extend the application of in-cell protein NMR. Revisionism is becoming increasingly important in the development of frameworks and 'dream experiments' relevant to in-cell physicochemistry.[90,91] Accordingly, in Chapter 1, I present a historical review outlining the landmark discoveries that have shaped our current perception of the organised cell interior. Useful model systems, experiments and philosophies are also suggested.

## Model Proteins

The interaction propensities of unrelated, single-domain proteins were studied under conditions akin to the *E. coli* cytoplasm. By using test proteins with different surface properties (Figure 2), general conclusions could be extracted regarding the origin of protein stickiness in complex environments (see Discussion). GB1 is the B1 domain of streptococcal protein G (Figure 2A).[92] GB1 and its mutants are model systems for studies of protein folding and design.[93–95] Owing to its high solubility and stability,[96] GB1 is a well-established solubility tag,[97] and can be heterologously expressed in *E. coli* to high concentrations (*i.e.* 100 mg/L culture). It is an αβ protein that contains a

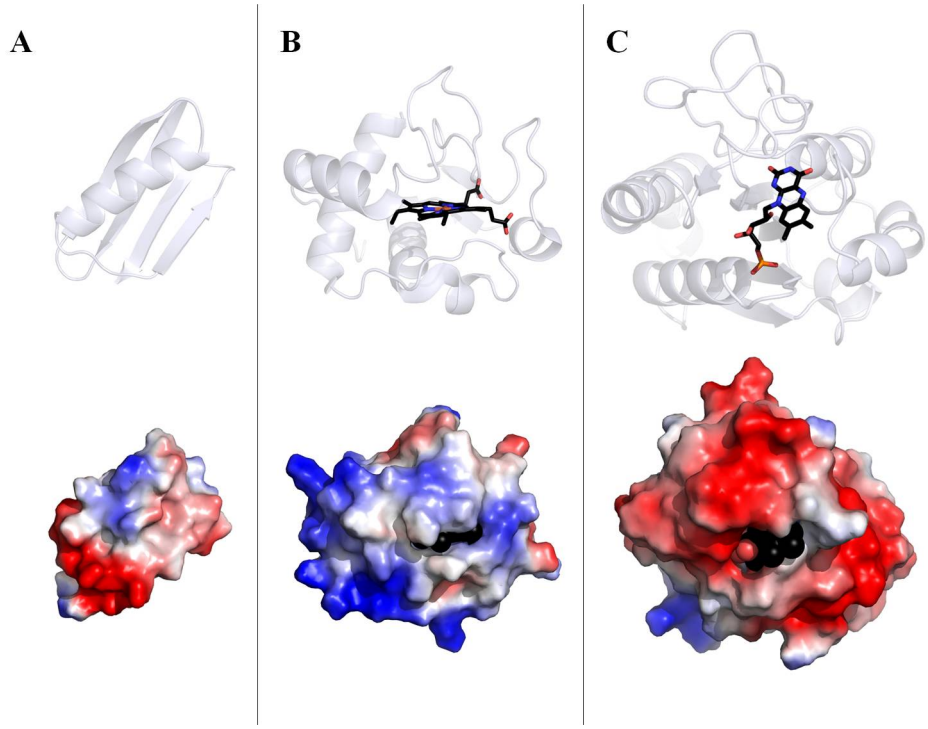



**Figure 2.** Cartoon and electrostatic surface representations of the model proteins **A** GB1; **B** cytochrome *c* and **C** flavodoxin. The haem and flavin mononucleotide cofactors of cytochrome *c* and flavodoxin are represented as black sticks or spheres.

four-stranded β-sheet and a four-turn α-helix.[98] GB1 contains 56 amino acids (6.23 kDa) and is anionic at neutral pH (p*I* = 4.3[96]). It is devoid of proline, serine, cysteine, histidine, arginine and prosthetic groups. Thus, its net charge is conferred by three protonatable residue types: aspartate (7); glutamate (3) and lysine (6). GB1 is an important model protein for *in vitro* and in-cell NMR studies.[73,74,96,99] It is considered a 'biologically inert' protein (*i.e.* unsticky surface),[74,99] possibly due to its small size and anionic nature. The GB1 mutant used herein (GB1-QDD) has the T2Q, N8D and N37D mutations, which prevents N-terminal methionine cleavage and deamidation across all pH values.[96] ΔTat-GB1 is an N-terminal fusion protein of GB1-QDD and the disordered, arginine-rich nuclear localisation signal of HIV-1 Tat (GRKKRRERRRA).[76,100] It is small (7.7 kDa) and has a theoretical p*I* of 8.2 and is undetectable by in-cell NMR spectroscopy,[76] indicating that it interacts extensively with cytoplasmic macromolecules.

Cytochrome *c*[101] is the most abundant protein in the mitochondrial intermembrane space (IMS) where it transfers electrons between complex III and IV of the respiratory chain to facilitate ATP synthesis.[102] It is also implicated in apoptosis where it oxidizes cardiolipin which leads to the release of proapoptotic

factors.[103] *Saccharomyces cerevisiae* cytochrome *c* is an all α protein composed of a series of α-helices and reverse turns (Figure 2B).[104] It contains 108 amino acids (12.6 kDa) and is lysine-rich (p*I* = 9.6[105]). Cytochrome *c* has 34 protonatable residues: aspartate (4); glutamate (7); lysine (16); arginine (3) and histidine (4). The majority of cytochrome *c*'s cationic residues are located in a large patch that surrounds the exposed haem edge (Figure 2B). This charged patch mediates cytochrome *c* interactions with small molecule (*e.g.* calixarene,[106] porphyrin[107,108]) and macromolecular[74,109,110] partners. As a model NMR and electron transfer protein, cytochrome *c* has been employed extensively in the NMR characterisation of transient protein-protein interactions.[74,111–114] Moreover, cytochrome *c*'s electron transfer capabilities and its ability to interact non-specifically with anionic molecules has led to its application in biosensor construction. For example, cytochrome *c* interactions with DNA,[109] bilirubin oxidase,[115] polyaniline sulfonic acid,[116] and *p*-sulfonatocalix[4]arene[117] have been used to generate electroactive assemblies on electrodes. Cytochrome *c* is not detectable by NMR in living *E. coli* cells or their extracts, owing to its pervasive, non-specific interactions with *E. coli* macromolecules.[74]

In *E. coli*, flavodoxin is an electron donor involved in the reductive activation of several enzymes including: pyruvate formate-lyase, anaerobic ribonucleotide reductase and $B_{12}$-dependent methionine synthase.[118] *E. coli* flavodoxin is an α/β protein containing five parallel beta strands flanked on both sides by helices (Rossman fold; Figure 2C).[118,119] It contains 175 amino acids (20.1 kDa) and is anionic at neutral pH (p*I* = 3.5[120]). Flavodoxin has 55 protonatable residues: aspartate (20); glutamate (17); lysine (6); arginine (4) and histidine (5). In flavodoxin, the non-covalently bound flavin mononucleotide (FMN) cofactor is enveloped by contiguous anionic and hydrophobic surface patches (Figure 2C). Chemical shift perturbation mapping experiments revealed that the FMN-containing face of flavodoxin mediates its interactions with partner proteins.[121] Given the abundance of proteins containing the ancestral α/β topology, flavodoxin is frequently employed in protein folding studies and, owing to its three-state folding mechanism, investigations of molten globules.[122–124] Notably, flavodoxin can fold in the absence of its FMN cofactor.[122] Thus, the apoflavodoxin-FMN interaction represents an important case study for protein-cofactor recognition.[125,126] Flavodoxin cannot be

detected by NMR in living *E. coli* cells (P.B. Crowley; personal communication) indicating that it is sticky, despite having a densely negatively charged surface expected to electrostatically repel the majority of *E. coli* macromolecules.

In this thesis, the NMR detectability of all three proteins (ΔTat-GB1, cytochrome *c* and flavodoxin) was tested inside living *E. coli* cells and cleared cell lysates (*i.e.* extracts). Extracts are crude mimics of the heterogeneous, complex cytoplasm[74,99,127] but can be manipulated to facilitate investigations of physiologically-relevant protein interaction propensities.[128] In Chapter 2 the interactions of ΔTat-GB1 in extract 'clusters' were investigated using a combination of NMR and size exclusion chromatography (SEC) studies. This system was used to address three central points: (i) what was the contribution of simple charge-charge interactions, (ii) did the arginine-rich motif interact preferentially with a specific class of macromolecule and (iii) could specific ions disrupt these ΔTat-GB1-containing assemblies?

The interactions of a second cationic protein, cytochrome *c*, were subsequently explored in *E. coli* extracts (Chapters 3 and 5). Previously, Crowley and co-workers used size exclusion chromatography (SEC) and mutagenesis experiments to demonstrate that cytochrome *c* formed high molecular weight complexes with *E. coli* macromolecules.[74] The role of simple charge-charge interactions in governing cytochrome *c* interactions was identified therein. This knowledge motivated a comparison of cytochrome *c* charge in dilute solution and in *E. coli* extracts at a physiological ionic strength (100 mM; Chapter 3). To this end, membrane confined electrophoresis (MCE)[129] and native gel electrophoresis was used. The MCE work was performed in collaboration with Ms. Kiara Jordan, Dr Dana Filoti and Prof. Thomas Laue at the University of New Hampshire. This work aimed to address the following questions (i) what is the effective charge of cytochrome *c* and charge inverted mutants (arginine/lysine→glutamate) at a physiological ionic strength, (ii) is cytochrome *c* charge modulated by different ions or (iii) sequestration by clusters in *E. coli* extracts (iv) can specific ions disrupt cytochrome *c* interactions with other macromolecules and (v) how does charge affect the 'stickiness' of the charge inverted cytochrome *c* mutants in *E. coli* extracts?

Native gel electrophoresis was also used to explore flavodoxin's net charge in (i) dilute solution, (ii) in the presence of cytochrome *c* and (iii) in extracts (Chapter 5). Intriguingly, in the presence of flavodoxin, cytochrome *c* charge is modified to an extent comparable to that observed in *E coli* extracts (Chapter 6). This led to the of the non-cognate, cytochrome *c*-flavodoxin interaction to represent a quinary[81] interaction. The effects of macromolecular crowding and gel confinement on the weak interaction were explored using the $^{1}H$, $^{15}N$ HSQC NMR experiment, which can discriminate between dynamic and permanent complexes (Chapter 6).

Recently $^{19}F$ NMR has emerged as a powerful method for studies of protein behaviour in complex, physiological environments owing to the fact that it is often amenable to HSQC-undetectable proteins.[130,131] Chapter 4 presents a novel strategy for protein labeling with fluorotryptophan in *E. coli* BL21, the organism of choice for recombinant protein production. The work aimed to identify whether (i) 5-fluoroindole, a fluorinated tryptophan precursor, could be incorporated into protein synthesis, (ii) whether inhibitors of aromatic amino acid biosynthesis were required, (iii) if this labeling strategy could be used to express proteins with a high tryptophan occurrence frequency or (iv) if this strategy was suited to cultures that constitutively express proteins.

# Chapter 1
# Grasping the Nature of the Cell Interior

The material in this chapter is from:

Kyne C, Crowley PB. Grasping the Nature of the Cell Interior: From *Physiological Chemistry* to *Chemical Biology*. *FEBS J.* **2015**, *In press*.

**Abstract**

Current models of the cell interior emphasise its crowded, chemically complex nature and dynamically organised structure. Although the chemical composition of cells is known, the cooperative intermolecular interactions that govern cell ultrastructure are poorly understood. A major goal of biochemistry is to capture these myriad interactions *in vivo*. We consider the landmark discoveries that have shaped this objective, starting from the vitalist framework established by early natural philosophers. Through this historical revisionism, we extract important lessons for the bioinspired chemists of today. Scientific specialisation tends to insulate seminal ideas and hamper the unification of paradigms across biology. Therefore, we call for interdisciplinary collaboration in grappling with the complex cell interior. Recent successes in integrative structural biology and chemical biology demonstrate the power of hybrid approaches. The future role of the (bio)chemist and model systems are also discussed as starting points for *in vivo* explorations.

## Introduction

Biochemistry is now resolutely focused on the exploration of biomolecules in physiological environments[23,90] and future advances require a comprehensive understanding of the emergent physicochemical properties of the cell interior. Remarkably, this grand challenge predates post-reductionist biochemistry, with musings of molecular organisation *in vivo* reported as early as 1917.[132] Herein, the roots of the current view of the cell interior are traced from the opposing vitalistic and mechanistic theories of early life scientists. Such an historic overview reveals how certain approaches have hindered (and continue to hinder) the discovery. For instance the significance of large, macromolecular machines and pervasive weak interactions were realised slowly, despite compelling evidence. Intensive scientific specialisation and a predominantly introspective focus apparently deterred the unification of biological frameworks. This contrasts sharply with exciting structural insights achieved recently from the harmonisation of experimental practices. We therefore emphasise the need for interdisciplinary collaboration and unified frameworks in the exploration of the complex cell interior. We describe useful model systems and future experiments for the biologically-inspired chemist exploring the molecular sociology of the cell.

## The Rise and Fall of Vitalism

**Organic Synthesis Redefines Chemistry** Natural philosophy defined living systems as possessing a mysterious, animating vital force. These vitalistic perceptions were challenged during the Scientific Revolution (1550–1700) when the increased rationality of modern science led to a new world view equating the universe to a machine.[133] Descartes conceived that organic bodies were complex mechanical systems that could be fully explained in terms of physics.[133] A quote from the 17th century physician Malphigi captures the mechanistic view:

> ‘Nature, in order to carry out the marvellous operations in animals and plants, has been pleased to construct their organised bodies with a very large number of machines, which are of necessity made up of extremely minute parts so shaped and situated, such as to form a marvellous organ, the composition of which are usually invisible to the naked eye, without the aid of the microscope.’[134]

The notion of organic machines was readily adopted in the field of anatomy, which explored the morphological structure and mechanical nature of the human body. However, failure to apply the machine concept to phenomena, such as development and reproduction, supported the notion of irreducible complexity in living systems. From the late 18th to the early 20th centuries, life was believed to arise from a 'vital force' that could not be understood in terms of fundamental physical laws.[91,135–137] Inaccessible to scientists, the vital force was believed necessary to synthesise organic compounds and to orchestrate biological processes. However, in the 1820s Wöhler prepared oxalic acid and urea from inorganic starting materials, indicating that the vital force was not critical for the synthesis of organic compounds.[137] Together with Kolbe's synthesis of acetic acid using coal in 1845,[138] a new definition of organic compounds emerged along with distinctions between organic chemistry and physiology.[136,137] Organic chemistry pertained to the study of compounds derived from nature while physiology involved the study of tissues. This compartmentalisation separated chemistry from studies of the living world.

**The Chemical Underpinnings of Vital Activity** With chemical endeavours redirected toward the development of purely chemical theories, such as elemental composition, isomerism and reactivity; the notion of vitalism persisted as an explanation for physiological phenomena.[136] The microscopic visualisation of cells led to the emergence of cell theory,[139,140] the discovery of cellular substructures and eventually questions on the nature of cellular fluids. For instance, in 1831 Brown identified the nucleus as an essential component of cells (Figure 1, top).[140] In 1844 von Mohl used the term 'protoplasm' to describe the active, fluid component of the cell.[140] The protoplasm was considered to be a reservoir of stored energy from which 'shapely forms' are constructed.[141] Around 1850, the protoplasmic theory of life emerged to postulate that all vital activities operate in the protoplasm.[142] Accordingly, the physical and chemical origins of life were sought through protoplasmic investigations. Studies by Hofmeister in 1867 further characterised the protoplasm as a viscous aqueous solution '…possessing in a remarkable degree the properties of a colloid. It is a mixture of different organic matters, among which albuminoids and members of the dextrine group are always present'.[141] In Goodale's 1889 review, the protoplasm was defined as a complex, fibrous mesh, containing 'numerous cognate proteids'.[141] While the protoplasm was defined as the medium in

which cellular activities are performed, aspects of the chemical nature of the protoplasm were identified and a physicochemical basis for vital activity was inferred. The field of colloid chemistry emerged with a view to advancing the knowledge of physiological processes.[132,143,144] However, the lack of basic chemical concepts (*e.g.* covalent bonding, electronegativity, pH) coupled with the absence of structural knowledge meant that the application of colloidal science to the protoplasm was slow to develop.[132] Thus, the metaphysical notion of vitalism endured, with declining levels of approval, until the 1920's.[145]

Crucial evidence against vitalism emerged in 1897 when Buchner identified zymase activity in yeast extracts.[144] The chemical activity of non-living, disrupted cells proved that a vital force did not govern life. This finding rendered cellular processes susceptible to scientific investigation. In 1912, Loeb dismissed vitalism as an anthropomorphic interpretation of life 'characteristic of all explanations of nature in the pre-scientific period' which survived due to the 'the interest of the layman (in questions of life and death) to a greater extent than possibly any other purely theoretical problem'.[146] Loeb consolidated several embryological studies to prove the physicochemical nature of egg activation by sperm in marine organisms *i.e.* the chemistry of the beginning of life.[146] Similarly, the *status quo* of protoplasmic understanding and vitalism is well-illustrated in Czapek's 1915 review of physiological phenomena in plants.[147] At this time, chemical analyses, by Reinke and Rodewald, revealed that proteins constitute 75% of the slime mould protoplasm while the remainder was composed of fatty acids, carbohydrates and inorganic compounds.[147] The lack of novel chemical components suggested that the 'peculiarities of the protoplasm' arose due to its integral, colloidal structure combined with its chemical nature. Such prescience was supported by efforts to dispel Reinke's claims that the vital properties of the protoplasm are lost upon cell death. The interplay of chemistry, 'correlation' and 'regulation' were deemed important in governing cellular processes.[147] A polemical rebuke of vitalism was presented by Troland in 1917 who found that 'more frequently than during the nineteenth century, men eminent in biology seem to quail before the complexity and delicacy of the life process, and, while uttering mechanistic truths about life, to offer them as sacrifices to a spirit of vagueness and discouragement'.[132] Inspired by the 'intellectual avalanche' effected by progress in theoretical physics, Troland

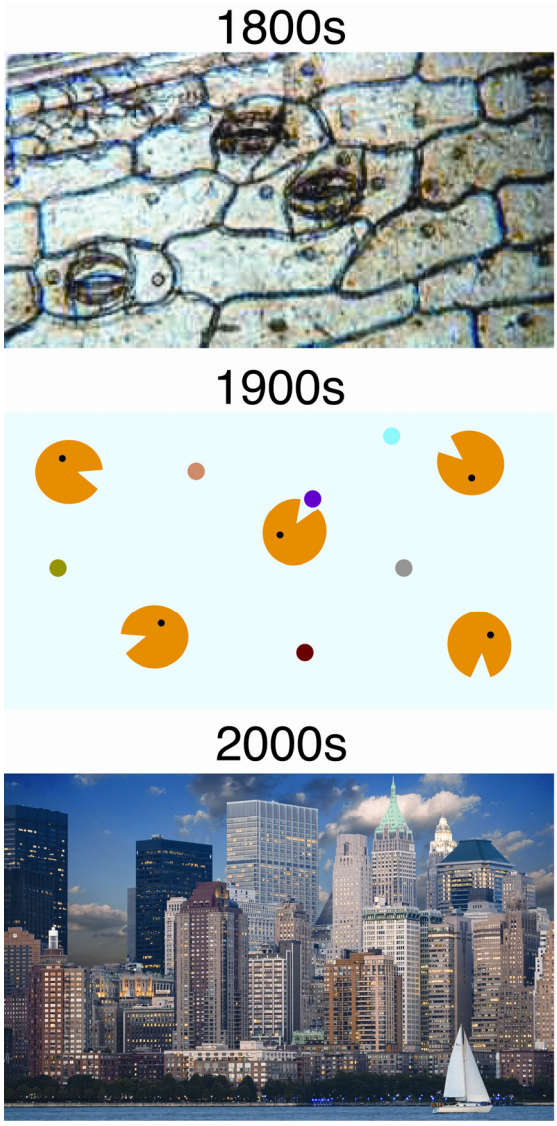



**Figure 1.** Changing perceptions of the cell interior: orchid epidermal cells viewed by microscope helped to establish cell theory (top; figure reproduced with permission from Professor Brian J. Ford);[140] the 'bag of enzymes' view of the cell prevailed for much of the 20th century; (middle), the cell interior is regarded today as a bustling metropolitan city (bottom).[148]

attempted to unite biological phenomena like chromatin 'reduplication' with physicochemical concepts such as surface heterocatalysis.[132] The 'enzyme theory of cell life' was thus established (see below). His call for new 'physico-chemical conceptions in terms of which a chaos of biological facts, already at hand, can be explained, or systematized' captured the increased consciousness of the need for chemisty in protoplasmic exploration.[132] Accordingly, the vitalist philosophy depreciated rapidly and the field of biochemistry advanced toward studies of enzymes.

### Opposing Views of the Cell Interior

**The Protoplasm: Bag of Enzymes or Organised Laboratory?** Although formerly thought to play a minor role in extracellular hydrolysis only, the crucial importance of enzymes in protoplasmic activity was established by Buchner's cell-free fermentation.[144,149] The putative role of enzymes in other physiological processes, such as germ-cell development,[150] gained momentum. The enzyme theory of cell life emerged to propose that life 'is fundamentally a product of catalytic laws acting in colloidal systems of matter' (Figure 1, middle).[141,147] Buchner's findings also demonstrated the possibility to study physiological processes outside cells. This subverted the protoplasmic theory of cell structure in which physiological processes were thought to occur only in the intact protoplasm. Indeed, the emerging view of the cell tended to disregard

protoplasmic singularity and suggested that biological reactions were governed by random encounters and simple, mass action relationships (*i.e.* the bag of enzymes view; Figure 1).[151] Meanwhile, several preeminent scientists proposed the drastically different notion that the molecular-level organisation of the cytoplasm supported cellular complexity. For example, upon establishing the enzyme theory of cell life Troland affirmed that 'it can hardly be doubted that the complex mechanisms of mitosis rests upon definite structural machinery, established by long periods of evolution'.[132] In 1919, Ritter suggested that the cell interior was replete with 'highly differentiated' molecules of varying molecular weight dispersed throughout co-existing phases.[152] The resulting 'metaplasmic constituents' acted simultaneously to generate 'special apparatus and an organised laboratory'.[152] Peters later elaborated on this view by suggesting that surface effects direct the formation of 'an organised network of protein molecules, forming a three dimensional mosaic extending throughout the cell'.[153] The mosaic was purported to constrain proteins, regulating their activity and allowing for rapid information transduction.

**Experimental Support for Molecular Organisation *In Vivo*** Several lines of experimental evidence later supported the theories of Troland, Ritter and Peters. For example, the importance of subcellular structure in maintaining the integrity of metabolic phenomena was demonstrated by Krebs.[154] In 1938, upon combining oil drops with lysed *Asterias* oocytes, Kopac noted the formation of a protein film at the oil-protoplasm interface (the oil droplet surface became crinkled, *i.e.* the Devaux effect).[155] A similar phenomenon was observed for oil droplets in aqueous protein solutions. However, in unlysed cells, the injected oil droplet remained spherical and the Devaux effect was never observed suggesting the lack of freely diffusing proteins in the intact protoplasm. Thus, protein confinement *in vivo* was demonstrated. Working around the same time, Cori *et al.* found that glycogen synthesis from glucose-1-phosphate could not be studied in crude muscle extracts due to a competing reaction which converted glucose-1-phosphate to glucose-6-phosphate.[156] This indicated the sensitivity of glycogen synthesis to cell disruption. Green isolated mitochondria in the late 1940s and identified the tricarboxylic acid cycle proteins within.[91] This finding suggested the ubiquity of protein confinement and co-localisation *in vivo*. Similarly, in the 1960s, whole cell centrifugation studies by Zalokar[157] and Kempner and Miller[158] demonstrated the reversible stratification of

the cell contents. Notably, the cytoplasm stratum was devoid of free proteins and ribonucleic acid which were instead found in the mitochondrial, ribosomal or nuclear strata. This finding suggested that macromolecules did not diffuse randomly in the cytoplasm. Kempner and Miller, like Zalokar, concluded that 'the entire macromolecular apparatus is associated with large particulates as a basic structure of cellular organisation'.[158] Moreover, the authors suggested the advantages of such organisation in reducing the dimensionality and directing substrate diffusion directly between metabolically-related enzymes. This supported Yanofsky's earlier studies of the two-enzyme tryptophan synthase system in which the intermediate, indole, was never detected free in solution during the reaction sequence.[159] In the 1960s and 1970s, Srere noted that the high apparent concentration of various enzymes *in vivo* was equal to or greater than the corresponding substrate concentration and thus Michaelis-Menten assumptions (*i.e.* excess of substrate) were not valid.[151,160] The growing knowledge of enzyme-effector interactions also suggested the complex, competitive nature of each step in a metabolic pathway, warranting studies of increased physiological relevance.[161] Srere postulated that enzymes assembled into multimeric complexes to facilitate the vectorial flow of substrates and/or intermediates along a metabolic pathway. Enzyme organisation was predicted to decrease the diffusion time of substrates and/or intermediates between pathway steps, to stabilise transient intermediates and to minimise cross-reactions with competing pathways.[162] The kinetic advantages of an immobilised system of malate dehydrogenase and citrate synthase over the free enzyme system was subsequently demonstrated.[163] In 1982, McConkey used O' Farrell gels to compare the molecular weight and isoelectric points of soluble protein mixtures extracted from CHO and HeLa cells.[81] Over 50% of the hamster cell proteins had human homologues. This result indicated the evolutionary tendency toward isoelectric point conservation suggesting that the rate of cellular protein evolution involving charged amino acid mutations are lower than expected.[81] McConkey predicted that most cellular proteins interact extensively and transiently with other macromolecules to produce a functional 'quinary' structure that constrains protein evolution by only accommodating mutations with no functional implications *i.e.* non-interacting sites.[81]

## The Reductionist Roots of Structural Biology

**Protein Crystallisation Engenders Structural Biology** Despite strong evidence to support the abundance of functional macromolecular assemblies *in vivo*, biochemistry developed mainly from the view that cells were membrane-enclosed bags of homogenous cytosol in which polymers and metabolites diffused randomly to interact with each other.[164] The lack of order in this 'random' cytoplasm led to the idea that it was unnecessary to study protein behaviour in the context of the intracellular environment. Buffered aqueous solutions became the preferred milieu for the characterisation of pure proteins. Thus, biochemistry tended toward reductionist studies focused on a single macromolecule in ideal solutions. For instance, in 1892 Osborne published protocols for protein purification from plant seeds.[165] Crystallisation, typically limited to haemoglobin and albumin, was used extensively by Osborne as a means of protein purification for further chemical analyses.[166] Additionally, the large number of proteins purified by Osborne disproved the notion that only a few distinct proteins existed in nature. Protein crystallisation facilitated seminal studies of protein behaviour. For example, in 1926 Sumner used crystallisation to purify jack bean urease.[167] The redissolved urease catalysed urea degradation, indicating that the catalytic activity of enzymes arises from the protein and not from co-purified organic compounds.[167] The first diffraction pattern from a protein crystal (pepsin) was reported by Crowfoot and Bernal in 1934[168] and the crystal structure of myoglobin[169] was solved thereafter. X-ray crystallography gave birth to the structural biology revolution, which focused on understanding macromolecular structure, function and interactions at the atomic level. The tens of thousands of macromolecular structures solved by X-ray crystallography[170] have revealed the roles of non-covalent interactions,[171] the hierarchical order of protein structure,[172] trends in molecular recognition[82,83,173] and insights into protein function.[174] In the 1950s Linderstrøm-Lang identified that amide hydrogens exchange constantly with solvent hydrogens.[175] Linderstrøm-Lang measured the rate of hydrogen-deuterium exchange for several proteins *in vitro* to explore the extent of hydrogen bonding within protein structures. The dependence of amide hydrogen exchange on conformational fluctuations of the protein was inferred, providing evidence for the dynamic nature of proteins.[175] The horse oxyhaemoglobin crystal structure refined by Perutz *et al.* contained regions devoid of

electron density, supporting further the notion of protein structural flexibility.[176] Similarly, significant conformational differences were observed in the early crystal structures for deoxy and oxyhaemoglobin, indicating the role of conformation change and dynamics in protein function.[176] Thus, solution-based studies of protein structure were required to reinforce the structural information derived from crystals and to explore the dynamic nature of proteins.

**Solution-State Studies and Supramolecular Machines** In the 1970s Wüthrich established two-dimensional nuclear magnetic resonance (NMR) spectroscopy as a powerful tool for solution studies of protein structure.[177] Early NMR structure determinations focused on the *Antennapedia* homeodomain (HD) and the operator DNA.[178] The importance of dynamic structural features in protein-DNA recognition was revealed, representing some of the earliest insights into the biological significance of disordered motifs.[179] However, studies of protein structural disorder (intrinsically disordered proteins) and the consequent transformative impact on the protein-structure function paradigm would not emerge for another 10-20 years.[5,179,180] Instead, NMR studies of protein-DNA interactions marked a surge toward the exploration of DNA-processing assemblies. Interest in such supramacromolecular assemblies arose from the newfound appreciation of their critical role in the orchestration of cellular processes. An early demonstration of the autonomous, modular and cooperative nature of functional assemblies or 'protein machines' was made by Alberts who presented the DNA replication apparatus (comprising > 40 proteins; Figure 2B) at a Cold Spring Harbour Symposium on Quantitative Biology.[181] Alberts urged biochemists to 'think of molecular genetics in terms of "protein machines" rather than in terms of sequential reactions that are carried out by individual proteins'.[181] This provocative insight emphasised the need for ordered, sophisticated machinery in complex living systems which resulted in 'freeing the biological process from the necessity of relying on random motions'.[181] In essence, Alberts' work helped re-establish a mechanistic view of the cell. The advantages of elaborate protein machines were implied. For example, the spatiotemporal clustering of functionally related macromolecules increases the probability and rate of productive encounters.[164] The enhanced specificity afforded by the cooperative interactions of multi-component protein machines resolved questions of how weak, moderately specific interactions (*e.g.* transcription factor-

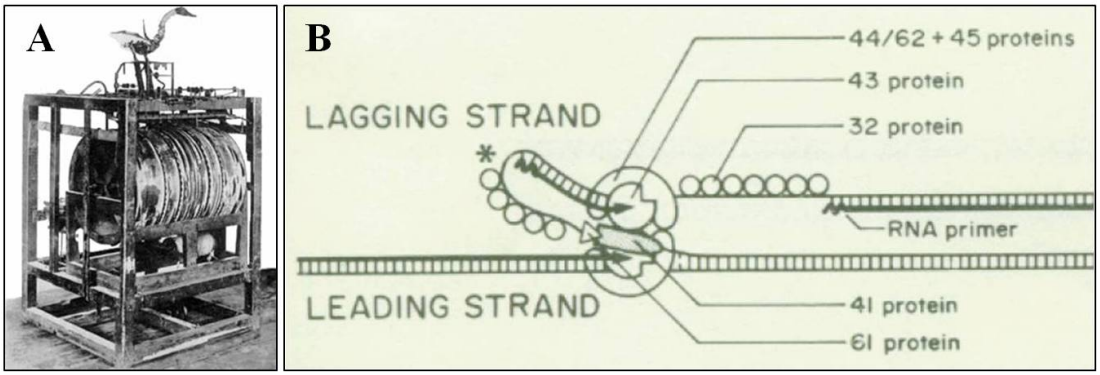



**Figure 2. Development of the 'biological machine' concept: A**, Vaucanson's mechanical duck (1739) was a symbol for the Cartesian machine. Built from hundreds of interacting parts, this automaton pretended to mimic digestion. **B**, Alberts' representation (1984) of a macromolecular machine, the T4 replication fork. Figure reproduced with permission from reference 181, copyright Cold Spring Harbour Laboratory Press.

DNA binding) became functional *in vivo*.[181] Conversely, the advantages of weak interactions for the regulation and reorganisation of protein machines were realised. Thus, although the importance of macromolecular machines was well-appreciated by 1990, an understanding of the mechanisms underlying their concerted function remained elusive.[164]

## Lessons from Macromolecular Machines: Shaping Future Experiments

**Integrated Approaches for the Interpretation of Macromolecular Machines** In the past 20 years, the ubiquity of macromolecular machines has been demonstrated. The advent of tandem affinity chromatography coupled with mass spectrometry has served to identify hundreds of protein machines in yeast[182] and other organisms.[183] Size exclusion chromatography[74,100,184] and fluorescence detected analytical ultracentrifugation studies[185] have also confirmed the abundance of high molecular weight assemblies in concentrated *E. coli* extracts. Fluorescence studies of protein and metabolite diffusion *in vivo* has demonstrated the sieve-like properties of the cytoplasm due to volume exclusion imparted by intracellular assemblies and microenvironments.[21,70] In-cell NMR spectroscopy has also indicated that high molecular weight assemblies pervade the cytoplasm. These assemblies hamper the detectability of most globular proteins by standard NMR experiments.[71,73,74,186] Several lines of experimental evidence supports the weak, responsive nature of these machines which cannot withstand even gentle lysis protocols.[81] For example, Benknovic and co-workers used fluorescence microscopy to identify cytoplasmic clusters containing enzymes involved in *de novo* purine biosynthesis in HeLa

cells.[187] Although difficult to reconstitute *in vitro*, the 'purinosome' is dynamically regulated *in vivo* by changes in cytoplasmic purine levels. Advances in cryo-electron microscopy (cryo-EM),[188] small angle X-ray scattering (SAXS),[189] fluorescence-based techniques,[190] single molecule microscopy[191] and electron tomography[192] have revealed, in exquisite detail, the structural and mechanical aspects of several large macromolecular machines such as the ribosome, spliceosome, proteasome, nucleosome and cytoskeletal filaments. Common mechanistic elements, such as NTP hydrolysis-coupled conformational changes, have also emerged.[193] Although mainstream structural biology methods have revealed snapshots of the interactions that glue specific macromolecular machines at distinct points in their operation, investigations of these interactions have yet to extend beyond phenomenological analyses. A major goal of biomacromolecular science now is to systematically interpret these myriad, time-evolving interactions to establish a hierarchy of attractive and repulsive physicochemical forces that shape cellular machinery. This knowledge will progress our understanding of biomolecular assembly *in vivo*. Exciting technological advances and the harmonisation of experimental practices promise to fulfil this objective. For instance, the recent inaugural wwPDB Hybrid/Integrative Methods Task Force workshop addressed the development, validation and curation of integrative structural models.[194] The standardisation of methods will foster the use of integrative approaches to explore macromolecular machines. Since a structural model can be generated from each experimental dataset, an ensemble of structural models can be developed. Accordingly, the accuracy and precision of the structural models can be assessed to facilitate an efficient, unbiased investigation of physiologically-relevant features of machine architectures.[194] Similarly, hybrid mass spectrometry experiments can be encoded into structural restraints for modelling analysis.[195] These models can be integrated with EM envelopes or crystal structures to enhance model accuracy by revealing regions of structural flexibility and transient complexes or intermediates.[195]

**Capturing Quinary Structure** Intracellular assemblies are not wholly represented by the permanent machinery characterised thus far. Indeed, 'molecular sociology' [192] in the cell is dominated by transient, quinary[81] interactions that link and extend machines to produce functional molecular communities or hyperstructures.[196,197] These architectures span nanometer to micrometer length-scales with lifetimes

ranging from nanoseconds to several minutes. Thus, quinary structure serves to bridge the molecular and cellular worlds. To understand quinary structure the emergent physicochemical properties of the cell interior must be identified. This cannot be accomplished by reductionist dissections. The grand challenge is to capture dynamic intracellular structures *in vivo* directly as they form and function. Indeed, the power of in-cell NMR spectroscopy for the characterisation of intracellular properties and global features of quinary structure has been demonstrated recently. For example, the Pielak group have illustrated the role of charge (and cytoplasmic pH) in governing quinary structure in *E. coli*.[87,198] Majumder and co-workers found that protein-RNA interactions comprise a major component of quinary structure.[79] Several systems have potential as useful models of quinary structure. For instance, membraneless organelles are intracellular phase-separated liquid droplets enriched in specific molecules.[46] Recently, Molliex *et al.* identified that weak interactions of the low complexity domain (LCD) of hnRNPA1 can orchestrate the dynamic assembly of cytoplasmic stress granules, a form of membraneless organelle.[24] The mechanism of granule assembly and cytoplasmic structuring appears to be linked to the LCD interaction propensities. Similarly, Nott and co-workers have identified that Ddx4-mediated droplet formation is dependent on the interactions of "clustered charge motifs" enriched in either cationic or anionic residues.[49] Thus, low-complexity systems which are critically important for the assembly of microscale structures *in vivo* are ideal candidates for the exploration of quinary structure.

Methods that illustrate the entire constellation of biomolecules *in vivo* will enhance our understanding of the "biological glue".[90] Notably cryo-EM[199] and cryo-electron tomography (ET)[200] have revealed aspects of vitrified cell ultrastructure. In particular, major advancements in specimen preparation and electron detection methods allow for the acquisition of 3D structural information on the macromolecular architecture of cells to nanometer resolution by cryo-ET.[201] For instance, Asano and co-workers recently used cryo-ET to compare the conformations of single- and double-capped 26S proteasomes in vitrified hippocampal neurons.[202] The authors identified that only 20% of proteasomes were engaged in substrate processing, which suggests that the capacity of the proteasome system is not fully exploited in healthy cells.[202] The ability to capture and compare the macromolecular

architectures comprising unique cellular states has led to the use of cryo-ET as a molecular diagnostics tool.[203] Despite the obvious promise of such techniques, further experimental progress is required to capture the co-operative molecular (re)arrangements that define functional regions in living cells.[90,197] These prospective techniques must reveal dynamic information about the emergent, biased molecular distributions that arise due to the integrated array of quinary interactions. To acquire this knowledge, hybrid methods will be necessary and the collaboration of physicists, cell biologists and (bio)chemists must be forged.[23,90,197] Mathematical analyses will be required to integrate changes in molecular distributions and interactions with specific cellular outputs (*e.g.* energy consumption, local pH changes). Advances in computational horsepower and algorithms showcase the utility of *in silico* methods for exploring macromolecular interactions leading to microenvironment formation and cytoplasmic physicochemistry. In 2010, McGuffee and Elcock used atomically detailed structures in the construction of a model containing 50 different types of *E. coli* macromolecules and GFP.[204] Intermolecular interactions were calculated on an atom-by-atom basis. Notably, the known translational and rotational diffusion coefficients for GFP could not be reproduced by modelling the electrostatic and steric interactions within the system.[204] Realistic GFP diffusion co-efficients were obtained only when short-range attractions between exposed hydrophobic groups were included in the model.[70,204] Similarly, by accounting for steric and hydrodynamic interactions, Hasnain's coarse-grained models of the *E. coli* cytoplasm accurately reproduced *in vivo* diffusion coefficients of GFP.[205] More recently an atomistic model of the *Mycoplasma genitalium* cytoplasm was constructed using physiologically-relevant concentrations of (homology-modelled) proteins, RNA, metabolites and ions.[206] This system represents a useful starting point for future simulations.

**The Role of the (bio)Chemist in Post-Reductionist Protein Science**

Though considered irrelevant to studies of life processes in the 1800s, chemistry continues to play a central role in the elucidation of physiological events today. Indeed, over the past two decades, chemistry has merged with cell biology, biochemistry and pharmacology to facilitate the molecular-level exploration of biological processes.[207] The resulting field of chemical biology has produced salient technologies including the RNA interference revolution,[208] genome mining[209] and

bioorthogonal chemistry.[210] Chemical genetics, the small molecule-based phenotype screening methodology, has identified novel therapeutics[211] and provided structural and functional insights into intracellular machinery.[212] Future developments in the field will involve harnessing the explorative role of small molecules.[213] To date, compound libraries used for high-throughput screens consist of simple, heteroaromatic molecules suited to a limited number of macromolecular targets. By synthesising compounds with increasing structural complexity, akin to natural products, a broader range of macromolecules may be targeted. The scope of the small molecule approach must also extend beyond druggability. For instance, the interaction propensities of several supramolecular receptors (*e.g.* calixarenes,[106] cucurbiturils,[214] molecular tweezers[215]) have been identified. These may be applied as *in vivo* probes of distinct binding epitopes and quinary interactions.

Just as supramolecular chemistry is inspired by biology,[216] there now arises the need for cell biologists to observe chemical systems for rudiments of emergent physicochemical characteristics. Supramolecular chemistry is veering toward the development of increasingly complex, far-from-equilibrium systems akin to those observed in nature.[216] For instance, the methylation of a dicarboxylic acid eliminates electrostatic repulsion between the molecules to facilitate gelation in the presence of a methyl iodide 'fuel'.[217] Gel disassembly occurs when the MeI is consumed. Given the abundance of D/E-methylation in eukaryotic proteins, including intermediate filament and cytoskeleton remodelling proteins,[218] studies of this artificial system may enhance our understanding of how methylation modulates supramacromolecular assemblies and cytosolic phase separation. The use of structural cell mimics, such as the crowded or confined polymersomes presented by Marguet and colleagues,[36] could be used for bottom-up investigations of assembly and emergent properties in microspaces. For instance, a series of different crowders could be used in such systems to approximate the effects of different microenvironments (*e.g.* cationic, anionic, hydrophobic) on the biophysical properties of a test protein (*e.g.* diffusion, apparent molecular weight, thermodynamic stability). In addition to yielding information on the emergent properties of a cell-like system, this approach may also be used to advance macromolecular crowding theory,[219–221] and to address the significance of the depletion effect *in cellulo*.[222] Increasingly complex physicochemistry could be achieved in cell mimics by incorporating phase-

separating macromolecules.[35,223] Apart from the aforementioned droplet-forming proteins,[24,46,49] other simple, liquid-liquid phase separating systems could be employed in such studies. Aumiller's cationic peptide/RNA droplets are an especially attractive system owing to the ability to control droplet assembly by peptide phosphorylation.[56] Similarly, Koga's poly-α-L-lysine/mononucleotide droplets are sensitive to pH, temperature and ionic strength and, thus, provide a means to explore molecular organisation across a variety of conditions.[52,224]

**Past and Present Philosophies: Notes for the Bioinspired Scientist**

A cursory comparison of the literature from the early 20th and 21st centuries reveals important philosophical differences in the treatment of scientific data. Early biochemists such as Loeb, Troland and Peters consolidated findings from several domains (*e.g.* genetics, chemistry, colloidal science, enzymology) to extrapolate broad biological implications. Troland emphasised the value of logical speculation in solving fundamental biological questions. Speculation overcomes 'the inertia of the human imagination, which does not immediately outline a plan for a machine to accomplish a definite purpose, even when it is provided with all of the principles of mechanics'.[132] Similar views were held by Peters and Loeb who, like Troland, also called for comprehensive physiocochemical frameworks in elucidating cellular processes.[132,143,153] By the mid 20th century, developments in chemical theories (*e.g.* atomic theory, covalent and non-covalent bonding, thermodynamics) and protein and nucleic acid analyses led to an unprecedented breadth of biochemical data. As a result, several biochemical fields emerged (*e.g.* enzymology, metabolism, genetics, pharmacology, neurochemistry and cell, molecular and structural biology) each with a unique intellectual framework and goals. Specialisation and 'fervent reductionism'[89] led to increasingly introspective themes and pragmatic approaches toward theory development. Parsimonious data interpretation substituted speculative abstraction to reveal many (isolated) intricacies of biochemical phenomena. Although insightful, such inward focus precluded the deduction of broad implications and the unification of biological frameworks. Thus, although now 'provided with all of the principles of mechanics' biochemists cannot currently 'outline a plan for a machine to accomplish a definite purpose' *in vivo*.[132]

In the advent of post-reductionist biochemistry, we recognise the need for harmonised experimental practices and intellectual frameworks in the exploration of the cell interior. Revisionism also emphasies the importance of sharing hypotheses in the development of new collaborations and frameworks. An especially poignant example of the power of abstraction is found in Spitzer and Poolman's electrochemical model of the cytoplasm.[63,65] By considering known biophysical parameters (*e.g.* intracellular macromolecular concentrations, proteome p*I* distributions) and clues from bacterial osmosensors, Spitzer and Poolman propose that macromolecular crowding and charge-charge repulsion structure the cytoplasm. This model has helped to establish a physicochemical framework on which future advancements can be built. This model has also helped to foster our current perception of the cell as a bustling metropolitan city,[148] maintained by myriad discrete interactions and dynamic infrastructures (Figure 1, bottom). This industrio-urban perception captures the triumphs of human society engendered by communication and imagination. This image also symbolises the promise of our future in-cell endeavours, which rely on the extensive integration of several scientific communities and creative approaches toward grasping the cell's molecular sociology.

# Chapter 2

# Specific Ion Effects on Macromolecular Interactions in *E. coli* Extracts

The material in this chapter is from:

Kyne C, Ruhle B, Gautier VW, Crowley PB. Specific Ion Effects on Macromolecular Interactions in *Escherichia coli* extracts. *Protein Sci.* **2015**, *24*, 310-318.

**Abstract**

Protein characterisation *in situ* remains a major challenge for protein science. Here, the interactions of ΔTat-GB1 in *E. coli* cell extracts were investigated by NMR spectroscopy and size exclusion chromatography (SEC). ΔTat-GB1 was found to participate in high molecular weight complexes that remain intact at physiologically-relevant ionic strengths. This observation helps to explain why ΔTat-GB1 was not detected by in-cell NMR spectroscopy. Extracts pretreated with RNase A had a different SEC elution profile indicating that ΔTat-GB1 predominantly interacted with RNA. The roles of biological and laboratory ions in mediating macromolecular interactions were studied. Interestingly, the interactions of ΔTat-GB1 could be disrupted by biologically-relevant multivalent ions. The most effective shielding of interactions occurred in $Mg^{2+}$-containing buffers. Moreover, a combination of RNA digestion and $Mg^{2+}$ greatly enhanced the NMR detection of ΔTat-GB1 in cell extracts.

## Introduction

As early as 1930, it was surmised that biomolecules assemble into 'organized networks' that form 'a three dimensional mosaic extending throughout the cell.'[153] Despite such prescience, the nature of intracellular macromolecular organisation remains poorly understood.[15,19,23,65,75,89,91,225,226] This knowledge gap is a result of the reductionist approach of investigating pure proteins under dilute, artificial conditions.[23,89] While the study of pure proteins remains important it is now appreciated that the biophysical properties of proteins inside cells are modulated by the supercrowded, heterogeneous cytoplasm.[65,75,86,89,226] Though daunting, the need for intracellular protein characterisation is apparent. Over the past decade or so, in-cell NMR spectroscopy has emerged to address this challenge.[73–75,77,78,128,227–237] Interestingly, in-cell NMR has confirmed the presence of abundant non-specific interactions inside cells, which can hamper the detection of proteins by $^{1}H$,$^{15}N$ HSQC and related experiments.[73–75,77,78,229,233–237] Furthermore, NMR studies have suggested that extensive interactions also prevail in cell extracts, which mimic the heterogeneous, complex nature of the cytoplasm.[73,74,77,127,232,238,239] For instance, Sarkar and co-workers describe the preparation of a dialysed, lyophilised *E. coli* extract and its effect on barley chymotrypsin inhibitor 2 (CI2) stability at extract concentrations of 100 and 130 g/L.[127] Mass spectrometric and proteomic analysis revealed that the 'reconstituted cytosol' contained 233 proteins.[127] This mixture included 33 of the most abundant *E. coli* proteins,[127] making it a reasonable representation of the *E. coli* proteome. CI2 was destabilised by 0.5-0.8 kcal/mol in the extracts due to weak, non-specific interactions with *E. coli* macromolecules.[127] Later, Sarkar also showed that the destabilisation effected by *E. coli* proteins was mitigated by physiologically-relevant concentrations of glycine betaine.[238] Thus, extracts retain essential characteristics of the cytoplasm and can be treated with exogenous components to foster an improved understanding of intracellular physicochemistry.

Size exclusion chromatography (SEC) of whole cell extracts provides a simple means to dissect macromolecular interactions under native-like conditions.[74,240] Importantly, the elution buffer can be selected to test the effects of biologically-relevant ions, ionic strength and pH. Recently, Crowley used SEC to characterise the interactions of cytochrome *c* and GB1[30] in concentrated *E. coli*

extracts.[74] GB1 eluted from the SEC column at a volume corresponding to the pure protein, indicating that GB1 does not interact with *E. coli* cytosolic macromolecules.[74] Under the same conditions, cytochrome *c* eluted with an apparent molecular weight > 150 kDa indicating its participation in high molecular weight complexes.[74] When the SEC buffer contained 200 mM NaCl, cytochrome *c* eluted at its expected position suggesting that charge-charge interactions drive the complexation process.[74] Similarly, SEC revealed that point and multiple charge-inverted mutants (arginine/lysine→glutamate) of cytochrome *c* eluted with lower apparent molecular weights (*i.e.* reduced complexation) than the wild type.[74] This difference in protein 'stickiness' tallied with observations by in-cell NMR spectroscopy. While in-cell $^{1}H$, $^{15}N$ HSQC spectra can be obtained on GB1,[73,74,229–231] cytochrome *c* and its mutants were undetectable.[74] However, the triple cytochrome *c* mutant was detectable by NMR in *E. coli* extracts. Taken together these data show that extensive complexation rendered cytochrome *c* undetectable by in-cell NMR.[74] Moreover, the utility of combining SEC and NMR for dissecting the interaction mechanisms of proteins in cell extracts was demonstrated.

Here, the influence of the nuclear localisation signal (NLS) of HIV-1 Tat[241] on protein interactions in *E. coli* cell extracts has been explored. An 11 residue Arg-rich sequence based on the Tat NLS was fused to the N-terminus of GB1 to form ΔTat-GB1. Despite high expression levels it was not possible to obtain a $^{1}H$, $^{15}N$ HSQC spectrum of ΔTat-GB1 either in cells or in untreated cell extracts. SEC studies revealed that ΔTat-GB1 interacted extensively with *E. coli* macromolecules. To determine the nature of these interactions different SEC buffers were employed. Combinations of 16 ions were tested at the physiologically-relevant ionic strength of 100 mM.[61,242–245] Previously reported differences in the 'shielding' properties of physiological and laboratory ions[74,244–246] prompted the investigation of both types and the results have been analysed in light of the Collins model.[246–248] Hofmeister effects, which occur at ionic strengths > 200 mM, were not relevant to the present study.[246,249] The digestion of nucleic acids revealed that ΔTat-GB1 was predominantly bound to RNA and these interactions could be further disrupted by $Mg^{2+}$ or succinate ($Succ^{2-}$) at an ionic strength of 100 mM (35 mM $MgCl_2$ or $Na_2Succ$). Interestingly, ΔTat-GB1 could be detected by NMR in extracts that were treated with RNase A and 10 mM $MgCl_2$. This supports recent evidence that the

NMR detection of proteins in cells is hampered by non-specific interactions[74,75127,228] akin to those expected to promote intracellular organisation.[23,65,81,89,90,226] Herein, cell extract NMR is demonstrated as a viable alternative to in-cell methods, especially for the investigation of macromolecular interactions.

**Materials and Methods**

**Mutagenesis** The 11 amino acid sequence GRKKRRERRRA (based on the NLS of HIV-1 Tat)[241] was introduced at the N terminus of GB1 between Met1 and Gln2 to yield ΔTat-GB1. The PCR product was digested with NdeI and SacI, cloned into a modified pET3a vector28 and sequence verified to yield pΔTat-GB1.

**Protein Expression** *E. coli* BL21 (DE3) was transformed with the vector pΔTat-GB1, which contains the ampicillin resistance gene. Cultures were grown as described previously[74] and $^{15}$N-enriched ΔTat-GB1was produced using a two step expression procedure.[74,226] Briefly, a single colony was used to inoculate an overnight Luria Bertani (LB) preculture supplemented with 75 μg/mL carbenicillin and 2 mM $MgCl_2$ and grown overnight at 30 ˚C with 300 rpm shaking. The preculture was diluted 1:100 with LB to produce the day culture. Note that all cultures described hereafter were supplemented with carbenicillin and $MgCl_2$ at the concentrations described above and incubated at 37 ˚C with 300 rpm shaking. When the optical density of the culture at 600 nm ($O.D._{600}$) reached 0.6-0.9 the cells were harvested by centrifugation (4500 rpm, 10 ˚C, 25 minutes) and resuspended in 1 L minimal medium (50 mM $Na_2HPO_4$, 50 mM $KH_2PO_4$, 2 g/L $(^{15}NH_4)_2SO_4$, 2 g/L D-glucose, 20 mM sodium citrate pH 7.0, 20 mM sodium succinate pH 7.0, 1 x 5052,[250] 30 mg thiamine). 30 minutes after resuspension and reincubation, ΔTat-GB1 over-expression was induced with 1 mM isopropyl-β-D-1-thiogalactoside (IPTG). After 4 hours of over-expression the culture was divided between 20 x 50 mL falcon tubes and the cells harvested by centrifugation (4500 rpm, 10 ˚C, 25 minutes). Each cell pellet was resuspended to 1 mL in 20 mM $KH_2PO_4$, 50 mM NaCl, pH 6.0 before freezing. Extracts containing unlabelled ΔTat-GB1 were prepared by directly inducing the LB day culture with 1 mM IPTG after an $O.D._{600}$ of 0.6-0.9 was reached. After 4 hours of over-expression the cells were harvested in 50 mL aliquots (as described above) and resuspended to 1 mL in 20 mM Tris-HCl, 50 mM NaCl, pH 7.0 before freezing.

**Cell Extract Preparation for SEC** Cell extracts were prepared as described[74] with the following changes: Lysis was completed by thawing the cell suspension, adding a few crystals of DNase I or RNase A, and sonicating.[74] The extract was obtained by centrifugation (14,000 rpm for 20 min). The stability of the extracts was temperature dependent. Precipitation occurred after 30 min incubation at 30 ˚C while the extracts remained stable for > 6 hours at 20 ˚C. Sample stability was optimised by preparing the extracts at 20 ˚C directly prior to use.

**Size Exclusion Chromatography** SEC was performed on an Åkta FPLC at 21 ˚C using an XK 16/70 column (1.6 cm diameter, 65 cm bed height) packed with Superdex 75 (GE Healthcare).[18] A continuous flow rate of 1.5 mL/min was employed and the column was equilibrated with 150 mL buffer prior to sample injection. Cell extract samples (850 μL) were injected onto the column and 1 mL fractions were collected. SEC experiments were performed on cell extracts grown on rich or $^{15}N$ labelled medium. Sample elution was monitored at 280 nm. The elution buffer was 20 mM Tris-HCl plus a salt at 100 mM ionic strength (unless otherwise indicated). The pH was adjusted to 7.0 and the buffers were filtered and degassed. To confirm reproducibility, the SEC experiments were performed at least three times for each condition tested.

**Cell Slurry Preparation for In-Cell NMR** Four hours post-induction, 50 mL *E. coli* culture aliquots expressing ΔTat-GB1 were harvested (1500 rpm, 20 ˚C, 15 minutes) and the pellet was gently resuspended in a final volume of 0.5 mL containing 20% $D_2O$.

**Cell Extract Preparation for NMR** The extracts used for the NMR experiments were produced from a 1 L stock culture. Here, the cell pellet was resuspended in 20 mM $KH_2PO_4$, 50 mM NaCl, pH 6.0. Owing to the temperature-dependent stability of the extracts, a fresh sample was used for each HSQC measurement (acquisition time ~20 min at 30 ˚C) and precipitation was avoided during data acquisition. Nucleic acid degradation was achieved by adding a few crystals of DNase I and/or RNase A to the extracts and incubating for 30 min at 20 ˚C prior to spectral acquisition. The extracts were spiked with the following salts: $MgCl_2$, $Na_2Succ$ or $Na_3Cit$ at an ionic strength of 100 mM. 1.5-3.0 M stock solutions of the salts were used so that < 15 mL

volumes were added and the extract was diluted by < 3%. Extracts typically had a pH of 6.5 (± 0.05 pH units).

**NMR Spectroscopy** $^{1}H$, $^{15}N$ HSQC (watergate) spectra[251] were acquired at 30 °C with 8 scans and 64 increments on a Varian 600 MHz Spectrometer equipped with a HCN probe or a cold probe. Data processing was performed in Biopack (with linear prediction in the $^{15}N$ dimension) and NMRPipe.[43] The spectra were analysed using CCPN.[252] All spectra were contoured identically (Figures 1, 6-8). In-cell $^{1}H$, $^{15}N$ HSQC spectra were acquired as described.[74] To avoid leakage, careful sample handling was required to prepare the cell slurry.[74,75,230,232]

**Gel Electrophoresis** 15% polyacrylamide gel electrophoresis (SDS-PAGE) with Coomassie blue staining was used to analyse the protein content of the cell extracts and the SEC fractions. Cell extract samples were diluted fourfold while the SEC fractions were concentrated threefold prior to loading.

**Results and Discussion**

**ΔTat-GB1 is Undetectable by In-Cell NMR** Figure 1A shows the in-cell $^{1}H$, $^{15}N$ HSQC spectrum of over-expressed ΔTat-GB1 in the *E. coli* cytosol. The cell suspension contained ~1 mM ΔTat-GB1, which is ~fiftyfold greater than the detection limit. Unlike GB1,[73,74,229–231] ΔTat-GB1 was not detected in cells or cell extracts even with sensitivity enhanced measurements[251] on a 600 MHz spectrometer equipped with a cold probe. The peaks observed in the in-cell spectrum for ΔTat-GB1 are due to mobile side chain amides and/or $^{15}N$- metabolites.[226] Purified ΔTat-GB1 yielded a high quality spectrum similar to that of GB1 (Figure 1B). We, and others, have shown the deleterious consequences of high molecular weight complex formation for in-cell NMR.[73–75,77,78,228,233–237] The lack of an in-cell spectrum for ΔTat-GB1 suggests that it interacts with macromolecules in the *E. coli* cytosol.

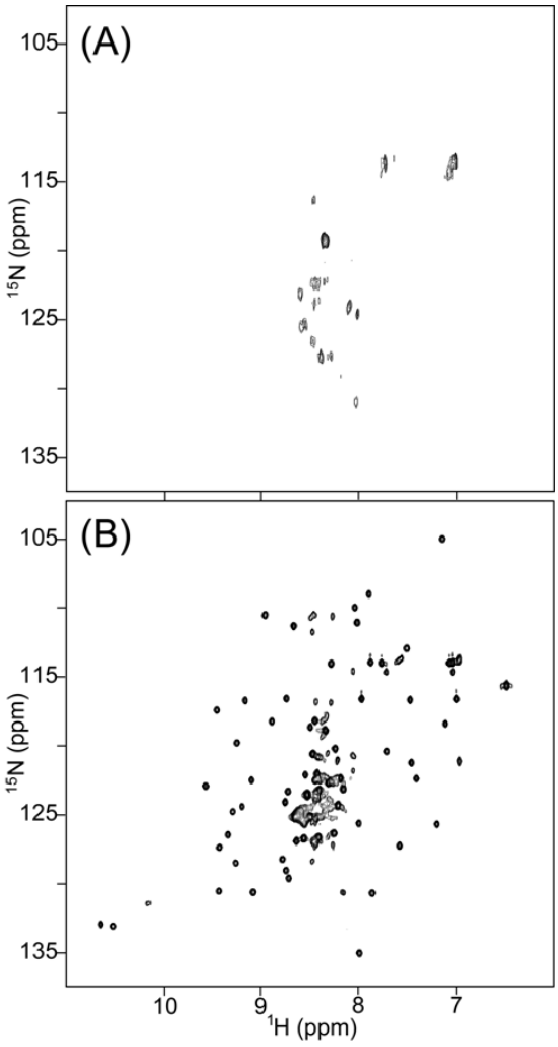
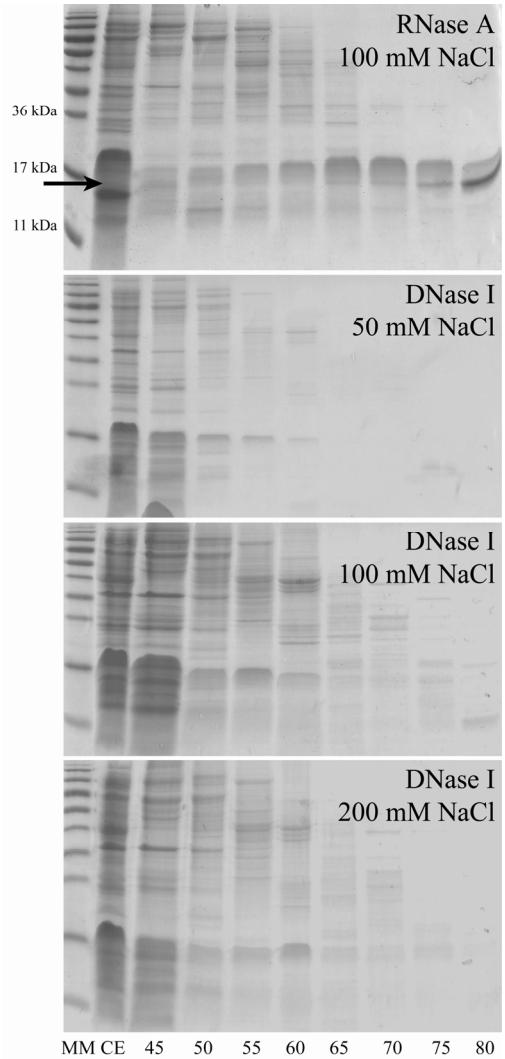



**Figure 1.** $^{1}$H, $^{15}$N HSQC spectra of (A) a suspension of *E. coli* cells containing over-expressed ΔTat-GB1 and (B) purified ΔTat-GB1.

**Interactions of ΔTat-GB1 with Cytosolic Macromolecules** To investigate the nature of ΔTat-GB1 interactions, SEC was performed on cell extracts containing the over-expressed protein. The total macromolecular concentration of the extracts was comparable to that of the *E. coli* cytosol, thus preserving the effects of macromolecular crowding on protein interactions.[74] Pure ΔTat-GB1, a 7.7 kDa protein, eluted from the Superdex G75 column at a volume of ~80 mL, in agreement with molecular weight standards. However, the SEC elution profile for ΔTat-GB1 in cell extracts revealed a different pattern. ΔTat-GB1 interactions with nucleic acids were probed by SEC studies on cell extracts that were digested with (i) RNase A or (ii) DNase I (the cell lysis step involved sonication, which also

fragments DNA). When extracts were treated with RNase A, the SEC elution profile revealed ΔTat-GB1 in many fractions with the greatest amount in fractions 65-75 close to where small proteins (20-30 kDa) typically elute (Figure 2). In contrast, when extracts were treated with DNase I, the majority of ΔTat-GB1 eluted in the void volume (45 mL fraction) corresponding to high molecular weight complexes > 150 kDa (Figure 2, 100 mM NaCl data). It can be concluded that a substantial portion of ΔTat-GB1 binds to RNA and/or ribosomes (the most abundant ribonucleoprotein, containing ~60% ribosomal RNA) in *E. coli*. Although this system lacks physiological Tat partners, ΔTat-GB1 interactions with RNA still occurred. This is consistent with Tat NLS binding to the HIV 'TAR' RNA.[253–255] Notably, the interactions remained intact after cell lysis and exposure to the SEC column, suggesting relatively high affinity complexes. Hereafter, all SEC experiments were performed on cell extracts treated with DNase I only. This reduced the viscosity of the samples (necessary for SEC) without compromising prominent ΔTat-GB1 interactions with RNA.




**Effect of Ionic Strength on ΔTat-GB1 Interactions** To test the effects of biologically relevant ionic strengths, including those observed under osmotic stress,[16,19] SEC was performed at 50, 100, and 200 mM NaCl. As the NaCl concentration was increased, the amount of ΔTat-GB1 in the 45 mL fraction decreased and there was a concomitant increase of ΔTat-GB1 in the 50-60 mL fractions. At 200 mM NaCl, fraction 60 contained the largest amount of ΔTat-GB1 (Figure 2).

**Figure 2.** SEC elution profiles of *E. coli* cell extracts containing over-expressed ΔTat-GB1. The extracts were treated with RNase A or DNase I and the SEC buffers contained 50, 100, or 200 mM NaCl. The gel lanes are labelled; MM: molecular weight marker; CE: cell extract; 45–80: fraction volume (mL). The arrow marks the migration position of ΔTat-GB1.

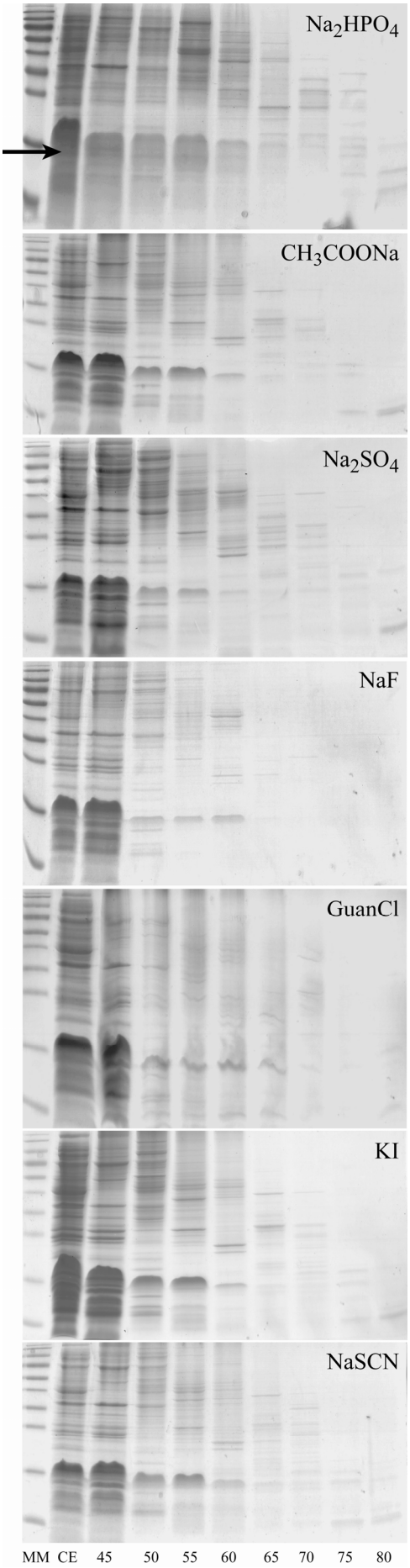



**Figure 3.** SEC elution profiles of *E. coli* cell lysates containing over-expressed ΔTat-GB1. The SEC buffers contained simple physiological and common laboratory ions such as sodium phosphate, sodium acetate, sodium sulfate, sodium fluoride, guanidine hydrochloride, potassium iodide or sodium thiocyanate at an ionic strength of 100 mM. The salts are presented in order of decreasing anionic kosmotropism (see discussion on Collins model).[37-39]

Although the ΔTat-GB1 interactions with RNA were sensitive to ionic strength the protein was never observed in the 80 mL fraction, where pure ΔTat-GB1 eluted. Thus, even at increased concentrations, NaCl was incapable of completely liberating ΔTat-GB1 from its interaction partners. This observation suggests that the complexes are not simply charge-charge in nature. It is likely that additional non-covalent interactions stabilise complexes containing ΔTat-GB1. This is consistent with the fact that Tat NLS is rich in arginine,[241] a uniquely versatile residue in its contribution to protein interactions.[83,255–257]Apart from charge-charge interactions, the guanidinium group is a pentadentate hydrogen bond donor that forms cation-π bonds with aromatic groups including the nucleotide bases in RNA.[256,258]

**Specific Ion Effects on Complexes Containing ΔTat-GB1** To identify which ions, if any, disrupt the interactions of ΔTat-GB1 with *E. coli* RNA, SEC buffers containing different salts were tested. For the majority of the salts investigated, at an ionic strength of 100 mM [$Na_2HPO_4$, $CH_3COONa$,

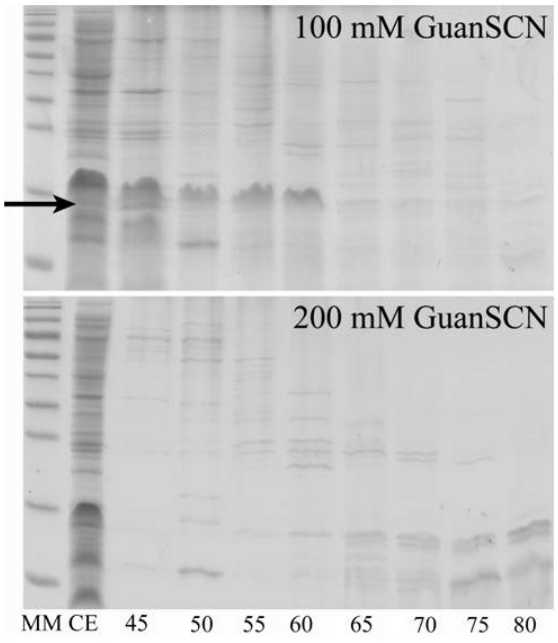



**Figure 4.** SEC elution profiles of *E. coli* cell lysates containing over-expressed ΔTat-GB1. The SEC buffer contained either 100 or 200 mM guanidine thiocyanate. The migration pattern is distorted by the guanidine.

$Na_2SO_4$, NaF, guanidine hydrochloride (GuanCl), KI, NaSCN (Figure 3), guanidine thiocyanate (GuanSCN, Figure 4), $NH_4)_2SO_4$, $NH_4Cl$, (data not shown), KGlu (Figure 5)], the elution profiles were closely similar to the 100 mM NaCl data (Figure 2). In many profiles the amount of ΔTat-GB1 in fraction 45 was identical, within error, to the cell extract (Figures 2-5). Even buffers containing 100 mM $Guan^+$ salts, which might be expected to compete with the arginine-driven interactions, failed to dissociate the complexes (Figures 3 and 4). Interestingly, disruption did occur at 200 mM GuanSCN (Figure 4). Notably, at this ionic strength the Hofmeister effect may contribute to the observed decrease in ΔTat-GB1 complexation. $Guan^+$ and $SCN^-$ are Hofmeister chaotropes capable of unfolding macromolecules[249] and thus, altering their interaction propensities. KGlu is an abundant salt present at concentrations upwards of 100 mM in *E coli*.[3,65,86] In buffers containing 100

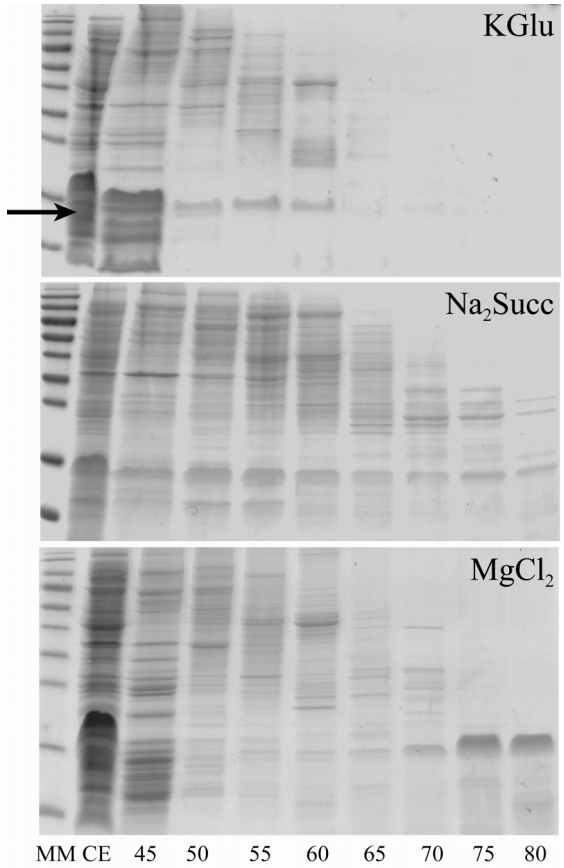



**Figure 5.** Elution profiles for SEC buffers containing physiological ions potassium glutamate, sodium succinate, or magnesium chloride at an ionic strength of 100 mM. The arrow marks the migration position of ΔTat-GB1.

(Figure 5) or 200 mM KGlu (not shown), ΔTat-GB1 eluted in the high molecular weight fractions. The lack of dissociation in the presence of an abundant cellular ion supports the idea that ΔTat-GB1 interacts extensively with *E. coli* macromolecules inside cells.[74] The most effective shielding of ΔTat-GB1 interactions occurred with divalent cations. In $Mg^{2+}$-containing buffers, ΔTat-GB1 eluted in the 75-85 mL fractions corresponding to the pure protein (Figure 5). The identity of the protein was confirmed by NMR spectroscopy; the 75-85 mL fractions yielded a $^{1}H$, $^{15}N$ HSQC spectrum similar to that in Figure 1B. A similar result was observed in $Ca^{2+}$-containing buffers (data not shown). The elution profiles obtained in buffers containing di/tri-carboxylates were also noteworthy for the absence/reduction of ΔTat-GB1 in the 45 mL fraction (See data for $Succ^{2-}$, Figure 5). Perhaps di/tricarboxylates are effective at disrupting ΔTat-GB1 interactions because they compete with the backbone phosphates of nucleic acids for the Tat NLS. Herein, the effects of both physiological (*e.g.* $K^{+}$, $Mg^{2+}$, $Glu^{-}$, $Succ^{2-}$) and laboratory ions (*e.g.* $Guan^{+}$, $Na^{+}$, $Cl^{-}$, $F^{-}$, $I^{-}$, $SCN^{-}$) were tested by SEC. Although common laboratory ions are often used to adjust the ionic strength of buffers in biological studies, recent investigations have shown that they do not accurately mimic the effects of physiological ions.[74,244–247] The laboratory ions tested by SEC were ineffective at shielding ΔTat-GB1 interactions with *E. coli* RNA. Although similar results were observed for the major cellular ions $K^{+}$ and $Glu^{-}$, the multivalent ions $Mg^{2+}$, $Ca^{2+}$, $Succ^{2-}$ and citrate ($Cit^{3-}$) shielded ΔTat-GB1 interactions. Thus, it is important to consider the effects of numerous (biological) ions in the assessment of macromolecular interactions.

**Cell Extract NMR Studies Recapitulate the Nature of ΔTat-GB1 Interactions**

While SEC can reveal the individual mechanisms governing a protein's interaction propensities in extracts, NMR can reveal how these mechanisms interrelate and manifest. Thus, ΔTat-GB1 interactions with nucleic acids and specific ions were also investigated in cell extracts by NMR (Figure 6). The extracts contained over-expressed ΔTat-GB1 and were untreated, treated with a nuclease (DNase I or RNase A) and/or spiked with a salt ($MgCl_2$, $Na_2Succ$ or $Na_3Cit$) at an ionic strength of 100 mM. Careful extract preparation was implemented to avoid sample precipitation (see Materials and Methods). Owing to the predominance of ΔTat-GB1 interactions with

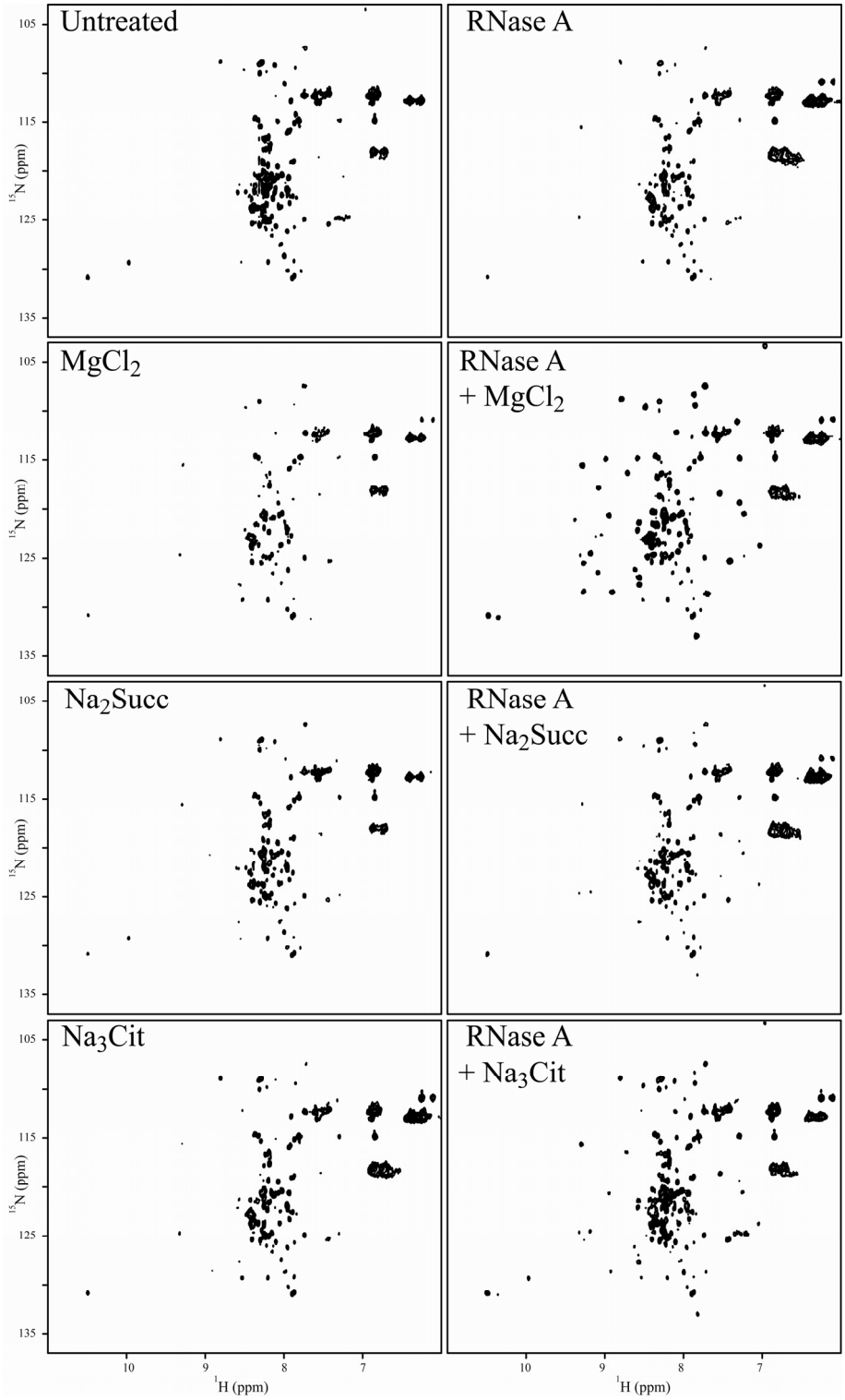



**Figure 6.** $^{1}H$, $^{15}N$ HSQC spectra of *E. coli* cell extracts containing over-expressed ΔTat-GB1. The left column shows spectra of extracts that were untreated or spiked with a salt. The right column are extracts that were RNase A treated and spiked with a salt. All of the salts were added to an ionic strength of 100 mM.

RNA, the effect of RNase A incubation times on ΔTat-GB1 interactions was also assessed by NMR. Spectra were obtained on extracts after incubation with RNase A for 30 and 60 min at 20 ˚C (Figure 7). The results were closely similar suggesting

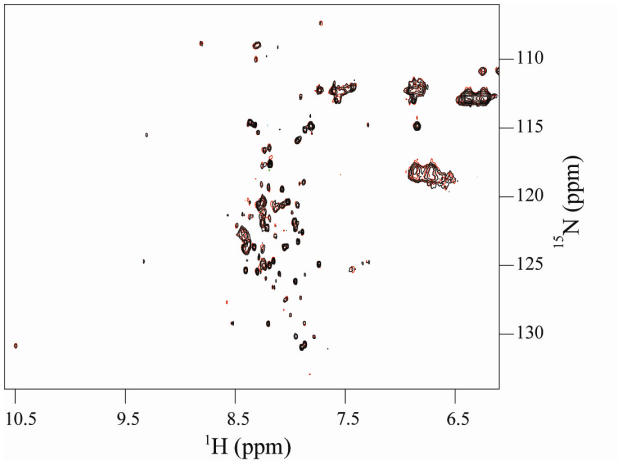



**Figure 7.** $^{1}H$, $^{15}N$ HSQC spectra of cell extracts containing ΔTat-GB1 that were pre-treated with RNase A for 30 or 60 minutes (black and red, respectively). The spectra were contoured identically and are closely similar. This indicates that the effects of RNase A activity on ΔTat-GB1 behaviour have culminated by 30 minutes.

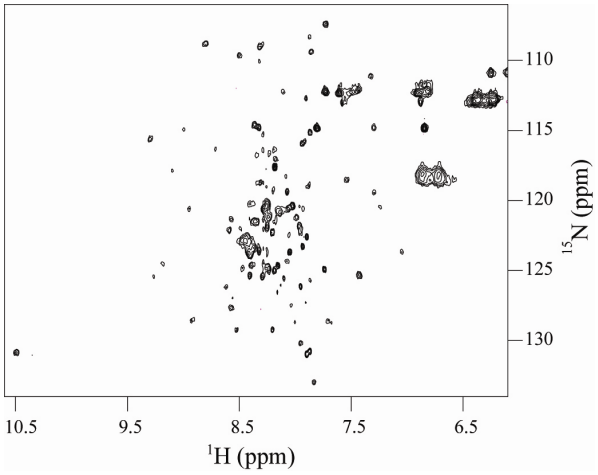



**Figure 8.** $^{1}H$, $^{15}N$ HSQC spectrum of an *E. coli* extract containing ΔTat-GB1 that was pre-treated with RNase A and spiked with 10 mM $MgCl_2$

that the effects of RNase A digestion had culminated by 30 min. Figure 6A shows the $^{1}H$, $^{15}N$ HSQC spectrum for ΔTat-GB1 in an untreated cell extract. Compared to the in-cell data (Figure 1A), a marked improvement in spectral quality was evident in the extract. However, the spectrum did not resemble that of purified ΔTat-GB1 (Figure 1B) because many amide resonances were broadened beyond detection. The spectra of extracts after RNA digestion or spiking with $MgCl_2$, $Na_2Succ$ or $Na_3Cit$ (Figure 6) were comparable to the spectrum from the untreated extract. This indicates that ΔTat-GB1 interacts with RNA despite digestion or the addition of 'shielding' ions. Spectra were also obtained on extracts treated with RNase A and $MgCl_2$, $Na_2Succ$ or $Na_3Cit$ (Figure 6). The presence of $MgCl_2$ resulted in a dramatic improvement in spectral quality, yielding a HSQC spectrum similar to purified ΔTat-GB1 (Figure 1B). Importantly, 10 mM $MgCl_2$ was sufficient to effect this substantial improvement in the spectrum (Figure 8). Similar spectral improvements were not effected in extracts spiked with 1 mM $MgCl_2$, which corresponds to the concentration of free $Mg^{2+}$ in the *E. coli* cytoplasm.[7] The carboxylate anions $Cit^{3-}$, and to a lesser extent $Succ^{2-}$, also gave rise to spectral improvements with the appearance of additional ΔTat-GB1 resonances and narrower line widths (Figure 6). Together, these data confirm that ΔTat-GB1 interactions with

RNA (and the formation of high molecular weight complexes) render it undetectable by in-cell NMR.

**The Interaction Propensity of ΔTat-GB1 in Extracts** The grand challenge of protein science is to decipher how macromolecular interactions drive subcellular architectures and biological processes. Such knowledge relies on *in vivo* studies. However, semi-reductionist approaches are also helpful to determine the physicochemical basis of macromolecular interactions in biological milieus.[73,74,77,78,230,233–236] Herein, SEC and NMR were used to investigate the interaction behaviour of ΔTat-GB1 in *E. coli* extracts. Importantly, ΔTat-GB1 and *E. coli* RNA are not natural partners. This facilitated the assessment of pervasive, non-specific interactions involving this model RNA-binding motif[255,258] in a biological

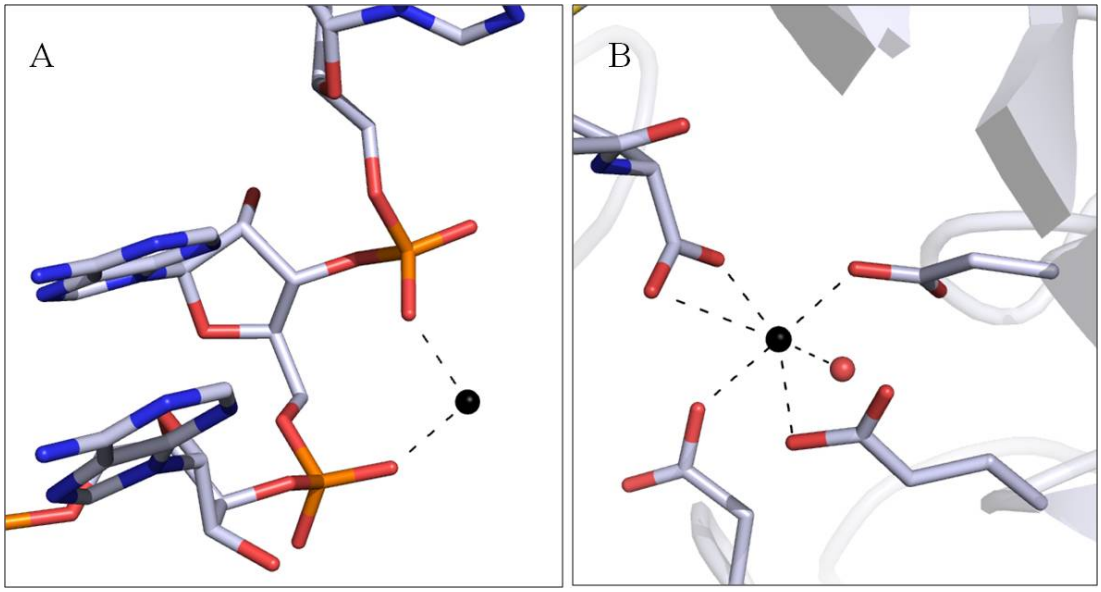



**Figure 9.** $Mg^{2+}$ chelation by (A) adjacent RNA phosphates oxyanions and (B) side chain carboxylates (PDB ID = 1jj2[259] and 1sjc,[260] respectively). $Mg^{2+}$-binding nucleotides and residues are shown as sticks. Atoms are coloured grey, C; blue, N; orange, P; red, O. The black and red spheres represent the $Mg^{2+}$ and water oxygen, respectively. Dashed lines indicate interatomic distances < 2.7 Å.

environment. The stickiness of the Tat NLS was demonstrated in that its interactions were dissociated in the presence of $Mg^{2+}$, which interacts strongly with biological oxyanions (*i.e.* phosphates and carboxylates).[225,247,248] For instance, a frequent mode of $Mg^{2+}$-phosphate interaction involves bidentate chelation by phosphate anions of adjacent nucleotides (Figure 9A).[259]

$Mg^{2+}$-binding structural motifs in proteins are not well understood but $Mg^{2+}$-binding sites are typically observed at loops and involve clusters of anionic sidechains which complement the octahedral co-ordination shell of $Mg^{2+}$, usually in combination with water oxygens (Figure 9B) or amide carbonyl oxygens.[260,261]

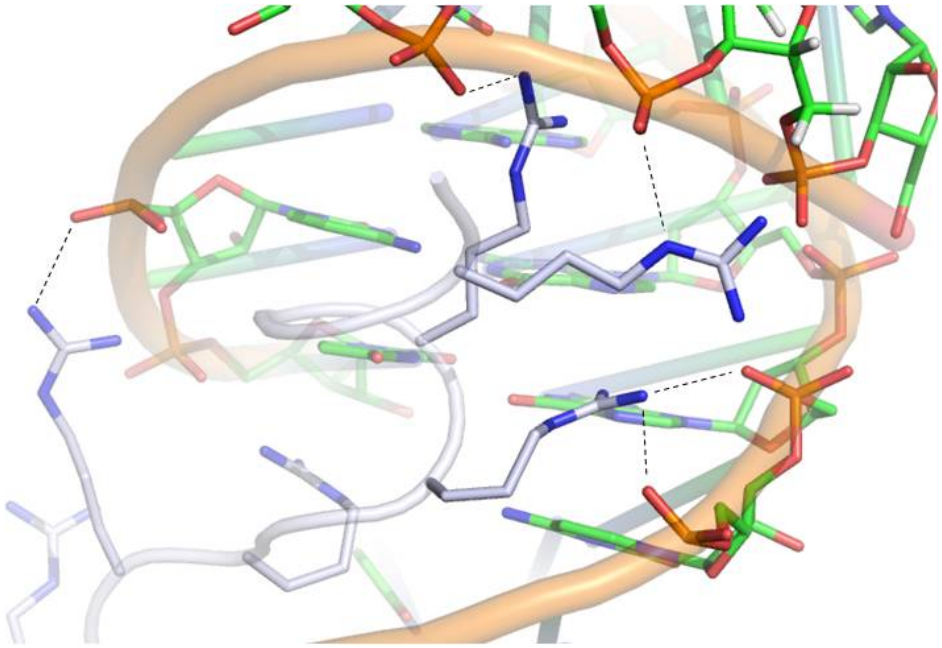


**Figure 10.** Interaction of the arginine-rich HIV-1 REV motif with an RNA aptamer (PDB ID = 484d).[59] Several arginine guanidinium groups are < 3.5 Å of RNA backbone phosphates indicating that they are engaged in charge-charge interactions. The dashed lines indicate potential salt bridge pairs.

The disruptive effect of $Mg^{2+}$ emphasises the importance of Arg-oxyanion salt bridges in driving ΔTat-GB1 interactions and suggests that Arg-phosphate salt bridges drive complexation between ΔTat-GB1 and RNA in *E. coli*. Together with $Mg^{2+}$ addition, RNA digestion was necessary for the NMR detection of ΔTat-GB1 in cell extracts. This supports the idea that multiple RNA phosphates neutralise the disordered, poly-cationic Tat NLS.[262] Notably, this binding mode differs from that reported for the specific Tat-TAR complex.[253–255] Although Arg/Lys-phosphate contacts have been linked with 'nonspecific' protein-RNA binding (Figure 10),[255,258] recent evidence shows that some RNA-binding proteins interact at sites lacking defined recognition elements.[263]

**Preferential Ion Pairing in Cell Extracts** To identify whether the specific ion effects are corroborated by established ion-pairing rules at physiological ionic strengths, our findings are assessed in light of the Collins model.[246–248] Hofmeister

effects (which occur at ionic strengths > 200 mM) are not relevant to the data presented here and will not be discussed.[246,249] Collins' "Law of matching water affinities" classifies ions as kosmotropes (small, charge dense and strongly hydrated) and chaotropes (larger, less charge dense and weakly hydrated by rapidly exchanging water molecules). Preferential interactions are observed between species of equivalent hydration and opposite charge. Kosmotrope-kosmotrope interactions are more favorable than kosmotrope-water interactions due to the high charge density and small ionic radii of the kosmotropic pair. Chaotropes interact weakly with water and so, preferentially pair, allowing for the formation of more energetically favourable water-water interactions. Kosmotrope-chaotrope pairs are unfavourable compared to their hydrated ionic counterparts and therefore tend not to interact.

As per the Collins model, the guanidinium groups in the Arg-rich Tat NLS are chaotropic (weakly hydrated) and should preferentially interact with weakly hydrated anions. $I^-$ and $SCN^-$ might therefore be expected to dissociate complexes containing ΔTat-GB1. However, at 100 mM ionic strength these anions failed to compete with *E. coli* RNA for binding to ΔTat-GB1 (Figure 3 and 4). Similarly, efforts to break ΔTat-GB1 interactions using simple physiological anions[3,225,245] such as $HPO_4^{2-}$, Acetate ($Ac^-$; Figure 3) and $Glu^-$ (Figure 5) were also ineffective. By comparison, di/tricarboxylate anions such as $Succ^{2-}$ (Figure 5) disrupted a substantial proportion of the ΔTat-GB1 complexes. Correspondingly, the majority of *E. coli* macromolecules contain negatively charged patches that, like polycarboxylate anions, are kosmotropic and multivalent.[3,65,68,225] $Mg^{2+}$, the most strongly hydrated of all biological cations, is a 'Collins counterion' compatible with biological phosphate and carboxylate oxyanions (Figure 9).[225,245] The disruptive effect of $Mg^{2+}$ is therefore consistent with the idea that the Tat NLS binds RNA phosphates in *E. coli* extracts. While the preferential $Mg^{2+}$-phosphate interaction is consistent with the Collins model, the Arg-oxyanion salt bridge represents a chaotrope-kosmotrope ion pair that is deemed energetically unfavorable.[246–248] We surmise that the high effective concentration[247,264] of positive charge (8 Arg and Lys residues) renders the NLS an effective oxyanion (*i.e.* kosmotrope) binder. Our data are part of a growing repertoire that suggests that preferential ion interactions between macromolecular charge clusters deviate from those predicted by the Collins model, which holds well for simple systems.[244] A more advanced model that incorporates not only the ion

properties, but also the charge density/distribution at the protein surface is necessary to describe systems of biological complexity.

**Conclusions**

All cellular processes are governed by the intricate interplay of electrolyte and macromolecular interactions.[23,65,89,91,225] The abundance of high molecular weight assemblies *in vivo* is well-appreciated and the inability to detect most globular proteins by in-cell NMR supports this idea.[73,74,77,78,128,226–237] Herein, ΔTat-GB1 stickiness was explored in *E. coli* cells and extracts by an enhanced sensitivity HSQC experiment. The protein was undetectable in these physiological environments (Figure 1A, 6-8) indicating that it interacts pervasively and non-specifically with *E. coli* macromolecules. A previously described SEC technique[74] showed that ΔTat-GB1 interactions were disrupted upon RNA digestion (Figure 2), suggesting that the protein forms high molecular weight complexes with RNA in the extracts. These interactions likely prevail in living *E. coli* cells and render the protein undetectable by in-cell NMR spectroscopy. Changes in the SEC buffer salts and ionic strengths were used to explore the physicochemical origin of ΔTat-GB1 interactions. Complexes containing ΔTat-GB1 remained intact in the presence of a range of different ions including the biologically abundant potassium glutamate (Figures 2-5). Likewise, high concentrations (200 mM) of the laboratory salt NaCl did not liberate ΔTat-GB1 (Figure 2). Instead, low concentrations of $Mg^{2+}$ were effective in disrupting the complexes (Figure 5). Divalent cations interact strongly with biological oxyanions[247] apparently dissociating Tat NLS-oxyanion interactions in the process. Similarly, di/tricarboxylate anions successfully competed with the anionic surface patches of *E. coli* macromolecules to interact with the Tat-NLS (Figure 5). Importantly, ΔTat-GB1 was NMR detectable in extracts only after the addition of $> 10$ mM $MgCl_2$ and RNA digestion (Figure 6).

Together, these data suggest an interaction mechanism for the Tat NLS was in the *E. coli* extracts. For example, the need for RNA digestion implies that the chelate effect drives ΔTat-GB1 interactions with *E. coli* RNA. The need for the phosphate binding cation, $Mg^{2+}$, suggests that Arg-phosphate salt bridges mediate the interactions of ΔTat-GB1 with RNA. These data contrast with the Collins model of preferential ion pairing[246–248] and suggest that specific ion effects must be re-

evaluated to explain the differential binding phenomena observed for densely charged protein surfaces at physiological ionic strengths in crowded cell environments.[65,225,244] SEC rivals other interactomic techniques in that it can dissect the interaction propensities of a protein in the context of specific ion effects. Cell extract NMR can reveal how the interaction mechanisms of a protein are borne out in a cell-like environment. When combined, these techniques allow for the extensive exploration of molecular 'cross talk' under native-like conditions which, in turn, help us to delve deeper into the post-reductionist era of protein science.

# Chapter 3

# Charge Determination of Cytochrome *c* and its Mutants: Implications for Cell Extract Interactions

*Manuscript in preparation*

**Abstract**

A major goal of protein science is to develop electrochemical frameworks that describe protein interactions in the complex, volume-occupied cell interior. Such frameworks warrant a comprehensive understanding of protein electrostatics which give rise to the potential energies governing non-covalent interactions. Charge is the only electrostatic property that can be accurately measured. Thus, studies of protein charge are an important foundation on which electrochemical frameworks can be built. Here, the charge of cytochrome *c* and charge-inverted (R/K→E) mutants are reported. The apparent charge of cytochrome *c* is salt-dependent and preferential cytochrome *c*-anion interactions are revealed. Native gel electrophoresis, together with knowledge of the cytochrome *c* and mutant charges, have been applied to explore the origin of cytochrome *c* stickiness in *E. coli* cell extracts.

## Introduction

**Protein Charge Measurements are Necessary to Elucidate Colloidal Protein Properties** Protein interactions mediate all cellular processes. Protein-macromolecule interactions are stabilised by several types of non-covalent interactions (*e.g.* charge-charge, hydrogen bonding, van der Waals interactions and the hydrophobic effect).[83] The stability and specificity of complexes are determined by different relative contributions of each interaction type.[265,266] However, the role of attractive charge-charge interactions in electrostatically steering oppositely charged partners toward productive docking geometries (*i.e.* increasing $k_{on}$) is well-established *in vitro*.[267,268] The physiological significance of charge-charge interactions is implicit in their long-range nature and the abundantly Coulombic character of physiological environments.[63,64] For instance, protein charge produces a potential energy 'halo' that extends from the van der Waals surface of the protein for up to 10 Å at physiological ionic strengths.[23] A similar distance separates macromolecules in high-concentration systems such as the cytoplasm. Thus screened charge-charge interactions are of biological importance.[23,64] Furthermore the *E. coli* cytoplasm is replete with anionic macromolecules suggesting that charge-charge repulsion is critical in stabilising the cytoplasm against random aggregation.[63]

At a local level, the high surface-to-volume ratios and chemically heterogeneous nature of the cytoplasm promotes multifarious, attractive macromolecular interactions. These co-operative interactions lead to the formation of macromolecular clusters[65] or microenvironments[5] and can be conducive to phase separation. Indeed, the discovery of cytoskeletal structures[269] and proteins containing membraneless organelle-forming motifs[49] in the *E. coli* cytoplasm supports the idea that it is a sol/gel (*i.e.* colloid).[65] Accordingly, Laue and Demeler considered aspects of colloid and surface chemistry as a starting point for tackling the 'Gordion Knot of *in vivo* chemistry'.[23] Colloidal chemistry typically explores the electrostatic properties of the system components to elucidate its physicochemical characteristics (*e.g.* stability).[270] Proximity energies represent the sum of the potential energies arising due to the electrostatic properties of macromolecular surfaces.[23] To date, charge is the only proximity energy that can be measured accurately.[271] Knowledge of protein charge is therefore fundamentally important in developing the "proximity

energy framework"[23] describing protein behaviour in complex, colloidal systems like the cytoplasm.

**Exploring Protein Charge using Membrane Monfined Electrophoresis** Several free-solution electrophoresis methods exist for protein charge determination including capillary electrophoresis (CE),[272] electrophoretic light scattering (ELS)[271] and membrane confined electrophoresis (MCE).[129] MCE is the most straightforward method for the accurate determination of protein electrophoretic mobility (μ) and effective electrical valence ($Z_{eff}$).[129,271,273] Protein valence is the ratio of its Coulombic charge, Q, to the elementary proton charge, $Q_p$ (both in Coulombs). The effective electrical valence is the formal valence reduced by the shielding effects of the Debye-Hückel counter-ion cloud and the electrophoretic effect.[271,274] MCE instrumentation facilitates two types of experiment: steady state electrophoresis (SSE) and real-time electrophoretic mobility (REM). In both experiments, macroions are confined within a quartz cuvette which is sealed at either end by semi-permeable membranes. An electric current is passed through the cuvette, establishing an invariant electric field along its length. In the applied electric field, macroions migrate toward the oppositely charged electrode. At regular time intervals, fixed wavelength absorbance scans are acquired at distinct points across the cuvette. The resulting absorbance profile shows the macroion concentration distribution along the cuvette. In SSE experiments, the magnitude of the applied electric field is relatively low (*i.e.* 0.05-0.2 V/cm) and macroion flux in the electric field is countered by macroion flux due to diffusion.[129] When the steady state is reached (10-12 hours), the flux of ions due to electrophoresis is balanced by the flux due to diffusion and a stable macroion concentration gradient forms in the cuvette. At the steady state all system properties, including the macroion concentration gradient, are invariant over time and the macroion no longer contributes to the net current at any point in the cuvette. The macroion concentration gradient is exponential with respect to the x-axis (position in cm) and can be fit to the following equation to yield the effective charge:

$$c(x) = c_o \exp\hat{}[Z_{eff}\,[E/(k_BT)](x-x_o)]$$

where: $\sigma$= baseline offset ($cm^{-1}$), $c_o$ = concentration at reference point $x_o$ (mg/mL), $Z_{eff}$ = effective electrical valence, $x_o$ = arbitrary initial reference point (cm), T = temperature (K), E = electric field (V/cm), and $k_B$ = Boltzmann's constant (1.3807 x $10^{-23}$ J/K).

REM experiments are faster than SSE experiments (~40 minutes versus 10-12 hours) and determine the electrophoretic mobility of the macroion by applying a strong electric field (*i.e.* 2.5-25 V/cm) forcing the macroion toward one end of the cuvette. The electrophoretic mobility ($\mu$) of the macroion is its velocity per unit electric field (in $cm^2/V.s$) and can be calculated by measuring sequential intensity scans and subtracting each scan from the previous one to yield the time-derivative of the macroion concentration distribution. $Z_{eff}$ can be calculated from $\mu$ as:

$$Z_{eff} = \mu f$$

where: $f = 6\pi\eta R_s$, $\eta$= solvent viscosity (g/cm.s), and $R_s$ = stokes radius of the protein (cm).

**Ion Binding Modulates Protein Charge** MCE determines the 'charge structure' of a protein,[271] which is the macromolecule's formal charge modified by bound ions. Two types of ion binding contribute to protein charge structure. Firstly, ions can interact with specific sites on a protein surface. For instance, the intracellular $Ca^{2+}$ receptor calmodulin has four well defined binding sites in which $Ca^{2+}$ is co-ordinated in a pentagonal bipyramidal geometry by specific main and side chain oxygens (Figure 1).[275] The second type of ion binding is non-specific or 'territorial' and is not typically evident in protein structures determined by X-ray crystallography. Territorially-bound ions do not occupy a specific site on the protein but instead bind diffusely to regions of high charge density on the protein surface.[273] Charge dense protein surfaces may be characterised by the presence of at least two charges separated by < 3 Å at physiological ionic strengths.[271] Territorial ions can exchange with solvent ions but no net change in protein charge is effected due to this exchange. Territorial ions reside within the protein solvation layer and diffuse with the protein. Thus, because proteins continuously bind to and release counter-ions, several distinct charged species can predominate during the course of the MCE experiment. Therefore, the MCE-determined charge value reports on the average charge of the macroion. Note, the Debye-Hückel counter-ion cloud refers to the

imbalance of co- and counter-ions surrounding the protein. The counter-ion cloud does not contribute to the charge structure of a protein but assembles as a result of it.[271,273] Thus, Debye-Hückel counter-ions are distinct from site-bound and territorial ions. The Debye-Hückel-Henry approximation can be used to adjust for the effects of the Debye-Hückel counter-ion cloud yielding the formal or Debye-Hückel-Henry valence:

$$Z_{DHH} = \mu f\,[(1+k_D\,a)/(H\,Q_p)]$$

where: $k_D$ = inverse Debye length ($cm^{-1}$), α = Stokes radius of protein + counter-ions (cm), H is Henry's function (unitless) which accounts for the electrophoretic effect.[276] Two phenomena comprise this effect (i) the distortion of the electric field lines due to the presence of a non-conducting species such as a protein and (ii) momentum transfer from the flow of counterions proximal to the non-conducting species.[277]

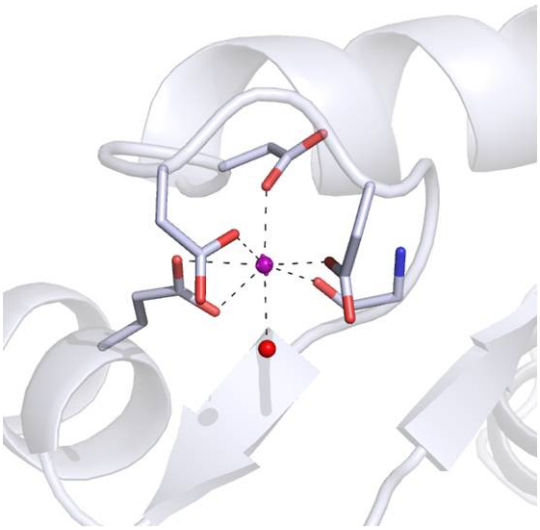


**Figure 1.** $Ca^{2+}$ binds specifically to an EF hand domain in calmodulin (PDB ID = 1cll).[278] $Ca^{2+}$-binding residues are shown as sticks. Atoms are coloured grey, C; blue, N; red, O. The purple and red spheres represent the $Ca^{2+}$ and water oxygen, respectively. Dashed lines indicate interatomic distances < 2.6 Å.

Currently, all free-solution electrophoresis techniques are applicable only to pure, dilute protein samples and are unsuited to studies of complex, volume-occupied solutions. However, changes to the electrophoresis buffer salts, ionic strength and pH can reveal variations in the protein charge resulting from ion-protein interactions. Given the ubiquity of ion-protein interactions in nature,[61,87,279] such knowledge has significant physiological implications. For instance, cytochrome *c* is the most abundant protein in the mitochondrial intermembrane space (IMS)[280] where it transfers electrons between complex III and IV of the respiratory chain. Porin channels (or voltage-dependent anion channels) in the mitochondrial outer membrane regulate the flux of metabolites and

ions (< 5 kDa; *e.g.* ATP and $Ca^{2+}$) between the cytoplasm and the IMS.[281] Thus, cytochrome *c* functions across a range of ionic strengths (100-150 mM) and metabolite compositions.[282] Indeed, the role of electrostatics in modulating cytochrome *c* interactions with its cognate[283–285] and non-physiological partners[74,106,109,112,286] is well established.

**Charge Governs Cytochrome *c* Stickiness in *E. coli* Extracts** Previously, Crowley and co-workers used size exclusion chromatography (SEC) to demonstrate that charge-inverted mutations (K/R→E) decreased the stickiness of cytochrome *c* in *Escherichia coli* extracts.[74] The three charge inverted mutants were: a single (R13E), double (R13E/K73E) and triple (R13E/K73E/K87E) mutant where residues 13, 73 and 87 comprise the known binding site of cytochrome *c* (Figure 2).[74,106,109,114,287] When WT cytochrome *c*-containing extracts were loaded on to an SEC column equilibrated in low ionic strength buffer (20 mM $KH_2PO_4$, 50 mM NaCl, pH 6.0), cytochrome *c* (12.7 kDa) eluted with an apparent molecular weight > 44 kDa.[74] Thus, cytochrome *c* co-eluted with other *E. coli* macromolecules as part of a high molecular weight assembly. In the same SEC buffer the single mutant eluted in fractions corresponding to high and low molecular weight species while the double mutant eluted with an apparent molecular weight of 10-44 kDa.[74] The triple mutant eluted at the expected volume (75-80 mL) corresponding to the monomeric molecular weight of cytochrome *c*.[74] These elution patterns indicate that the net charge of cytochrome *c* governs the extent to which it interacts with other macromolecules. WT cytochrome *c,* the most cationic of the series, is the stickiest. The authors rationalised this finding by considering the electrostatic nature of the *E. coli* cell interior, which is replete with anionic macromolecules that likely interact non-specifically with cytochrome *c*.[74] The decrease in cytochrome *c* charge effected by the mutations decreases the protein's propensity to interact with anionic macromolecules. Accordingly, the triple mutant did not interact with *E. coli* macromolecules. This result indicates that cytochrome *c* complexation in extracts is governed by attractive charge-charge interactions. Similarly, WT cytochrome *c* eluted in the 75-80 mL fractions when extracts were treated with a higher ionic strength buffer (20 mM $KH_2PO_4$, 200 mM NaCl, pH 6.0). The disruption of cytochrome *c* interactions effected at increased ionic strengths supports the idea that

cytochrome *c* interacts with anionic macromolecules via simple charge-charge interactions.

Herein, MCE was used to determine the charge of cytochrome *c* in the presence of different salts, at the physiologically-relevant ionic strength (*I*) of 100 mM. Importantly, the cytochrome *c* charge was found to be substantially lower than that predicted from its primary structure. Preferential ion interactions with the cytochrome *c* surface were also identified. Notably, sulfate anions bind tightly to cytochrome *c*, forming a charge-neutral species. This finding coincides with a known sulfate binding site in the crystal structure for cytochrome *c*.[287] The charge-inverted cytochrome *c* mutants were investigated by MCE also. The decrease in cytochrome *c* charge effected by each mutation correlated with the decreased stickiness of the mutants in concentrated *E. coli* extracts.[74] The charge and interactions of cytochrome *c* and the mutants were explored by using native gel electrophoresis and SEC both in dilute solution and in *E. coli* extracts. The effects of a physiologically relevant-milieu on cytochrome *c* charge were revealed, and the implications for charge-based intracellular assemblies were considered.

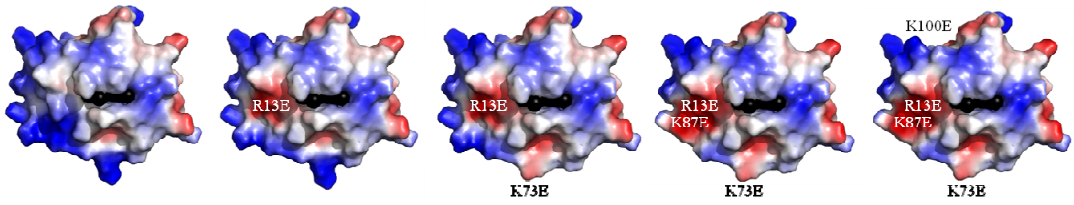



**Figure 2.** Electrostatic surface representations of WT cytochrome *c* and the charge inverted mutants, where positive and negative potentials are coloured blue and red, respectively. The exposed haem edge is shown as black spheres. The position of lysine residues are indicated.

## Materials and Methods

**Protein Production and Purification** *S. cerevisiae* cytochrome *c* and its mutants were expressed and purified according to previously described methods.[74,287] The single and triple cytochrome *c* mutants were provided by Madeleine Mallon and Róise McGovern. The $^{1}H,^{15}N$ HSQC NMR spectra of the mutants are comparable to that of WT cytochrome *c* indicating that no gross structural changes occurred due to mutagenesis.[74]

**MCE Buffer Preparation** Buffers were prepared on the day of use with distilled, deionised water. Each buffer contained 10 mM Bis-TRIS Propane (BTP) plus salt at an ionic strength of 100 mM. The pH was adjusted to 7.0 before the buffers were filtered and degassed. Buffer conductance was determined at 20 ˚C using a VWR1054 conductivity meter with a platinum probe. Average conductivities for the KCl, $KNO_3$, $K_2SO_4$ and $MgCl_2$-containing buffers are given in Table 1.

**Table 1.** Average conductivity of MCE buffers: 10 mM BTP, 100 mM salt, pH 7.0 at 20 ˚C.

| MCE Buffer Salt ($I$ = 100 mM) | Average Conductivity (mS/cm) |
|---|---|
| KCl | 12.00 |
| $KNO_3$ | 11.73 |
| $MgCl_2$ | 7.44 |
| $K_2SO_4$ | 7.22 |

**MCE Membrane Preparation** Spectra/Por Biotech Regenerated Cellulose Dialysis membranes with a MCWO = 8-10 kDa were soaked in H2O for 5 minutes prior to cutting with a stainless steel punch. Membrane discs of 5 mm diameter were thus generated. The membrane disks were boiled in H2O for 10 minutes then cooled and stored at 4 ˚C.

**Membrane Confined Electrophoresis** Comprehensive descriptions of the MCE instrument have been reported previously.[129,271,273,274] All MCE experiments were performed at 20 ˚C following at least 2 hours of complete system equilibration in which a constant flow of buffer (at 10 mL/hour) was used to condition the sample-confining membranes. Protein samples were typically studied at a concentration of 1 mg/mL. During REM experiments, currents of ± 1 mA were applied across the cuvette such that electric fields of ± 2-2.5 V/cm were achieved. Initially, both positive and negative electric fields were applied to each sample and 250 absorbance scans (at 230 nm) were acquired for each electric field. The values presented herein represent an average of at least two measurements with an error of < 30%. Proteins that did not form concentration gradients in response to positive or negative electric fields at magnitudes of either 2-2.5 or 20-25 V/cm or during SSE experiments were taken to have a charge of ~0. MCE is unsuited to systems in dynamic equilibrium.[271] Thus, to ensure that the charge-neutral proteins were monomeric during the MCE experiments, measurements were also performed at concentrations below the critical

aggregation concentration of most proteins *i.e.* 0.1 (8.3 nM) and 0.05 mg/mL (4.2 nM). SSE was also performed on the proteins that appear to be charge-neutral by MCE.[129,271] SSE experiments were performed in electric fields of ± 187 and 374 mV/cm over 24 hours and absorbance scans were acquired every 15 minutes.

The MCE data were analysed using the Spin Analytical MCE Analysis Software. The electrophoretic mobility was converted to $Z_{eff}$ and Z using ZUtilities (available at http://www.rasmb.org/). A radius ($R_s$) of 1.65 nm[105,288] was used for cytochrome *c* and the charge inverted mutants. Moreover, anionic radii of 0.121, 0.129 and 0.230 nm were used for $Cl^-$, $NO_3^-$ and $SO_4^{2-}$, respectively.[246]

**Sedimentation-Velocity Analytical Ultracentrifugation** SV-AUC experiments were performed using a Beckman Coulter XLA analytical ultracentrifuge. Data were acquired at 50,000 rpm at 20 ˚C in 100 mM KCl, 10 mM BTP, pH 7.0. Samples were equilibrated to 20 ˚C prior to ultracentrifugation. Sedimentation was monitored using absorbance scans at 550 nm and a pathlength of 1.2 cm. At this wavelength, the reduced cytochrome *c* haem has an extinction coefficient of 27.5 $mM^{-1}cm^{-1}$.[289] Sample concentrations of 0.5, 0.15 and 0.05 mg/mL were studied. The data were fit using SEDFIT[290] and plotted in GUSSI.

***E. coli* Extract Preparation** Cell pellets for SEC were prepared as previously described[74] but with the following changes: the cell pellet from 50 mL of $^{15}N$ minimal medium culture was resuspended in 1 mL of 20 mM $Na_2HPO_4$, pH 7.0 and frozen overnight. After thawing, cell lysis was completed by sonication. The cell slurry was treated with 10 µg/mL DNase I or RNase A and centrifuged for 30 minutes at 20,000 x g at room temperature. The supernatant contained cytochrome *c* and was loaded onto a SEC column. For agarose gel electrophoresis, DNase I-treated *E. coli* extracts were prepared as above but from 50 mL of saturated cultures (LB medium, $O.D._{600}$ = 6-7).[291] The extract samples were spiked with cytochrome *c* or a mutant at a working concentration of 0.2 mM. Extracts prepared from cells over-expressing cytochrome *c* were unstable and precipitation was observed after incubation for 40 minutes at room temperature or 30 ˚C. The extracts were stable for over 6 hours when stored on ice. Extracts prepared from saturated *E. coli* cells were stable for up to 8 hours when stored on ice or at room temperature. The extract was

less stable after the addition of cytochrome *c i.e.* precipitation was observed after 1.5 hours at 30 ˚C.

**Native Gel Electrophoresis** 20 μL samples of 0.2 mM cytochrome *c* or the charge inverted mutants were analysed in 2% agarose gels (13.5 cm x 14.0 cm) prepared in 20 mM $Na_2HPO_4$, or 20 mM Tris-HCl at pH 7.0. Gels were equilibrated in the electrophoresis buffer for 30 minutes at 4 ˚C prior to running at a constant voltage (100 V) for 30 minutes at 4 ˚C. The buffer temperature was 6-12 ˚C after electrophoresis, indicating that the gels were not heated substantially during the experiment. The gels were imaged directly after electrophoresis using a flatbed scanner. Each gel image was processed identically using Adobe Photoshop. Importantly, changes to the colour balance, saturation, brightness and contrast were applied uniformly across the images. Tools that skew the tonal range or colour balance within an image (*e.g.* exposure, levels or curves) were avoided.

**Size Exclusion Chromatography** The SEC buffer preparation, column and experimental conditions have been described previously (Chapter 2).[74,100]

**Cell Extract NMR Spectroscopy** Cell extracts were prepared for NMR as described above from pellets obtained from 50 mL cultures of cytochrome *c*-expressing cells and also from saturated *E. coli* cultures spiked afterwards with 0.3 mM $^{15}$N-labelled cytochrome *c*. The addition of 35 mM $MgCl_2$ was achieved by spiking the extracts with 6.4 μL of a 3 M $MgCl_2$ stock solution. $^{1}$H, $^{15}$N TROSY-selected HSQC spectra[292] were acquired at 30 ˚C with 16 scans and 64 increments on a Varian 600 MHz Spectrometer equipped with a HCN cold probe. The spectra were processed as previously described (Chapter 2).

**Results and Discussion**

**Cytochrome *c* Charge is Lower than Calculated & Salt-Dependent** REM-MCE was used to determine the electrophoretic mobility of cytochrome *c* in 10 mM BTP at pH 7.0 in the presence of KCl, $KNO_3$, $MgCl_2$ or $K_2SO_4$ at an ionic strength of 100 mM. Table 2 lists the μ, $Z_{eff}$ and $Z_{DHH}$ and corresponding charge (Q) values for cytochrome *c* in each buffer. The $Z_{eff}$ was approximately 3 times less than the $Z_{DHH}$ as predicted from MCE theory.[271]

**Table 2.** Average values for cytochrome *c* μ, $Z_{eff}$, $Z_{DHH}$, $Q_{eff}$ and $Q_{DHH}$ determined at 20 ˚C from at least 2 measurements. The values have an error of < 30 %.

| MCE Buffer Salt (*I* = 100 mM) | μ ($cm^2$/V.s) | $Z_{eff}$ | $Z_{DHH}$ | $Q_{eff}$ (C) | $Q_{DHH}$ (C) |
|---|---|---|---|---|---|
| KCl | $5.82 \times 10^{-5}$ | 1.08 | 3.10 | $1.73 \times 10^{-19}$ | $4.97 \times 10^{-19}$ |
| $KNO_3$ | $5.47 \times 10^{-5}$ | 1.04 | 3.00 | $1.67 \times 10^{-19}$ | $4.81 \times 10^{-19}$ |
| $MgCl_2$ | $4.58 \times 10^{-5}$ | 0.87 | 2.50 | $1.39 \times 10^{-19}$ | $4.01 \times 10^{-19}$ |
| $K_2SO_4$ | ~ 0 | ~ 0 | ~ 0 | ~ 0 | ~ 0 |

The Innovagen Protein Calculator (http://pepcalc.com/protein-calculator.php) was used to calculate the cytochrome *c* valence at pH 7.0 using model p$K_a$ values for the ionisable side chains (Table 3). A value of 8.2 was obtained which deviates substantially from the effective and Debye-Hückel-Henry valence determined by REM (Table 2). The relatively high errors (15-30 %) associated with the measurements reported herein supports the idea that cytochrome *c* charge is low at physiological ionic strengths. Typically, the error associated with MCE measurements is < 10%.[271,273] However, cytochrome *c*'s low charge causes its mobility due to diffusion to counter its mobility in the applied electric field under the standard MCE conditions (see Materials and Methods). This leads to the formation of broader boundaries. The disparity between the calculated and experimentally determined charge values arises from two sources.[16] Firstly, amino acid p$K_a$'s can shift from model values due to the local electrostatic potential effected by protein folding.[293] Secondly, valence calculations are based on the incorrect assumption that only proton binding modulates protein charge.

**Table 3.** Model p$K_a$ values for ionisable groups in proteins.[293,294]

| Amino Acid Side Chain | p$K_a$ |
|---|---|
| N-terminus | 8.96 |
| Arginine | 12.50 |
| Lysine | 10.54 |
| Histidine | 6.04 |
| Glutamic Acid | 4.31 |
| Aspartic Acid | 3.90 |
| C-terminus | 2.16 |

Previously, Winzor reported an effective valence of 9.7 for *Equus caballus* cytochrome *c* in 10 mM acetate at pH 4.7.[295] *E. caballus* and *S. cerevisiae* cytochrome *c* display > 60% sequence homology and are expected to have similar charge structures. The difference in buffer conditions (ionic strength and pH) used herein and by Winzor likely give rise to the dramatic valence differences. This

suggests that cytochrome *c* valence is mitigated by interactions with anions. Accordingly, the valence of cytochrome *c* was determined in different salts at identical ionic strengths (100 mM; Table 2) and its valence decreased in the order $KCl \approx KNO_3 > MgCl_2 > K_2SO_4$. These data illustrate a progressive increase in anion association to cytochrome *c*. Importantly, interactions with the divalent $SO_4^{2-}$ neutralise the protein. The preferential sulfate-cytochrome *c* interaction coincides with the sulfate binding site observed in the crystal structure of cytochrome *c* (PDB ID = 1ycc;[104] Figure 3A). This structural evidence indicates that the sulfate is a site-bound ion. Several cytochrome *c* structures contain site-bound sulfates. For example, the crystal structure of *Crithidia fasciculata* cytochrome *c* is noteworthy for the presence of seven site-bound sulfates which mediate trimer formation in the asymmetric unit (Figure 4).[296] The neutralisation of cytochrome *c* by interactions with sulfate ions is reminiscent of cytochrome *c* camouflage effected by weak interactions with *p*-sulfonatocalix[4]arene.[106] Interestingly, a similar, lysine-rich site

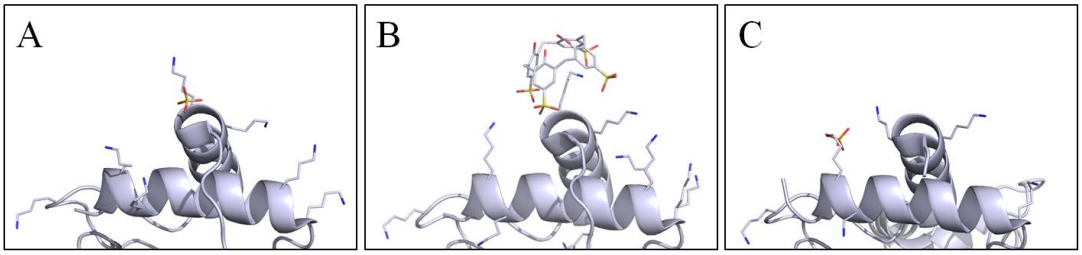



**Figure 3.** Crystal structures of a lysine-rich patch of *S. cerevisiae* or *E. caballus* cytochrome *c* showing similar **A.** sulfate (1ycc), **B.** *p*-sulfonatocalix[4]arene (3tyi) and **C.** phosphate binding sites (3nbs)[298]. Atoms are coloured as per Figure 1 but with yellow representing sulfur. Water molecules are omitted for clarity.

is occupied by the sulfate anion (PDB code = 1ycc; Figure 3A) or a sulfonate group from *p*-sulfonatocalix[4]arene (PDB code = 3tyi;[106] Figure 3B) in both crystal structures. Efforts to determine the cytochrome *c* valence by MCE in a buffer containing 6.7 mM ($I$ = 100 mM) *p*-sulfonatocalix[4]arene were hampered by the fact that the calixarene interacted strongly with the confining membrane, obstructing ion flux. Similarly, the effects of the physiologically-relevant phosphate and polycarboxylate anions could not be studied.[271] However, the data indicate the significance of direct, site-specific sulfate interactions with the cytochrome *c* surface. Free solution electrophoresis also identified preferential interactions between sulfate and the cationic proteins RNase[276] and lysozyme from hen egg white[246] or T4

bacteriophage.[297] The significance of tight, direct sulfate interactions with cationic surface patches is thus suggested.

Owing to their tetrahedral geometry, similar radii[299] and anionic nature, sulfates and phosphates tend to interact with similar protein surface patches (Figure 5).[300,301] For instance, the $C^{\alpha}NN$ structural motif[302] recognises sulfates and phosphates as well as cofactor and nucleotide phosphates.[300] This motif consists of a tripeptide segment which co-ordinates the anions using three invariant main chain atoms. The first invariant position is occupied by the $C^{\alpha}$ atom of the first segment residue while the second and third positions are held by the backbone N atoms of the second and third segment residues. The main chain N-H forms hydrogen bonds with separate phosphate oxygens while the $C^{\alpha}$-H engages in a weak C-H…O hydrogen bond with the proximal phosphate oxygen. The $C^{\alpha}NN$ binding motif is often located at the amino terminal end of an α-helix which has a positive dipole moment.[302] Interestingly, a $C^{\alpha}NN$ binding motif involving Lys4 recognises the sulfate/sulfonate groups in the *S. cerevisiae* cytochrome *c* crystal structures (Figure 5A), suggesting that this site is optimised for anion recognition. Although phosphates have not been observed in a *S. cerevisiae* crystal structure, the domain-swapped dimer of *E. caballus* cytochrome *c* is noteworthy for the presence of a phosphate ion bound to the lysine-rich cytochrome *c* surface (Figure 3C and 5B).[298] The phosphate is co- ordinated by a salt bridge with

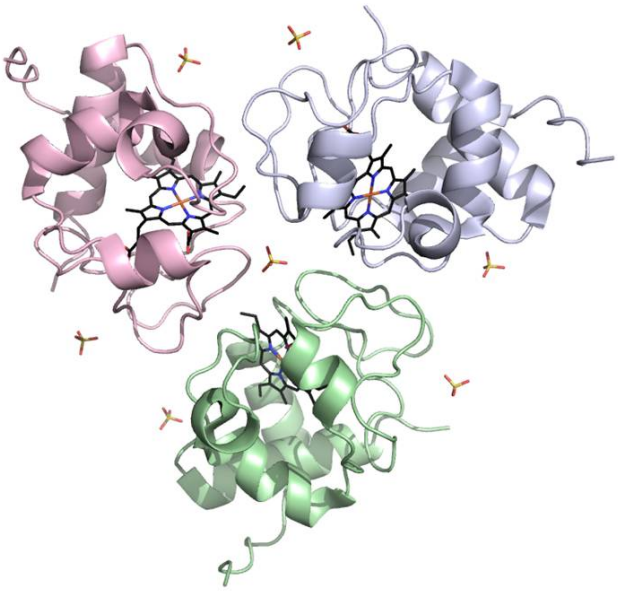


**Figure 4.** The asymmetric unit of *C. fasciculata* cytochrome *c* contains three protein molecules and seven sulfates (PDB ID = 2YK3).[296] The central sulfate appears to act as a bridging ion via salt bridge interactions to Lys90 in each protein. The haem carbons are coloured black.

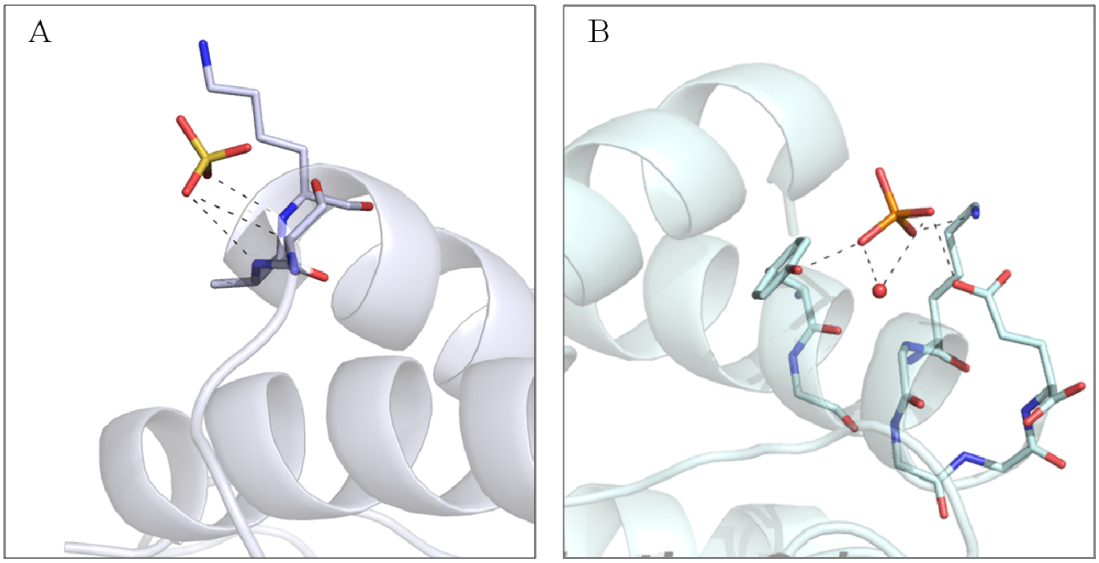



**Figure 5**. Anion binding sites in **A**. *S. cerevisiae* and **B** *E. caballus* cytochrome *c*. **A** $C^{\alpha}NN$ motif recognizes the sulfate in the *S. cerevisiae* structure. **B** At a proximal site in the *E. caballus* cytochrome *c* structure,[298] the phosphate is coordinated by side chain hydrogen bond donors and a water molecule. Dashed lines indicate interatomic distances < 3.5 Å.

K100 and hydrogen bonds with the side chains of Y97, E104 and a water molecule (PDB code = 3nbs;[298] Figure 5B). This site is in close proximity to the sulfate binding site in the *S. cerevisiae* structure (S-P distance = 11.8 Å in the aligned structures), indicating that the lysine-rich cytochrome *c* surface patch can recognise phosphates such as the phospholipid phosphate groups in the mitochondrial IMS.[303] Indeed, two phospholipid binding sites have been postulated for cytochrome *c*[303] with the lysine-rich surface patch is expected to interact with anionic phospholipid phosphates.

**Valence of Cytochrome *c* and its Charge-Inverted Mutants** Protein characterisation under conditions of reduced charge can reveal the importance of electrostatics in governing protein behaviour.[74,106,276,297] While pH changes are typically used to modulate protein charge, this can incidentally alter other physicochemical protein properties (*e.g.* hydrodynamic radius) producing results that cannot be wholly linked to protein surface charge.[276,304] Charge inverted mutants offer an elegant alternative by mutating selected charged residues on the protein scaffold to residues of opposite charge.[297] The effects of charge inversion on cytochrome *c* interactions in concentrated cell extracts were previously explored by Crowley and co-workers.[74] Replacing cationic residues for Glu reduced the extent to which the protein interacted with *E. coli* macromolecules. Indeed, the implications of

charge inversion are apparent during purification. The more negatively charged the cytochrome *c* mutant, the weaker it interacts with the negatively charged carboxymethyl cellulose (CM) resin. WT cytochrome *c* elutes from a CM column at ~300 mM NaCl, 20 mM $KH_2PO_4$ pH 6.0. The single and double mutants elute at ~200 and ~100 mM NaCl, respectively. The triple and quadruple mutants interact weakly with the carboxymethyl resin even in the absence of NaCl at 20 mM $KH_2PO_4$, pH 5.0 and elute when the pH is increased to 6.0. Given the obvious importance of charge in governing cytochrome *c* behaviour, the μ of each mutant was quantified by MCE in 100 mM KCl, 10 mM BTP, pH 7.0 (Table 4). KCl was selected as an electrolyte because the electrophoretic transport properties of $K^+$ and $Cl^-$ are equal in magnitude but opposite in sign thus preventing detrimental electroosmotic flows.[297]

**Table 4.** $Z_{cal}$ and the MCE-determined μ with corresponding $Z_{eff}$, $Z_{DHH}$, $Q_{eff}$ and $Q_{DHH}$ for cytochrome *c* and the charge-inverted mutants.

| Cytochrome *c* | $Z_{cal}$ | μ ($cm^2/V.s$) | $Z_{eff}$ | $Z_{DHH}$ | $Q_{eff}$ (C) | $Q_{DHH}$ (C) |
|---|---|---|---|---|---|---|
| WT | 8.2 | $5.82 \times 10^{-5}$ | 1.08 | 3.10 | $1.73 \times 10^{-19}$ | $4.97 \times 10^{-19}$ |
| R13E | 6.2 | $3.74 \times 10^{-5}$ | 0.71 | 2.02 | $1.14 \times 10^{-19}$ | $3.24 \times 10^{-19}$ |
| R13E/K73E | 4.2 | $2.42 \times 10^{-5}$ | 0.46 | 1.32 | $0.737 \times 10^{-19}$ | $2.11 \times 10^{-19}$ |
| R13E/K73E/K87E | 2.2 | ~ 0 | ~ 0 | ~ 0 | ~ 0 | ~ 0 |
| R13E/K73E/K87E/K100E | 0.2 | ~ 0 | ~ 0 | ~ 0 | ~ 0 | ~ 0 |

Table 4 summarises the μ, Z and Q values for cytochrome *c* and the charge-inverted mutants. The calculated valences ($Z_{cal}$) are also listed. Note that exact μ values for the triple and quadruple mutants could not be determined by REM or SSE. Both experiments rely on the formation of a macroion concentration gradient against a membrane boundary in response to an electric field.[129,273,274] Macroions with an average charge ~0 do not form electrophoretic concentration gradients. Their mobility is instead governed by thermal motion which disfavours concentration gradient formation. A range of electric fields were tested in MCE-REM studies of triple and quadruple cytochrome *c* mutants (see Materials and Methods) but a concentration gradient never formed. Similarly, a concentration gradient did not form during SSE experiments. Together, these findings suggest that the triple and quadruple mutants have an average charge of ~0. Proteins with a charge ~0 tend to self-associate in solution.[305] Sedimentation-velocity analytical ultracentrifugation

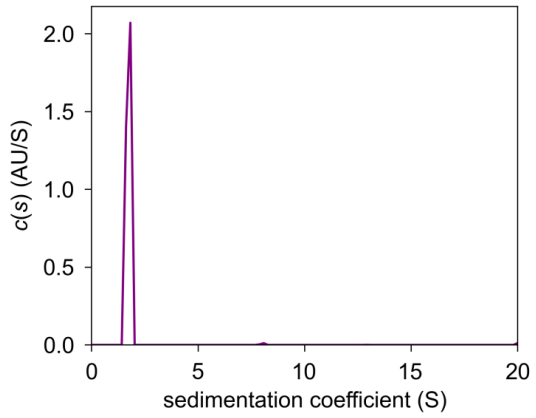



**Figure 6.** A representative sedimentation coefficient distribution acquired from a 0.5 mg/mL sample of the quadruple cytochrome *c* mutant in 10 mM BTP, 100 mM KCl, pH 7.0. The c(s) distribution shows a single peak at 1.69 s with no evidence of aggregation at higher s values.

was therefore used to determine the oligomeric state of WT cytochrome *c* and the double and quadruple mutants. Concentrations of 0.5, 0.15 and 0.05 mg/mL were studied for each protein in 10 mM BTP, 100 mM KCl, pH 7.0. In each case a single species was identified with an average sedimentation coefficient (s) of 1.74 that corresponds to the molecular weight of the protein monomers (Figure 6). This confirms that the charges of monomeric proteins are probed in this MCE study. The calculated and experimentally-determined valence for each protein reveals the same overall pattern: WT cytochrome *c* has the highest valence followed by the single, double and triple/quadruple mutants respectively. However, the calculated valences are substantially greater than the measured values and overestimate the extent to which the net valence of cytochrome *c* is reduced by each mutation. For instance, the $Z_{cal}/Z_{DHH}$ ratio for cytochrome *c* and the single and double mutants are 2.65, 3.07 and 3.18 respectively. The variation between the ratios indicates the poor correlation between predicted and measured valences, which are likely due to disparate charge distributions in the protein. The limitations of charge/valence predictions based on the primary structure and model p$K_a$ values are thus highlighted.

**Specific Ion Effects on Cytochrome *c* Interactions in *E. coli* Extracts** SEC was used to explore specific ion effects on WT cytochrome *c* interactions in *E. coli* cell extracts. In an SEC buffer containing 20 mM $Na_2HPO_4$, 35 mM $K_2SO_4$ ($I$ = 100 mM), pH 7.0 cytochrome *c* eluted mainly in the high molecular weight fractions (50-65 mL; Figure 7). Some cytochrome *c* is evident in the low molecular weight fractions (70 and 75 mL; Figure 7). A similar elution profile was obtained for SEC experiments performed in buffers containing 100 mM KCl and $KNO_3$ but there was no cytochrome *c* in the 70 and 75 mL fractions. Thus, it appears as though most of the charged-based interactions of cytochrome *c* with *E. coli* macromolecules are not

disrupted in the presence of 35 mM $K_2SO_4$. Importantly, cytochrome *c* is not naturally expressed in *E. coli* and therefore has no cognate partners in the extracts. It is likely that the high effective concentration of negative charge found at macroanionic surface patches combined with the high surface-to-volume ratios in cell extracts renders cytochrome *c* interactions with macromolecules more favourable than cytochrome *c*-sulfate interactions. Interestingly, in SEC buffers containing 35 mM $MgCl_2$ ($I$ = 100 mM) cytochrome *c* eluted exclusively in the low

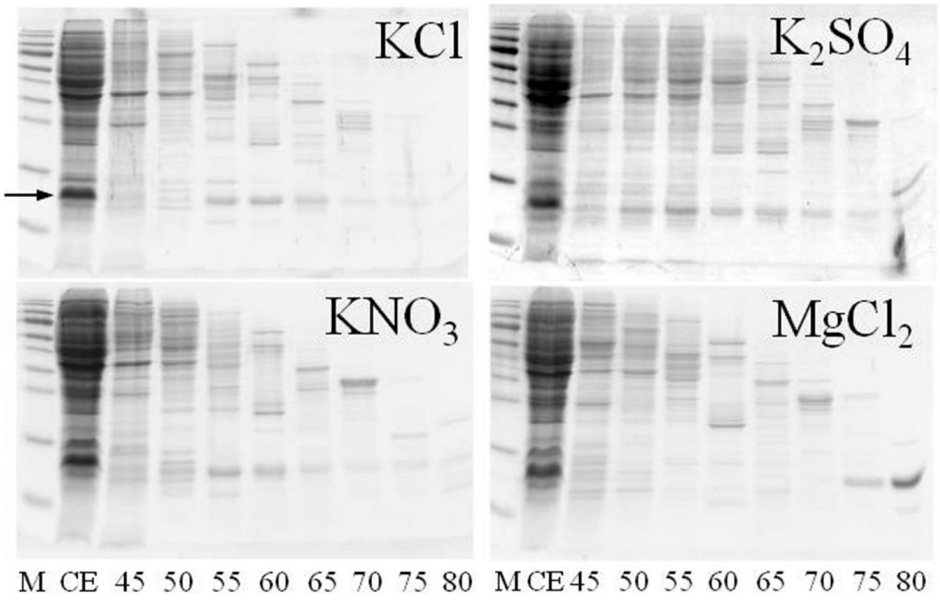



**Figure 7.** SEC elution profiles for cytochrome *c*-containing cell extracts that were pre-treated with DNase I. The buffers were 20 mM $Na_2HPO_4$ or Tris-HCl (pH 7.0) and contained KCl, $KNO_3$, $MgCl_2$ or $K_2SO_4$ at an ionic strength of 100 mM. The arrow marks the migration position of cytochrome *c*. The gel lanes are labelled; MM: molecular weight marker; CE: cell extract; 45–80: fraction volume (mL).

molecular weight fractions (75 and 80 mL). When these fractions were combined and exchanged into a typical NMR buffer (20 mM $KH_2PO_4$, 75 mM NaCl, pH 6.0), the spectrum for reduced cytochrome *c* was observed (Figure 8). The disruption of cytochrome *c* interactions with *E. coli* macromolecules in the presence of $Mg^{2+}$ is consistent with the liberation of ΔTat-GB1 from RNA-containing complexes in $Mg^{2+}$-containing buffers (Chapter 2). Owing to the large number of lysine residues at the cytochrome *c* surface, it is expected that lysine-phosphate and/or lysine-carboxylate interactions drive cytochrome *c* complexation with nucleic acids and/or proteins, respectively. To elucidate the origin of its stickiness in extracts, cytochrome *c* interactions with nucleic acids were probed by pretreating the extracts with

nucleases (RNase A or DNase I) prior to SEC in a buffer of 20 mM $Na_2HPO_4$, 100 mM KCl, pH 7.0. In each case, cytochrome *c* eluted in the high molecular weight fractions (Figure 7; RNase A-treated extract data not shown). The inability to disrupt cytochrome *c*-containing complexes by digesting nucleic acids indicates that it interacts mainly with anionic proteins. Thus, perhaps numerous lysine-carboxylate salt-bridge interactions mediate complexation. Although protein-protein interfaces are typically devoid of lysine residues,[83] the role of lysine in driving non-specific interactions is indicated here.

Interestingly, cytochrome *c* was never detectable in *E. coli* extracts by NMR spectroscopy, even when the extracts were spiked with 35 mM $MgCl_2$ ($I$ = 100 mM), DNase I and RNase A and studied by the TROSY-HSQC experiment which is

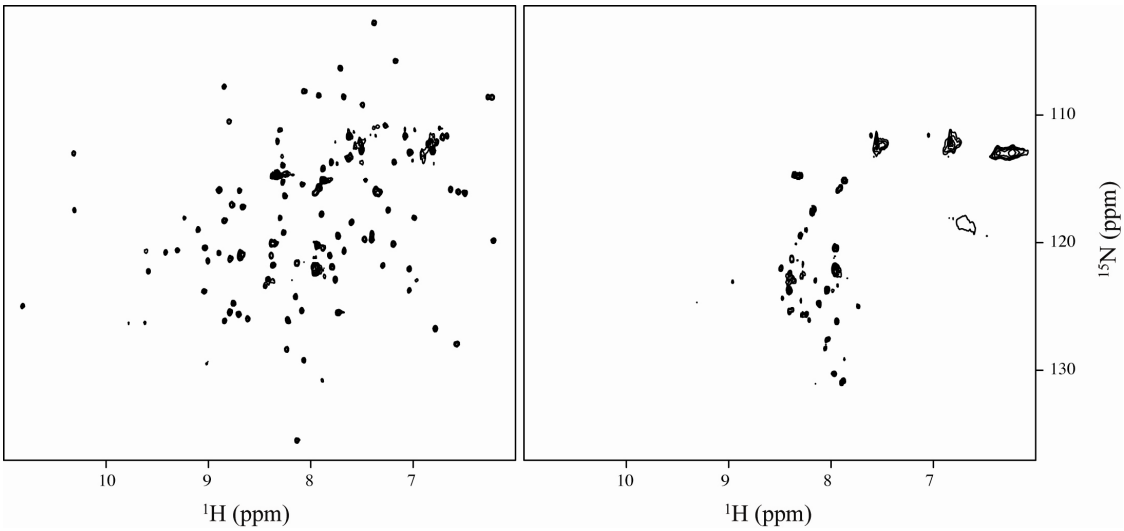



**Figure 8.** $^1$H, $^{15}$N TROSY-HSQC spectra obtained from: (left) the concentrated 75-85 mL fractions after SEC of cytochrome *c*-containing extracts in 20 mM $KH_2PO_4$, 35 mM $MgCl_2$ pH 7.0 and (right) concentrated extracts pretreated with RNaseA and DNase I and spiked with $MgCl_2$ to a final concentration of 35 mM ($I$ = 100 mM).

optimised for the detection of large biomolecules (Figure 8).[306] This result indicates that, in nuclease and $MgCl_2$-treated extracts, cytochrome *c* interacts with *E. coli* macromolecules to form high molecular weight assemblies that tumble too slowly to be detected by NMR. Notably, SEC involves sample separation. Perhaps cytochrome *c*'s interactions in the extracts may be weakened but not fully disrupted in the presence of $Mg^{2+}$ and the resulting transient (yet extensive) complexes cannot withstand exposure to the SEC column or electrophoresis.

**Electrophoresis Confirms Preferential Interactions of Cytochrome *c* and the Mutants** Previously, Crowley and co-workers used SEC and NMR to demonstrate that charge-inverted mutations decreased cytochrome *c* stickiness in *E. coli* extracts.[74] Although charge is a system property, the dilute solution MCE data acquired at a physiological ionic strength reveals novel aspects of the charge structure of cytochrome *c* and the mutants. For example, although cationic, WT cytochrome *c* charge is low and cytochrome *c* is preferentially solvated[307] by anionic proteins in *E. coli* extracts. The triple mutant is charge-neutral and is therefore expected to be the least soluble of the mutants and also preferentially solvated by macromolecules. Intriguingly, cell extract SEC studies revealed that the triple mutant elutes in the low molecular weight fractions indicating that is not preferentially solvated by other macromolecules, despite its low charge.[74] Native gel electrophoresis was used to explore the charge and interactions of cytochrome *c* and the mutants under conditions of greater physiological relevance (Figure 9). Protein migration was compared in dilute solution and in *E. coli* extracts in four different electrophoresis buffers. In keeping with the MCE data, pure cytochrome *c* always migrated toward the cathode with a migration distance that was short (*e.g.* 5 mm in 20 mM $Na_2HPO_4$, pH 7.0) despite the relatively high applied voltage and long measurement times (see Materials and Methods). Importantly, Crowley *et al.* found that pure $^{15}N$-labelled cytochrome *c* encapsulated in 1.5% agarose gels yielded a HSQC spectrum identical to that for cytochrome *c* in buffer, indicating that cytochrome *c*-does not interact with agarose gels.[308] Thus, cytochrome *c*-agarose interactions are unlikely to contribute to the poor migration of the protein. These data suggest that cytochrome *c* is weakly cationic (Figure 9). Flavodoxin, by comparison, is a highly anionic protein that migrates toward the anode with a migration distance that is ~ three times greater than that of cytochrome *c* (Chapters 5 and 6). The migration distance of cytochrome *c* was shorter in the presence of buffers containing $K_2SO_4$ and $KNO_3$ at an ionic strength of 100 mM compared to buffer alone (Figure 9). This result demonstrates the mitigation of cytochrome *c* charge by anions. The bands due to cytochrome *c* and the mutants were more diffuse in the presence of $KNO_3$ suggesting that $NO_3^-$ binds in a territorial manner to the proteins, producing several distinct charge species (Figure 9). The reduction of cytochrome *c* charge effected by charge inversion is also evident and the migration of triple and quadruple mutants is poor in the low ionic strength buffer and in the presence of $K_2SO_4$ and

$KNO_3$ (Figure 9). These data indicate that the triple and quadruple mutants are charge neutral (Figure 9). In the presence of $Mg(NO_3)_2$ cytochrome *c* and the mutants migrated toward the cathode. Indeed WT cytochrome *c* migrated further in a buffer containing 35 mM $Mg(NO_3)_2$ than in a buffer lacking salt (Figure 8). This finding suggests that $Mg^{2+}$ can bind to anionic residues in WT cytochrome *c* and the mutants, rendering the proteins more cationic.

Cytochrome *c* migration was drastically different in dilute solution and in cell extracts (Figure 9). Cytochrome *c* migrated toward the anode in concentrated *E. coli* extracts in the low ionic strength buffer and in the presence of $KNO_3$ and $K_2SO_4$ (at $I$ = 100 mM; Figure 9). This result supports earlier evidence suggesting that cytochrome *c* interacts extensively with anionic proteins in *E coli*.[74] The resulting assemblies have a net negative charge and are reminiscent of the anionic macromolecular clusters expected to prevail *in vivo*.[61,63] The diffuse nature of the cytochrome *c* band in cell extracts confirms that cytochrome *c* interacts non-specifically with numerous anionic proteins. The bands due to the anionic, cytochrome *c*-containing assemblies are generally less diffuse in the higher ionic strength buffers (where salt $I$ = 100 mM) than in the low ionic strength buffer (where salt $I$ = 0 mM) indicating that the anionic assemblies have a low net charge at physiological ionic strengths (Figure 9). WT cytochrome *c* interactions were never disrupted at higher ionic strengths, emphasising the significance of screened charge-charge interactions in driving macromolecular interactions in volume-occupied solutions. In the low ionic strength buffer, the single mutant migrates toward the anode as part of a macroanionic complex in the *E. coli* extract. When the buffers also contained $KNO_3$ and $K_2SO_4$, low concentrations of the unbound single mutant migrate toward the cathode also. This result indicates that the interactions between the single mutant and anionic *E. coli* proteins are weaker than those of WT cytochrome *c* at physiological ionic strengths. An even greater concentration of the double mutant migrated toward the cathode in low and higher ionic strength electrophoresis buffers indicating that its interactions with *E. coli* proteins were substantially weaker than those of the single mutant. The migration of the triple and quadruple mutants were closely similar in both dilute and cell extract samples irrespective of the electrophoresis buffer. This result confirmed that these mutants do not interact with *E. coli* macroions. Thus, they are preferentially hydrated in the

extracts. Importantly, the triple mutant contains 20 cationic residues (arginine, lysine, histidine) and 14 anionic residues (aspartate and glutamate). The heterogeneous charge distribution of the triple mutant surface (Figure 5) likely renders the protein highly soluble, despite its low net charge. Notably, glutamate and aspartate are among the most strongly hydrated side chains.[309] Indeed, the solubility of RNAse Sa was higher for the T76E mutant than the T76R or T76K mutants at pH 4.25 (protein p*I*) and 7.0.[310] This indicates that glutamate contributes more favourably to protein hydration than lysine or arginine. Relative to WT cytochrome *c*, the increasingly negative charge of the triple and quadruple mutant surfaces probably causes them to electrostatically repel anionic macromolecules in the cell extracts, rendering them preferentially hydrated in extracts.

In a buffer containing 35 mM $Mg(NO_3)_2$ ($I$ = 100 mM) the migration of WT cytochrome *c* and the mutants is similar in extracts and in dilute samples (Figure 9). This indicates that, in extracts, the interactions of WT cytochrome *c* and the mutants are disrupted in the presence of $Mg^{2+}$. The more diffuse nature of the red bands in the extract lanes indicate that several cationic species exist for cytochrome *c* and each mutant. This result may suggest that cytochrome *c* (and mutant) interactions with macromolecules persist. In the cell extract samples, heavy precipitates can be observed in the lane wells and migrating toward the anode in the $Mg^{2+}$-containing buffer. This suggests that $Mg^{2+}$ interactions with the anionic clusters in the extracts reduce the net charge of the assemblies which no longer repel each other electrostatically leading to haphazard aggregation (and cytochrome *c* liberation).

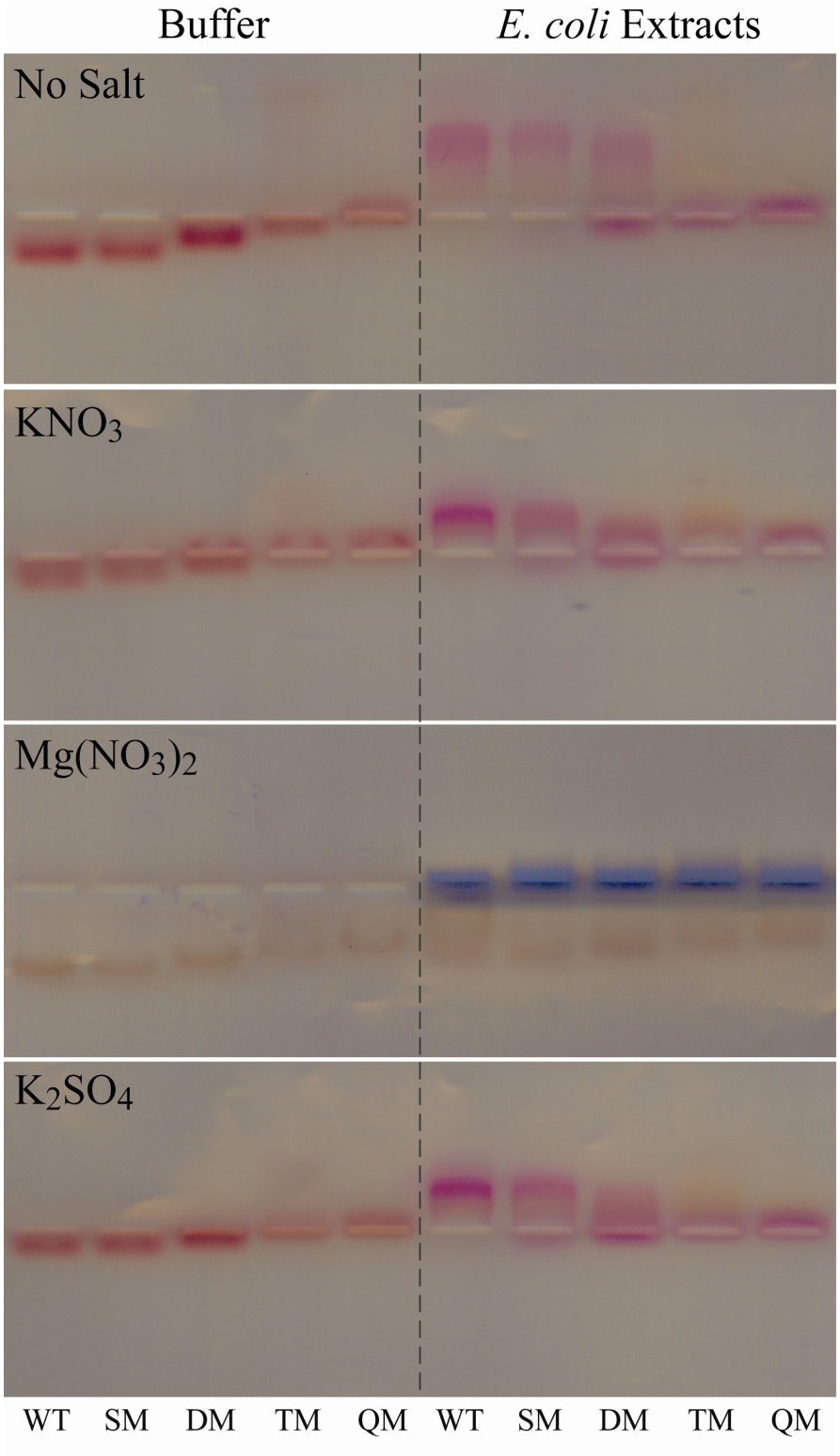



**Figure 9.** Native gels showing the migration of cytochrome *c* (WT) and the single (SM), double (DM), triple (TM) and quadruple mutants (QM) in buffer and in *E. coli* extracts (sample volume = 20 µL). The electrophoresis buffer contained 20 mM $Na_2HPO_4$ or Tris-HCl at pH 7.0 and the indicated salt at an ionic strength of 100 mM.

**Broader Implications**

Recently, NMR studies of the interaction between *E. caballus* cytochrome *c* and a 41 bp oligonucleotide revealed that the complex was attenuated at increased pH and ionic strengths.[109] These findings indicate that charge-charge interactions facilitate cytochrome *c*-DNA complexation.[109] Moreover, numerous cytochrome *c* lysines (K7, K8, K72, K73, K86, K87 and K100) were involved in the interaction with DNA,[109] suggesting the significance of lysine-phosphate interactions (*i.e.* charge-charge and salt bridge) in driving complex formation. Non-specific protein-DNA interactions, which typically precede specific complexation, are also characterised by a high proportion of charge-charge interactions.[83] Thus, the cytochrome *c*-DNA system may prove useful for studies of non-specific protein-DNA interactions and their role in DNA searching (*i.e.* facilitated diffusion)[311] and recognition.

As a small, lysine-rich protein, cytochrome *c* is a useful model for histones which form octameric assemblies that bind to DNA. The resulting nucleosomes regulate gene expression *in vivo*.[312,313] As universal eukaryotic DNA compaction factors, nucleosomes bind to DNA with relatively low sequence specificity[314] and these assemblies are highly dynamic.[315] The role of histone charge in regulating nucleosome stability is well-established.[316–318] For instance, acetylation of lysine 53 in the globular core of *S. cerevisiae* histone H3 facilitates binding of the switch/sucrose non-fermentable (SWI/SNF) nucleosome remodelling complex and altered gene activity.[319] Structural analysis of the yeast nucleosome core particle shows that lysine 53 of histone H3 is located at the entry and exit points of the wrapped DNA superhelix.[320] This indicates that the Lys53-DNA phosphate interaction is disrupted upon Lys53 acetylation (due to lysine charge neutralisation) which is expected to induce partial DNA unwrapping prior to the recruitment of the SWI/SNF complex which subsequently remodels the nucleosome.

Sequence analysis of globular histone folds[321] from over 500 proteins across 154 species' revealed that sequence variation within each histone class is relatively low and tends to preserve physicochemical properties.[322] Accordingly, lysine and arginine residues typically comprise ~20% of histone sequences.[323] Although amenable to CE studies,[324] histone charge has never been determined by CE or any other technique. However, lysine and arginine residues comprise ~18% of the *S.*

*cerevisiae* cytochrome *c* sequence which is closely similar to that of histones. The MCE-determined cytochrome *c* charge values may therefore also represent histone charge under physiologically-relevant ionic strengths. For instance, histone charge may also be substantially lower than calculated. Although weak and screened charge-charge interactions are significant in crowded and confined environments, a single charge inverted mutation at the lysine-rich cytochrome *c* binding site had a notable impact on its stickiness.[74] This supports the idea that a single, post-translational modification (*e.g.* acetylation, phosphorylation) may drastically attenuate histone-DNA interactions, destabilising the nucleosome.

**Conclusions**

It is well-established that the charge of cytochrome *c* directs interactions with cognate and non-specific partners.[18,21–23] MCE was used to determine the valence of cytochrome *c* in four different buffers containing salts at the physiologically-relevant ionic strength of 100 mM. Charge-inverted (R/K→E) cytochrome *c* mutants were also characterised by MCE. The effective valence of cytochrome *c* was substantially lower (~eightfold) than predicted, was salt-dependent, and decreased in the following order: KCl ≈ $KNO_3$ > $MgCl_2$ > $K_2SO_4$. In the $K_2SO_4$–containing buffer, cytochrome *c* was charge-neutral. This result indicates that sulfates interact directly with cytochrome *c* forming a neutral macroion. Other free solution electrophoresis studies have identified preferential cationic protein-sulfate interactions,[246,276] indicating the significance of tight, direct sulfate interactions with cationic surfaces.

Cytochrome *c* is preferentially solvated by anionic macromolecules in physiological solutions.[74] Native gel electrophoresis confirmed that cytochrome *c* interacted pervasively with *E. coli* macromolecules resulting in macroanionic assemblies akin to the clusters expected to prevail in the cytoplasm.[63] A previously described SEC technique[74] revealed that cytochrome *c* interacts preferentially with *E. coli* macromolecules and that the resulting assemblies were not disrupted by the presence of $SO_4^{2-}$. Thus, although the lysine-rich surface patch of cytochrome *c* recognises sulfates, non-specific interactions between the cationic residues of cytochrome *c* and clusters of negatively charged residues on anionic protein surfaces prevail in the crowded cell extracts. Cytochrome *c* interactions were not disrupted when DNA or RNA were digested, indicating that cytochrome *c* does not interact

preferentially with these macromolecules. Thus, lysine-glutamate and/or lysine-aspartate interactions likely drive cytochrome *c* complexation with anionic proteins. Although lysine is not frequently observed in specific protein-protein interfaces,[258] our data indicates that lysine may be important in facilitating weak, non-specific interactions (akin to quinary interactions) in biological environments. In keeping with earlier studies (Chapter 2),[100] the disruptive effect of $Mg^{2+}$ was observed by SEC and electrophoresis in extracts containing cytochrome *c*. The triple and quadruple mutants were approximately charge neutral and unsticky in extracts. These mutants therefore remain preferentially hydrated in extracts. This finding indicates that low net charge does not explicitly render proteins preferentially solvated by neighbouring macromolecules under native-like conditions. The increased negative electrostatic surface potential of the triple and quadruple mutants causes them to repel the anionic extract macromolecules. Thus, an intricate interplay of electrostatic forces governs the surface properties and non-specific interactions of the cytochrome *c* mutants. The triple and quadruple mutant data emphasise the role of screened, short-range charge-charge repulsions in preventing erratic aggregation in physiological environments. The need for more advanced models describing protein surface charge density/distribution and specific ion effects is apparent.

# Chapter 4

# Simple and Inexpensive Incorporation of $^{19}$F-Tryptophan for Protein NMR Spectroscopy

The material in this chapter has been published in part as:

Crowley PB, **Kyne C**, Monteith WB. (2012) Simple and Inexpensive Incorporation of $^{19}$F-Tryptophan for Protein NMR Spectroscopy. *Chem. Commun. 48*, 10681–10683.

**Abstract**

Fluorine-containing amino acids are valuable probes for biophysical studies of protein properties. $^{19}$F-labeled protein production can require expensive reagents, time-consuming genetic manipulation and compromised expression systems. Herein, a simple method of $^{19}$F-tryptophan labelled protein production is presented. This inexpensive approach involves the use of *E. coli* BL21 (DE3), the workhorse of protein production, which can utilise fluoroindole for the inducible over-expression of proteins containing $^{19}$F-tryptophan.

## Introduction

**Protein Labelling for NMR Spectroscopy** Advances in protein labelling strategies have led to a surge in isotopic labelling schemes, expanding the frontiers of protein NMR studies *in vitro* and in living cells.[227,325–329] Strategies for protein/peptide labelling have been described for several *in vitro* systems including cell-free synthesis,[330] total chemical synthesis[331] and chemical post-modification.[332] However, i*n vivo* strategies are often preferred because they are relatively inexpensive and facilitate in-cell protein characterisation. Heterologous over-expression of labelled proteins has been described in *E. coli*,[73,74,333,186] *Brevibacillus choshinensis,*[334] *Pichia pastoris,*[335] *Baculovirus*-infected insect cells[336] and mammalian cells.[235,337] Due to its well characterised genetics and the large number of strains and cloning vectors, *E. coli* remains the most extensively used system for the efficient production of recombinant proteins.

Uniformly labelled proteins can be produced in *E. coli* cells cultured on $^{13}$C and/or $^{15}$N-enriched minimal medium. Amino acid-selective isotope labelling can also be achieved by protein over-expression in *E. coli* grown on minimal medium supplemented with 50-100 mg/L of the labelled amino acid or biosynthetic precursor(s).[227] Non-uniform isotope labelling can simplify NMR spectra by eliminating peaks from unlabeled residues. Accordingly, complications in peak assignments that result from peak overlap or chemical shift degeneracy are reduced. The protein size limit of solution state NMR is also increased. For instance, $^{13}$C-methyl groups have emerged as particularly useful NMR probes due to their high mobility, inability to undergo chemical exchange with water and the high intrinsic sensitivity conferred by the three protons coupled to a single carbon nucleus.[227,328] Indeed NMR studies of the dynamics and interactions of the 20S proteasome core particle (620 kDa) were facilitated by the selective $^{13}CH_3$-enrichment of isoleucine, leucine, and valine side chains, high levels of protein deuteration (which reduces dipolar relaxation, enhancing resolution and sensitivity) and the use of TROSY-type pulse sequences.[338] Similarly, $^{13}CH_3$ Met-labelled calmodulin was used to characterise interactions with the drug phenoxybenzamine hydrochloride inside living *E. coli* cells.[227]

Amino acid-selective isotope labelling in *E. coli* is suited to amino acids which are the terminal products of their respective biosynthetic pathways (*e.g.* lysine, arginine, histidine, cysteine, methionine).[326] Aspartate, glutamate, glutamine, serine, glycine and threonine are precursors for the biosynthesis of other amino acids and the label can be transferred between amino acids to an undesired site (isotope scrambling).[326,339] Similarly, scrambling can occur when transaminases transfer the α-amino group from one amino acid to the α-keto precursors of other amino acids. Amino acid biosynthesis is regulated by feedback inhibition, however, and amino acid scrambling can be attenuated by supplementing the growth medium with 50-100 mg/L of the other 19 unlabelled amino acids, which suppresses endogenous amino acid biosynthesis.[329,340] Auxotrophic strains, which are rendered defective in the synthesis of a specific amino acid by genetic lesions,[326] can be used in the production of selectively labelled protein.[326,329] Auxotrophic *E. coli* strains must be grown in media enriched in the amino acid that cannot be biosynthesised. Some auxotrophic

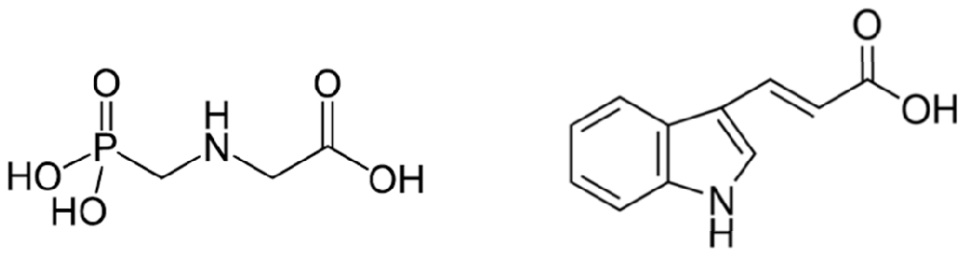



**Figure 1.** Structures of glyphosate (left) and indoleacrylic acid (right)

strains have slow growth rates and low yields[2] and the construction of multiply auxotrophic strains normally results in reduced cell viability.[329] A limited number of amino acid biosynthetic pathways can be suppressed by specific inhibitors leading to induced auxotrophy.[325,341] For example, glyphosate inhibits 5-enolpyruvyl shikimic acid-3-phosphate synthase which is essential for aromatic amino acid biosynthesis (Figure 1).[325] Indoleacrylic acid blocks tryptophan synthase (Figure 1).[342,343] Although mainly reported for *E. coli*, fluorinated recombinant proteins have also been expressed in *Lactococcus Lactis* by inducing aromatic amino acid auxotrophism with indoleacrylic acid or glyphosate.[344]

**Protein Labelling with Unnatural Amino Acids** The incorporation of unnatural (*e.g.* fluorinated, nitrobenzyl, pyridyl) amino acids by the *E. coli* expression machinery is typically achieved using an orthogonal tRNA/aminoacyl-tRNA synthetase pair that encode the unnatural amino acid.[327] The rare amber codon is recognised by the orthogonal tRNA and is introduced at a single site in the gene for the target protein.[339,345] For example, the $tRNA_{CUA}$/Tyr tRNA synthase pair were expressed in *E. coli* and used in the biosynthesis of trifluoromethyl-L-phenylalanine-labelled nitroreductase.[346] This technology[347,348] has led to a resurgence in NMR studies of $^{19}$F-labelled proteins, which were described as early as 1974.[349] Other strategies for $^{19}$F-labelling in *E. coli* routinely involve the incorporation of fluorinated aromatic amino acids by manipulating the Shikimate pathway of aromatic amino acid biosynthesis.[325,329,342] Fluorinated aromatic amino acids are useful NMR probes as the number of aromatic amino acids in proteins is usually low, leading to facile spectral interpretation. Moreover, *E. coli* strains that are auxotrophic for the aromatic amino acids have been developed and inhibitors exist for several enzymes in the aromatic amino acid biosynthesis pathway.[326] Additionally, *m*-$^{19}$F-tyrosine, 4-, 5-, and 6-$^{19}$F–labeled tryptophan, and *m*-, *o*-, and *p*-$^{19}$F-phenylalanine are commercially available.[350]

**$^{19}$F NMR of Proteins** Several properties of $^{19}$F make it an attractive NMR-active (spin ½) nucleus. Firstly, fluorine is rarely present in biological systems and therefore $^{19}$F spectra have no background signals.[351] Fluorine is small (van der Waals radius = 1.47 Å) and protein fluorination is therefore unlikely to induce large structural changes. $^{19}$F is 100% naturally abundant and has a high gyromagnetic ratio making it highly NMR sensitive on a par with hydrogen.[347,348] The $^{19}$F nucleus has a broad chemical shift range (> 300 ppm) and is sensitive to subtle changes in the local chemical environment including protein tertiary structure. For instance, Oldfield *et al.* showed that six 4-fluorotryptophan residues in hen egg white lysozyme (HEWL; 14.3 kDa) were well-resolved and spanned 16.8 ppm.[352] By comparison, only a single, broad peak was observed in guanidinium-chloride denatured 4-fluorotryptophan HEWL.[352] This finding emphasises the fact that hydrophobic environments generally deshield fluorine nuclei, eliminating chemical shift degeneracies. Notably, the 5 or 6-fluorotryptophan residues of native, folded HEWL spanned just 2.8 and 2.4 ppm, respectively.[352] $^{2}$H solvent isotope shift studies (in

which solvent exposed $^{19}F$ nuclei experience an upfield shift from the corresponding chemical shift in $H_2O$) and crystal structure analysis suggested that the decreased resolution observed for 5 or 6-fluorotryptophan, relative to 4-fluorotryptophan arose from their increased solvent exposure and consequent shielding.[352]

In the past decade, fluorine-containing proteins have found numerous applications including the study of high molecular weight proteins,[346,353] protein folding,[348,354–357] domain swapping,[358,359] and protein-ligand binding.[360,361] The $^{19}F$ nucleus has been particularly useful in characterising ligand binding by G protein–coupled receptors,[353] and overcoming the challenges of in-cell NMR.[184,346,357] For instance, Ye and co-workers recently demonstrated the advantages of $^{19}F$-labelling over $^{15}N$, $^{13}C$, and $^{2}H$ enrichment in *Xenopus laevis* oocytes for several test proteins.[362] Unlike the $^{15}N$, $^{13}C$, and $^{2}H$ enriched test proteins (GB1, ubiquitin calmodulin, protein disulfide isomerase and α-synuclein), all of the $^{19}F$-labelled test proteins were detectable in oocytes with spectra that were not compromised by background signals.[362] By probing the $^{19}F$ longitudinal relaxation time ($T_1$) of 5-fluorotryptophan or 3-fluorotyrosine labelled GB1 the authors determined that the viscosity of the oocyte cytoplasm was 1.2-1.3 times that of water.[362] The $^{19}F$ transverse relaxation rate ($R_2$) also reports on the viscosity imposed by weak interactions of the test protein with the surrounding medium. Thus, the difference in viscosities obtained from $T_1$ and $R_2$ measurements indicates the extent of test protein interactions with the cytoplasm. $R_2$ measurements suggest a viscosity twice that of water.[362] Considering the $R_2$-measured viscosity experienced by GB1 in *E. coli* was 6-11,[90,131] GB1 interactions with the oocyte cytoplasm are negligible. The authors performed $^{19}F$ line shape analysis to infer that the majority of GB1 resonance broadening arises due to chemical shift disparities imposed by the numerous microenvironments harbouring GB1 throughout the cytoplasm.[362] This $^{19}F$ NMR study represents the first quantitative NMR evidence of the effects of intracellular inhomogeneity on protein detectability in oocytes.

**Novel Routes for Producing $^{19}F$ Tryptophan-Labelled Proteins** Current methods to incorporate fluorine-containing amino acids are hampered by time-consuming genetic manipulation, compromised expression systems and require expensive amino acid analogues and inhibitors.[325,342,346,357,358,363] An alternative strategy is to use

fluorinated biosynthetic amino acid precursors. The use of labelled precursors (*e.g.* $^{13}$C pyruvate) is a well-established strategy for the production of selective $^{13}$C amino acid labelling.[227] Here, a simple method for the production of two proteins, GB1 and flavodoxin, containing $^{19}$F-tryptophan in the standard, prototropic expression strain *E. coli* BL21 (DE3) is described. The advantages of selecting Trp are its lower abundance (therefore giving rise to fewer signals relative to $^{19}$F-Tyr or $^{19}$F-Phe) and its frequent occurrence as a hot spot residue at protein interfaces.[364,365]

Precedent for this work comes from several lines of evidence: (1) the α-subunit of tryptophan synthase cleaves 3-indole-D-glycerol-3'-phosphate to release indole for subsequent attachment to an activated serine in the β-subunit. Crystal structure analysis reveals that the α-subunit accommodates a broad range of substrate mimics including fluorinated analogues.[366,367] (2) *E. coli* can incorporate isotopically-enriched ($^{13}$C) indole to yield protein that is suited to selective NMR detection of tryptophan side chains.[368] (3) *E. coli* M5219 (a tryptophan auxotroph) can use indole analogues (*e.g.* 7-azaindole, 4 and 5-fluoroindole) for tryptophan synthesis[369] and (4) halotryptophans can be produced by the addition of haloindole to *E. coli* cell lysates containing over-expressed tryptophan synthase.[370]

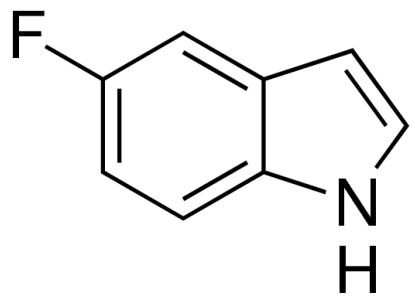



**Figure 2.** Structure of 5-fluoroindole

**Materials and Methods**

**Protein Production** The pET3a expression vector coding for GB1 (pGB1-QDD)[96] contained the T7/*lac* promoter and the ampicillin resistance gene. GB1 was produced by transforming the vector into *E. coli* BL21 (DE3). A single colony was used to inoculate a 5 mL LB preculture containing 75 μg/mL carbenicillin and 2 mM $MgSO_4$ and grown overnight at 30 ˚C. All cultures were supplemented with 75 μg/mL carbenicillin and 2 mM $MgSO_4$ and incubated with 300 rpm shaking. The preculture was transferred to a 1 L LB main culture and grown at 37 ˚C. When an optical density at 600 nm of 0.7 was reached the cells were harvested by centrifugation at 4500 rpm, 20 ˚C for 25 minutes. The cells were gently resuspended in 1 L minimal medium (50 mM $Na_2HPO_4$, 50 mM $KH_2PO_4$, 2 g/L $(NH_4)_2SO_4$, 2 g/L D-glucose, 20

mM sodium citrate pH 7.0, 20 mM sodium succinate pH 7.0, 1 x 5052,[250] 30 mg thiamine). After 30 minutes incubation at 37 ˚C the cells were split into 4 x 250 mL cultures (each in 2 L Erlenemeyer flask) which were either (A) unsupplemented or supplemented with (B) 120 mg/L 6-fluoro D/L-tryptophan, 60 mg/L L-tyrosine, 60 mg/L L-phenylalanine and 1 g/L glyphosate, (C) 60 mg/L 5-fluoroindole, and (D) 300 mg/L 5-fluoroindole. The 5-fluoroindole was first dissolved in 0.5 mL of dimethyl sulfoxide. GB1 over-expression was induced with 1 mM IPTG, the cells were harvested after 2 hours and resuspended in lysis buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.5) before freezing.

The pET3a expression vector coding for flavodoxin (pDH07)[121] also contained the T7/*lac* promoter and the ampicillin resistance gene. Flavodoxin was expressed as described for GB1 but with the following changes: the cells harvested from 400 mL of LB culture were resuspended in 500 mL minimal medium and split into 4 x 100 mL cultures (each in a 500 mL Erlenemeyer flask; the 5$^{th}$ 100 mL aliquot was discarded). After 30 minutes incubation the cultures were either (A) unsupplemented or supplemented with (B) 60 mg/L indole, (C) 60 mg/L 5-fluoroindole or (D) 30 mg/L 5-fluoroindole. Flavodoxin over-expression was induced with 1 mM IPTG and the cells were harvested after 4 hours before resuspending in the lysis buffer and freezing.

The pBPCYC1(wt)/3 expression vector codes for cytochrome *c*, cytochrome *c* haem lyase and the ampicillin resistance gene.[371] This plasmid was transformed into *E. coli* BL21 (DE3) and a single colony was used to inoculate a 5 mL LB preculture which was incubated for 3 hours at 37 ˚C. 1 mL of the preculture was used to inoculate each of the following 100 mL cultures: (A) TB (terrific broth), or minimal medium that was (B) unsupplemented or supplemented with (C) 60 mg/L indole (D) 60 mg/L 5-fluoroindole or (E) 20 mg/L 5-fluoroindole. Cytochrome *c* was constitutively expressed in each culture for 24 hours before harvesting, resuspending in lysis buffer and freezing.

**Protein Purification** Cell slurries containing GB1 were thawed and added to boiling, lysis buffer (30 mL buffer per cell suspension from 1 L culture), heated to 80 ˚C and then cooled on ice for 10 minutes before centrifuging at 11,000 rpm, 10 ˚C for 15 minutes. The supernatant was loaded onto a diethylaminoethyl (DEAE)

sepharose ion exchange column equilibrated in 20 mM Tris-HCl, 50 mM NaCl, pH 7.5. GB1 was eluted using a linear gradient from 0-1 M NaCl. GB1-containing fractions were concentrated and exchanged into a low ionic strength buffer before a second round of DEAE chromatography. The GB1-containing fractions were concentrated to < 800 μL before loading onto an XK 16/70 column packed with Superdex 75 equilibrated in 20 mM $KH_2PO_4$, 50 mM NaCl, pH 6.0. Protein purity was assessed by SDS-PAGE.

Cell slurries containing flavodoxin were thawed and sonicated before adding 10 μg/mL DNase 1, 100 mM $MgSO_4$ and centrifuging at 11,000 rpm 10 ˚C for 40 minutes. The supernatant was treated with ammonium sulfate at 30% saturation before centrifuging at 11,000 rpm 10 ˚C for 40 minutes to remove the precipitate. The supernatant was dialysed overnight at 4˚C against 2 L of 20 mM Tris-HCl pH 7.0. The cleared extract was loaded onto a DEAE column equilibrated in 20 mM Tris-HCl at pH 7.0 and flavodoxin was eluted using a linear gradient from 0-1 M NaCl. The flavodoxin-containing fractions were concentrated and exchanged into a low ionic strength buffer before a second round of DEAE chromatography in which flavodoxin was eluted using a linear gradient from 0-0.5 M NaCl. The flavodoxin sample was then purified by two rounds of SEC in 20 mM $KH_2PO_4$, 200 mM NaCl, pH 6.0. Flavodoxin-containing fractions (yellow) with an absorbance ratio $A_{274}/A_{467}$ of < 12 were pooled and exchanged into NMR buffer (20 mM $KH_2PO_4$, 75 mM NaCl, pH 6.0).

**Mass Spectrometry** The masses of unlabelled and fluorinated GB1 were determined using a Waters LCT Premier XE Time of Flight (TOF) mass spectrometer run in positive ion mode equipped with a Waters 2795 HPLC and MassLynx Software.

**NMR Spectroscopy** The typical NMR sample conditions comprised 0.3-0.5 mM $^{19}$F protein (GB1 or flavodoxin) in 20 mM $KH_2PO_4$, 50 or 75 mM NaCl, 10% $D_2O$ at pH 6.0. The in-cell and cell extract NMR samples were prepared as described previously (Chapter 2).[100] Samples of the starting materials, 6-fluoro D/L-tryptophan and 5-fluoroindole (~2 mM) were prepared in the same buffer with 10% $D_2O$. 1D $^{19}$F spectra were acquired at 25 ˚C on a Varian 500 MHz spectrometer equipped with a OneNMR probe operating at 470 MHz. GB1 spectra were obtained with 512 transients, 65k points, a spectral width of 230 ppm and a delay time of 1 second. The

measurement time was approximately 14 minutes. The cell lysate sample contained ~0.1 mM GB1 and the spectrum was obtained with 2048 transients. Spectra were referenced to a 10% trifluoroacetic acid sample. Flavodoxin spectra were acquired with 1024 transients, 5435 k points and a spectral width of 231.1 ppm. A relaxation delay time of 0.7 s and an acquisition time of 0.05 s were used. Baseline correction was applied where necessary. The $^{19}$F NMR data were analysed in MestReNova.[372] $^{1}$H, $^{15}$N HSQC spectra were acquired at 30 °C on a Varian 500 MHz spectrometer equipped with a HCN probe.

**Results and Discussion**

**Expression of $^{19}$F Trp-labelled GB1** GB1 contains a single Trp residue (Figure 3) and is a model for NMR studies.[74,76,96,99,358,362] $^{19}$F Trp-labelled GB1 was expressed in *E. coli* BL21 (DE3) cultured in minimal medium using two different labelling protocols. The fist protocol involved the addition of 1 g/L glyphosate to block the Shikimate pathway. Here, the medium was supplemented also with 120 mg/L 6-fluoro D/L-tryptophan, 60 mg/L L-tyrosine and 60 mg/L L-phenylalanine. The second labelling protocol involved the addition of 60 mg/L or 300 mg/L of the tryptophan precursor 5-fluoroindole.

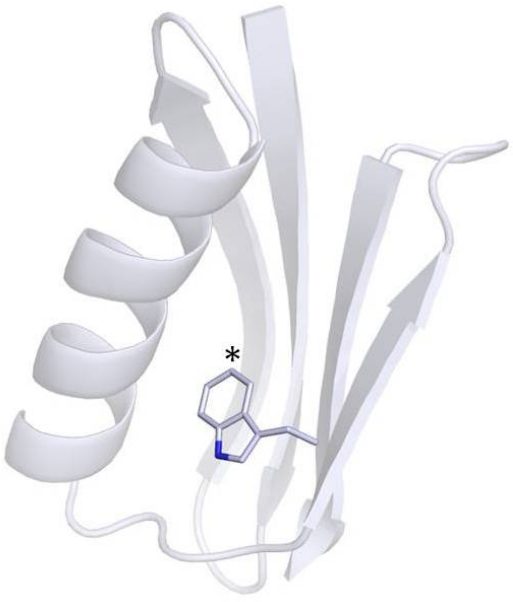



**Figure 3.** Crystal structure of GB1 with the Trp43 side chain shown as sticks. The fluorination site of 5-fluorotryptophan GB1 is indicated by an asterisk

Figure 4 shows the expression level of GB1 in minimal medium alone and in minimal medium supplemented as described. GB1 production was comparable in all cultures, except in minimal medium containing 300 mg/L 5-fluoroindole (Figure 4, lanes D). The low yield of GB1 was similar to that observed pre-induction (Figure 4, lane R) and was mirrored in a reduced total cell mass for culture D [~0.35 g (wet weight) compared to ~0.45 g for cultures A-C]. This finding suggests that high concentrations of 5-fluoroindole inhibit cell growth. Similarly, $^{19}$F-labelled aromatic amino acids can

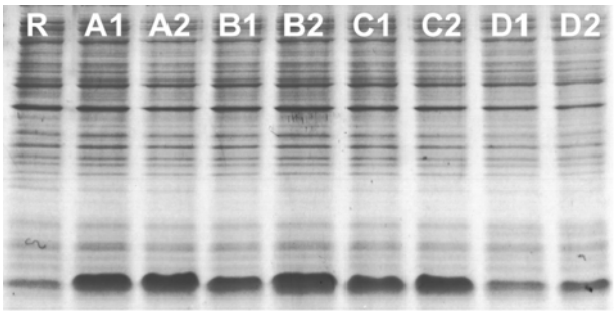



**Figure 4.** SDS-PAGE analysis of the total protein content of cell slurries prepared from minimal medium cultures (R) before and after induction. The minimal medium was (A) unsupplemented or supplemented with (B) 120 mg/L 6-fluoro D/L-tryptophan, 60 mg/L L-tyrosine, 60 mg/L L-phenylalanine and 1 g/L glyphosate, (C) 60 mg/L 5-fluoroindole, and (D) 300 mg/L 5-fluoroindole. Label numbers 1 and 2 indicate the number of hrs post induction.

inhibit *E. coli* growth.[325,350,363] Importantly, the yield of $^{19}$F-labeled protein from growth medium containing 60 mg/L 5-fluoroindole was comparable to that obtained from standard medium (Figure 4), suggesting that 5-fluoroindole was not toxic at this concentration. Pure protein samples were analysed by mass spectrometry and the observed masses (Table 1) indicated successful incorporation of 6-fluorotryptophan and 5-fluoroindole into B and C, respectively. $^{19}$F NMR spectra were acquired on the starting materials and the purified GB1 samples and were referenced to trifluoroacetic acid (Figure 5). The starting materials, 6-fluoro-tryptophan and 5-fluoroindole, gave rise to complex multiplets at -46.1 and -50.1 ppm, respectively (Figures 5A and 3B). GB1 produced from culture containing 6-fluorotryptophan gave rise to a signal at -45.1 ppm (Figure 5A) while GB1 produced from culture containing 5-fluoroindole gave rise to a signal at -48.2 ppm (Figure 5B), confirming the incorporation of the starting material. The chemical shifts were similar to previously reported values for proteins containing 6-fluorotryptophan[354] and 5-fluorotryptophan.[355,373,374]

**Table 1.** Observed and calculated masses for WT and $^{19}$F-GB1

| GB1 Purified from Culture | Calculated Mass (Da) | Observed Mass (Da) |
|---|---|---|
| A | 6224.8 | 6224.0 |
| B | 6242.8 | 6241.9 |
| C | 6242.8 | 6244.1 |

$^{19}$F NMR spectroscopy is useful for the detection of proteins inside cells.[339,362] The $^{19}$F in-cell NMR method has been developed using proteins labelled with 3-fluorotyrosine[357] and trifluoromethyl-phenylalanine.[346,357] These reagents are costly and the latter system is beset with the challenges of orthogonal codons. Thus,

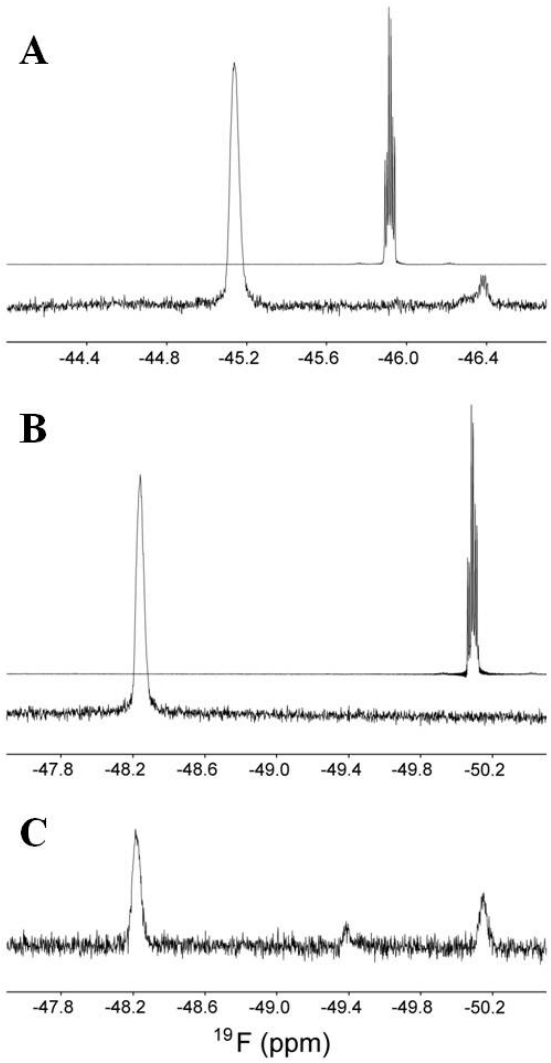



**Figure 5.** $^{19}$F NMR spectra of starting materials and fluorinated protein samples. (A) 6-fluoro D/L-tryptophan (upper trace) and pure GB1 produced on minimal medium containing 6-fluoro D/L-tryptophan (lower trace), (B) 5-fluoroindole (upper trace) and pure GB1 produced on minimal medium containing 5-fluoroindole (lower trace). (C) Cell lysate with over-expressed GB1 from 50 mL of *E. coli* culture grown on minimal medium containing 5-fluoroindole

in addition to the NMR spectroscopic characterisation of pure $^{19}$F-labelled GB1, cell extracts containing over-expressed $^{19}$F-GB1 were also studied. The cell extract gave rise to three signals in the 1D $^{19}$F spectrum (Figure 5C), which could be assigned to 5-fluorotryptophan-labelled GB1 and free 5-fluoroindole by comparison to the spectra for the pure compounds (Figure 3B). The peak at -49.4 ppm corresponds to free 5-fluorotryptophan.[362] The presence of a signal for $^{19}$F-GB1 in the cell extract demonstrates the utility of fluorotryptophan-labelled proteins for $^{19}$F-NMR detection of a specific protein in a complex mixture.

. A $^{15}$N/$^{19}$F-labelled GB1 sample was produced by expression on minimal medium containing both 5-fluoroindole and $(^{15}NH_4)_2SO_4$. The $^{1}$H,$^{15}$N HSQC spectrum (Figure 6) displayed minor chemical shift perturbations with respect to non-fluorinated GB1 confirming that 5-fluoro-tryptophan did not induce significant structural perturbations in the protein.[374]

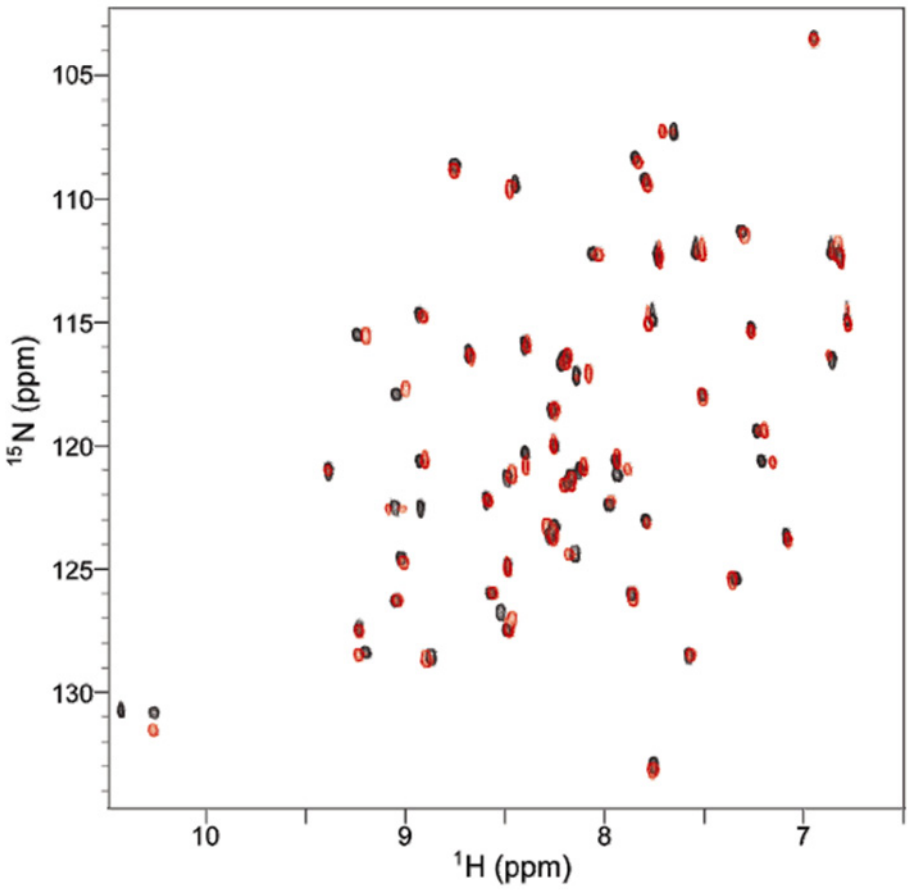



**Figure 6.** Overlaid $^{1}$H, $^{15}$N HSQC spectra of GB1 (black contours) and 5-fluorotryptophan GB1 (red contours).

**Expression of $^{19}$F Trp-labelled Flavodoxin** Flavodoxin contains 4 tryptophan residues (Figure 7) and has a higher tryptophan occurrence frequency (2.4%) than the average protein (1.4%).[375] To demonstrate the application of the $^{19}$F-Trp labelling strategy to proteins with more than 1 tryptophan, flavodoxin was used as a test case. Figure 8 shows the expression level of flavodoxin 0-4 hours post-induction in minimal medium alone and in minimal medium supplemented with 60 mg/L indole, 60 or 30 mg/L 5-fluoroindole. The total protein content was highest in the cell slurries prepared from normal minimal medium and decreased in the following order: 60 mg/L indole > 30 mg/L 5-fluoroindole > 60 mg/L 5-fluoroindole. Since reduced flavodoxin is blue-grey in colour, the difference in flavodoxin yield from the cultures grown on minimal medium supplemented with 30 and 60 g/L 5-fluoroindole was visible also in the cell pellets (Figure 9). The standard culture and the culture containing 60 mg/L indole had comparable total protein contents (Figure 8) indicating that indole is not toxic to *E. coli* at this concentration (0.5 mM). Indeed

indole is a well-established signalling molecule involved in plasmid stability, biofilm formation, stress responses, antibiotic resistance and virulence.[376] Previous studies of indole import and export in *E. coli* (using cultures supplemented with 2-10 mM indole) did not report any loss of cell viability.[376] This finding supports our observations. Although flavodoxin was expressed to low levels in *E. coli* cultured in minimal medium containing 60 mg/L 5-fluoroindole, the total protein content of the cell slurry remained low, even four hours post induction (Figure 8).

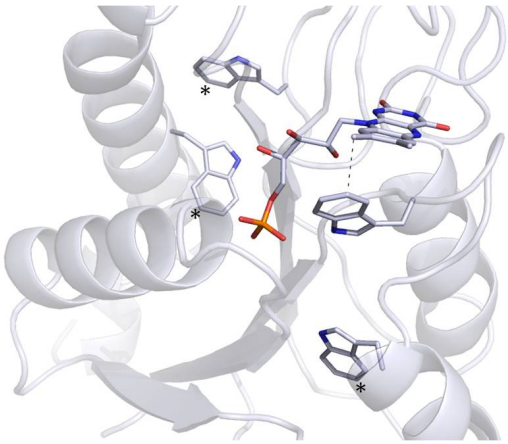


**Figure 7.** Tryptophan side chains of flavodoxin and the non-covalently bound flavin mononucleotide. Asterisks indicate the fluorination sites of W66, W120 and W160. The dashed line indicates that the fluorine of W57 is in van der Waals contact with the nucleotide.

A similar inhibitory effect was observed during the expression of the *Ralstonia solanacearum* fucose-binding lectin (RSL) in minimal medium containing 60 mg/L 5-fluoroindole (P.M. Antonik and P.B. Crowley, personal communication). The expression level of RSL was higher when the minimal medium was supplemented with 30 mg/L 5-fluoroindole compared to 60 mg/L 5-fluoroindole. Curiously, GB1 expression or cell growth was not significantly compromised in medium supplemented with 60 mg/L 5-fluoroindole. Indole, and probably fluoroindole, diffuses through the *E. coli* cell membrane without the need for endogenous transport machinery.[376] The high level of GB1 expression in the medium containing 60 mg/L 5-fluoroindole suggests that 5-fluoroindole is not toxic to *E. coli* cells at this external concentration. The disparity in cell toxicity effected in the presence of 60 mg/L 5-fluoroindole coincides with the number of tryptophans in the over-expressed test protein. GB1 has a single tryptophan while flavodoxin and RSL have 4 and 7 tryptophans, respectively. Owing to the known toxicity of fluorinated aromatic amino acids,[345,350,363] it is likely that the higher toxicity is effected by the increased cytosolic concentration of 5-fluorotryptophan produced from 5-fluoroindole in response to the over-expression of flavodoxin or RSL. Given that the culture contains 5-fluoroindole, the inhibition of

cell growth is unlikely to arise from hampered indole synthesis. Perhaps the suppressed *E. coli* growth in minimal medium containing 60-300 mg/L 5-fluoroindole occurred due to the suppression of another cellular process by 5-fluorotryptophan. The toxicity of 5-fluorotryptophan is expected to arise from its interference with the conversion of anthranilic acid to indole by anthranilate synthetase.[377]

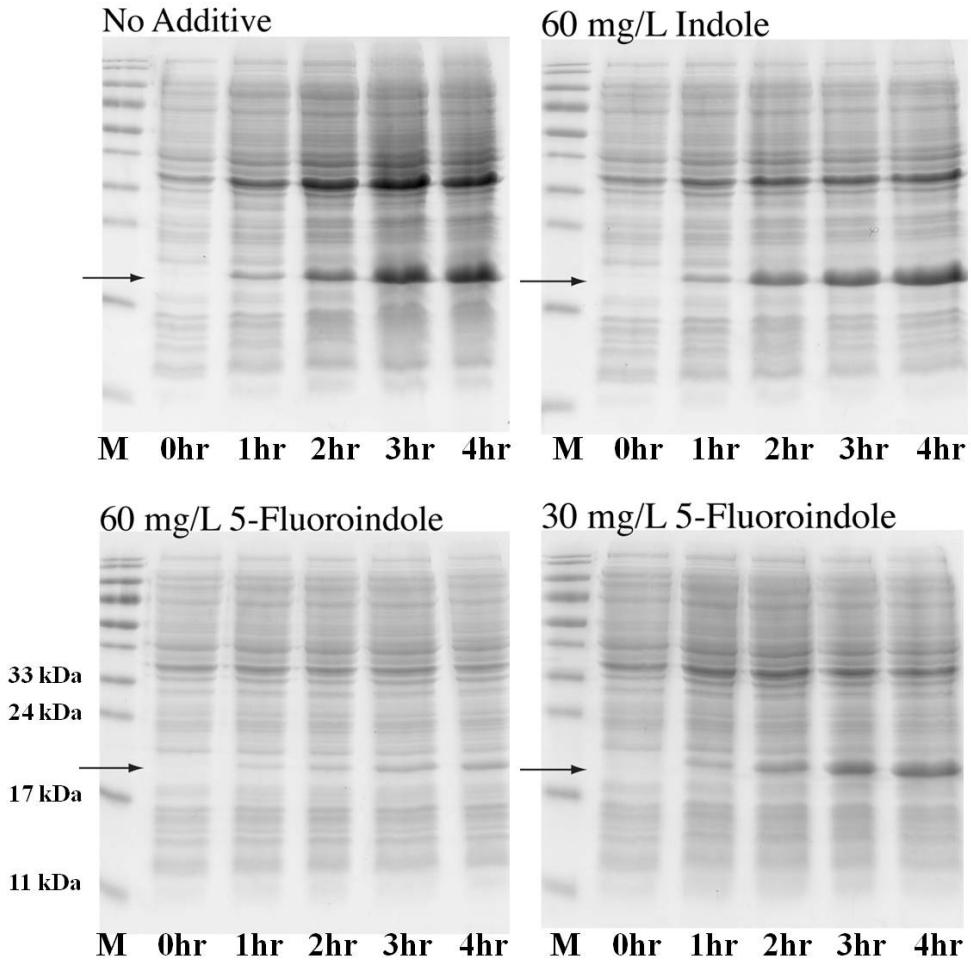



**Figure 8.** The total protein content of cell slurries prepared from minimal medium cultures 0-4 hours after the induction of flavodoxin overexpression. The minimal medium contained no additive or 60 mg/L indole, 60 mg/L 5-fluoroindole or 30 mg/L 5-fluoroindole. The gel lanes are labelled; M: molecular weight marker; 0hr-4hr: 0-4 hours post-induction. The arrow marks the migration position of flavodoxin.


**Figure 9.** *E. coli* pellets harvested from minimal medium supplemented with 60 or 30 mg/L 5-fluoroindole after 4 hours of flavodoxin over-expression. The pellet harvested from the culture containing 30 mg/L 5-fluoroindole is more intensely blue-grey due to the higher level of flavodoxin expression in this culture.

**$^{19}$F-labelled Cytochrome *c* Expression** Despite the fact that it contains a single tryptophan and expresses to high levels in *E. coli* BL21 (DE3), cytochrome *c* could not be expressed constitutively in *E. coli* BL21 (DE3) on minimal medium supplemented with 20 or 60 mg/L 5-fluoroindole. Figure 10 shows the total protein content of terrific broth (TB) and minimal medium cultures 24 hours after inoculation with 1 mL aliquots from the same preculture. Cytochrome *c* expression was evident in TB and in minimal medium that is unsupplemented or supplemented with 60 mg/L indole (Figure 10). The low total protein content of the minimal medium containing 20 and 60 mg/L 5-fluoroindole indicates that cell growth is inhibited by 5-fluoroindole in cultures with low cell densities. This finding was reflected in the optical density at 600 nm (Table 2). To test the effects of 5-fluoroindole on cell growth, the $O.D._{600}$ of low density minimal medium cultures containing untransformed *E. coli* BL21 (DE3) cells ± 5-fluoroindole was monitored over four hours. The minimal medium cultures were prepared by performing a 1 in 100 dilution of (A) a saturated LB culture ($O.D._{600} = 8$) or (B) a preculture grown for three hours. Similar dilution approaches are used in the preparation of minimal medium cultures from precultures for protein (A) over-expression and (B) constitutive expression, respectively (see Materials and Methods). Notably, the $O.D._{600}$ of the control cultures lacking 5-fluoroindole was comparable to that of the

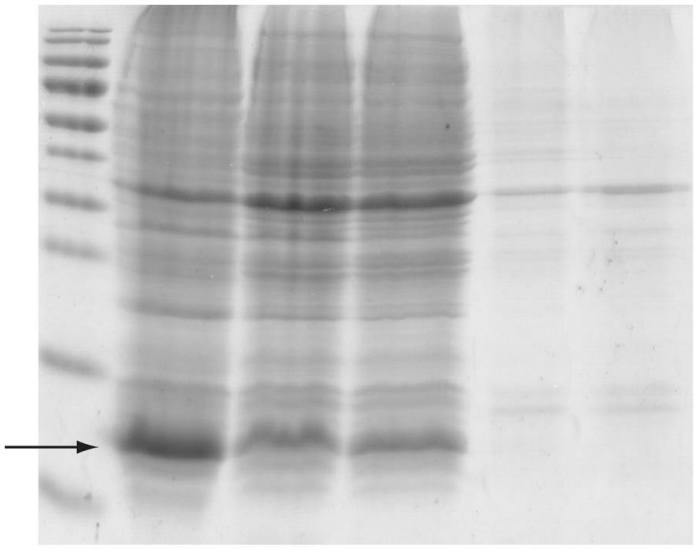



**Figure 10.** Total protein content of *E. coli* cell slurries constitutively expressing cytochrome *c* for 24 hours in terrific broth (TB) or minimal medium alone (MM) or minimal medium supplemented with 60 mg/L indole, or 60 or 20 mg/L 5-fluoroindole. M stands for molecular weight marker.

cultures containing 30 or 60 mg/L 5-fluoroindole (data not shown). This result suggests that the expression vector and the intracellular concentration of 5-fluorotryptophan likely affect the growth of *E. coli* cultures. It can be concluded that this labelling scheme is unsuited to *E. coli* cultures that constitutively express a protein of interest.

**Table 2.** O.D.$_{600}$ of *E. coli* cultures 24 hours post inoculation

| Culture | O.D.$_{600}$ |
|---|---|
| TB | 4.24 |
| Minimal medium | 3.69 |
| Minimal medium + 60 mg/L indole | 3.64 |
| Minimal medium 60 mg/L 5-fluoroindole | 0.29 |
| Minimal medium 20 mg/L 5-fluoroindole | 0.64 |

**In-cell and *in vitro* NMR of $^{19}$F-Trp labelled Flavodoxin** The incorporation of 5-fluoroindole into flavodoxin biosynthesis is evident from the $^{19}$F spectrum of the purified flavodoxin, which contains 4 well-resolved peaks at -120.46, -122.18, -124.30 and -125.24 ppm (Figure 11A). In-cell $^{19}$F NMR has been useful for the detection of 'sticky' proteins that are invisible by the HSQC experiment.[362,378] Flavodoxin was undetectable by $^{1}$H, $^{15}$N HSQC-TROSY NMR in cells and extracts

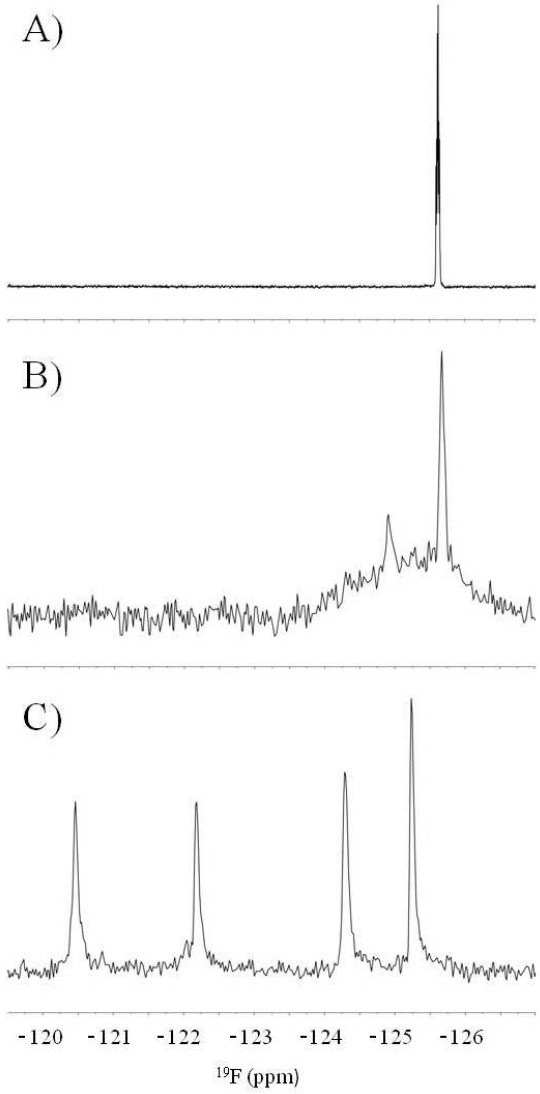



**Figure 11.** $^{19}F$ NMR spectra of fluorinated flavodoxin and starting materials: (A) pure 5-fluorotryptophan-labelled flavodoxin (B) in-cell NMR spectrum of $^{19}F$-labelled flavodoxin and (C) 5-fluoroindole

(Chapter 5). Accordingly, $^{19}F$-flavodoxin was studied by $^{19}F$ NMR inside *E. coli*. The $^{19}F$ spectrum obtained from a suspension of *E. coli* cells after 4 hours of flavodoxin over-expression is dramatically different to flavodoxin's *in vitro* spectrum (Figure 11B). The peak at -125.67 ppm corresponds to free 5-fluoroindole (Figure 11C). A second peak located 0.8 ppm downfield of the 5-fluoroindole peak corresponds to 5-fluorotryptophan (Figure 11B).[362] No other peaks were observed in the spectrum and thus, $^{19}F$ Trp labelled flavodoxin could not be detected by NMR inside *E. coli* cells despite high levels of over-expression (see blue-grey pellet; Figure 9).

**Conclusions**

Labelling of a target protein is imperative for NMR studies. Since the 1970s, $^{19}F$-labelled aromatic amino acids have been employed as NMR probes for investigations of biophysical properties. More recently, the application of $^{19}F$ NMR has tended toward protein recognition studies.[353,359,360,373,379,380] Herein, an inexpensive, facile method of $^{19}F$ Trp-labelled protein production is described which involves supplementing minimal medium with the fluorinated tryptophan precursor, 5-fluoroindole. In contrast to the standard method of fluorotryptophan incorporation, which relies on the use of tryptophan analogues and inhibitors,[325,374] our method does not require inhibitors of aromatic amino acid biosynthesis. The use of 5-fluoroindole instead of 6-fluoro-D/L-tryptophan represents a ~fifteenfold decrease in costs (based on gram prices from SIGMA). The cost advantage is substantially greater when the prices of the other

conventional additives, phenylalanine, tyrosine and glyphosate are factored in. The greater ease of production of fluorine-labelled proteins will lead to advancements in $^{19}$F protein NMR. In particular, the field of in-cell NMR spectroscopy[15] will benefit from the ability to produce $^{15}$N- and $^{19}$F-labelled protein samples by the simple addition of fluoroindole to the standard *E. coli* expression system. In combination with $^{19}$F NMR, 5-fluorotryptophan fluorescence could be exploited to yield a direct comparison of NMR and fluorescence-probed intracellular parameters (*e.g.* viscosity, pH), protein dynamics and interactions. The different timescales accessible to both techniques as well as their differential susceptibility to local intracellular environments could be harnessed to reveal novel aspects of quinary structure. The determination of NMR distance constraints for membrane protein structure elucidation may also benefit from more facile routes toward $^{19}$F-labelled protein production.[380] Finally, it is expected that this approach will facilitate the structural characterisation of fluorine-labelled proteins, leading to an improved understanding of non-covalent interactions involving fluorine. [379]

# Chapter 5
# Studies of Quinary-Like Interactions in E. *coli* Extracts

*Manuscript in preparation*

**Abstract**

Intracellular protein interactions structure the cytoplasm into functional zones. A detailed knowledge of protein interactions in physiological environments will therefore enhance our understanding of cytoplasmic structure. Although the interactions stabilising protein quaternary structure are well-established from *in vitro* studies, insights into the low-specificity protein-cytoplasm interactions that comprise 'quinary structure' are scant. This knowledge gap arises due to the significant difficulties associated with studying dynamic processes at the molecular-level *in vivo*. To understand the major physicochemical features governing quinary-like structures in physiological environments, the interactions of two sticky test proteins, cytochrome *c* and flavodoxin, were explored in *E. coli* extracts during a series of twofold serial dilutions. To this end NMR and native gel electrophoresis were employed. The inability to disrupt cytochrome *c*'s quinary-like interactions indicates that it forms relatively high affinity complexes with *E. coli* macroanions. Charge is not implicated in flavodoxin's attractive quinary interactions which were disrupted by a combination of cell lysis and stepwise dilutions. Together, the findings suggest that quinary interactions are governed by a complex interplay of non-covalent interactions that can, in the case of the anionic protein flavodoxin, be modulated by the level of macromolecular crowding.

## Introduction

Interactions beget complexity. Although often typified at the species level, this phenomenon also manifests *in cellulo* where myriad molecular interactions structure the cytoplasm into functional zones.[63,381] Intracellular interactions are not limited solely to quaternary structure formation.[83] Biomolecules tend to interact extensively *in vivo* to produce functional cytoplasmic meshworks that span distances far greater than the dimensions of oligomeric protein assemblies.[21,74,227,381,198] The term quinary structure[81] denotes the dynamic, weak interactions ($K_d$ > 1 μM)[73,382] that organise the cytoplasm. Considered less specific than the interactions that structure 'permanent' macromolecular machines,[198] quinary interactions likely govern inter-machine cross talk and cytoplasmic responsivity. Owing to their inherent transience, quinary interactions are readily perturbed and must therefore be captured *in vivo*.

Although NMR spectroscopy is well-suited to the residue-level characterisation of transient interactions in various milieus,[80,383] quinary interactions have confounded in-cell HSQC studies of globular proteins. The increased apparent molecular weights of proteins inside cells results in decreased rotational correlation times and reduced spectral quality. This shortcoming has been used to suggest the ubiquity of quinary interactions *in vivo*.[73,74,100,231,362] In-cell NMR studies of biologically inert ('unsticky') proteins and their mutants have been used to explore quinary structure. For instance, Monteith and co-workers used NMR-monitored hydrogen/deuterium exchange to study the effects of quinary interactions on the stability of GB1 and four of its point mutants in *E. coli* cells.[198] Measurements of mutant GB1 stability in buffer revealed that the average destabilisation effected by the mutations was low.[198] However, in the *E. coli* cytoplasm the mutations destabilised GB1 tenfold greater than in buffer due to changes in the quinary interactions of the mutated GB1.[198] This result indicates the role of quinary interactions in altering biophysical protein properties. Importantly, the most dramatic destabilisation was effected by the charge-inverted mutant GB1 D40K in *E. coli*,[198] suggesting the role of surface charge-charge contacts in modulating a protein's quinary interactions and, thus, stability *in vivo*. Subsequently, Cohen and co-workers showed that the quality of the GB1 K10H HSQC spectrum deteriorates with decreasing cytoplasmic pH, indicating that surface charge neutralisation extends quinary structure.[87] Barbieri *et al.* used mutagenesis and in-cell NMR to show that

the quinary-like interactions of the cationic human protein profilin 1 differed in human and *E. coli* cells.[384] Profilin 1 interactions were disrupted in *E. coli* cells by mutating 4 or 5 cationic residues to uncharged or anionic residues.[384] *E. coli* extract NMR studies indicated that, in the absence of physiological partners, profilin 1 interacted preferentially with *E. coli* RNA via arginine/lysine-phosphate salt bridge interactions.[384] Although studies of quinary structure are in their infancy, the role of attractive charge-charge interactions in driving quinary-like interactions in the cytoplasm has been established.[79,198,384] However, given the abundance of anionic proteins in many proteomes,[64,385] knowledge of non-covalent forces governing anionic protein stickiness is also required. Thus, investigations of the low-specificity interactions of both cationic and anionic test proteins under native-like conditions remain crucial for identifying the key physicochemical forces driving 'quinary-like' interactions.

Herein, *E. coli* extracts are used to mimic the complex, heterogeneous cytoplasm for investigations of the quinary-like interactions of two oppositely charged proteins, cytochrome *c* and flavodoxin. Neither protein is NMR detectable in living *E. coli* cells (Chapters 3, 4 and reference 74) indicating their sticky nature. NMR and native gel electrophoresis was used to explore the physicochemical basis of cytochrome *c* and flavodoxin interactions in extracts during two-stage, twofold serial dilutions. The results indicate differences in the strength of the attractive, quinary-like interactions involving either the cationic cytochrome *c* or the anionic flavodoxin. These findings are used to consider how different levels of macromolecular crowding might differentially tune the quinary interactions of cationic and anionic proteins.

## Materials and Methods

**Protein Expression and Purification** $^{15}$N-labelled and unlabelled *S. cerevisae* cytochrome *c*[371] and *E. coli* flavodoxin[121] were expressed in *E. coli* BL21 (DE3) and purified according to standard methods (Chapters 3 and 4).[74,371,119] Purified cytochrome *c* was oxidised with a 2-3 fold excess of $K_3[Fe(CN)_6]$ at 4 ˚C. Cytochrome *c* and flavodoxin samples were exchanged into NMR buffer (20 mM $KH_2PO_4$, 75 mM NaCl, pH 6.0) before freezing. Protein concentrations and purity were estimated from UV-vis spectra using the extinction coefficients $\varepsilon_{550}$= 27.5 mM$^{-}$

$^{1}cm^{-1}$ and $\varepsilon_{467}$ = 8.5 $mM^{-1}cm^{-1}$ for cytochrome *c*[74] and flavodoxin,[121] respectively. Protein purity was assessed also using SDS-PAGE.

**Cell Extract Preparation** Cell extract samples were prepared as previously described[100] from 50 mL LB cultures saturated with *E. coli* BL21 DE3 ($O.D._{600}$ = 7; Chapter 3).[291] The extracts were always pre-treated with DNase I and were stable for up to 8 hours when stored on ice. Directly prior to NMR or native gel electrophoresis the concentrated extracts were spiked with 0.3 or 0.2 mM of the $^{15}N$-labelled or unlabelled test protein, respectively. Cell extract dilutions were performed using 20 mM $Na_2HPO_4$, pH 7.0 and, for the NMR experiments, the test protein and $D_2O$ concentration was always restored to 0.3 mM and 10 %, respectively, before the next acquisition.

**NMR Spectroscopy** $^{1}H$,$^{15}N$ TROSY-selected[292] HSQC spectra were acquired at 30 ˚C with 16 scans and 64 increments on a Varian 600 MHz Spectrometer equipped with a HCN coldprobe. The data were processed in Biopack (using linear prediction in the $^{15}N$ dimension) and NMRPipe.[386]

**Native Gel Electrophoresis** 20 μL samples of 0.2 mM cytochrome *c* and/or flavodoxin were analysed in dilute solution and in *E. coli* extracts by native gel electrophoresis in 2% agarose gels (13.5 cm x 14.0 cm). The gels were prepared in 20 mM $KH_2PO_4$, pH 6.0 or 20 mM $Na_2HPO_4$, pH 7.0. 100 mM $KNO_3$ has a relatively high conductivity and is therefore a suitable alternative to NaCl for ionic strength studies (Chapter 3). 20 mM phosphate buffers containing 100 mM $KNO_3$ were therefore used to produce high ionic strength buffers at pH 6.0 or 7.0. Gels were equilibrated in the selected buffer for 30 minutes at 4 ˚C prior to running at a constant voltage (100 V) for 30 minutes at 4 ˚C. The gels were imaged using a flatbed scanner directly after electrophoresis. Each gel image was processed identically using Adobe Photoshop. Importantly, changes to the colour balance, saturation and contrast were applied uniformly across the images (Chapter 3).

**Results and Discussion**

**In-Cell NMR Recapitulates Cytochrome *c* and Flavodoxin Promiscuity.** In-cell NMR has emerged as an effective tool to gauge the extent of test protein quinary interactions.[74,79,87,100,184,237,238,387–390] Figure 1 compares the $^{1}H$, $^{15}N$ HSQC-TROSY

spectra for cytochrome *c* or flavodoxin in buffer or in live *E. coli* cells (the cytochrome *c* data were published previously[74]). Although both proteins give rise to high quality HSQC spectra in buffer they are undetectable *in cellulo.* Similarly, cytochrome *c* was not detectable in *E. coli* extracts, despite the use of $^{1}H$,$^{15}N$ TROSY-HSQC experiments (Figure 2).[292,391,109,392] Flavodoxin was weakly detectable in the TROSY-HSQC spectrum obtained from the concentrated extract (Figure 2). Together, these data indicate that cytochrome *c*[74] and flavodoxin interact extensively with cytosolic macromolecules.

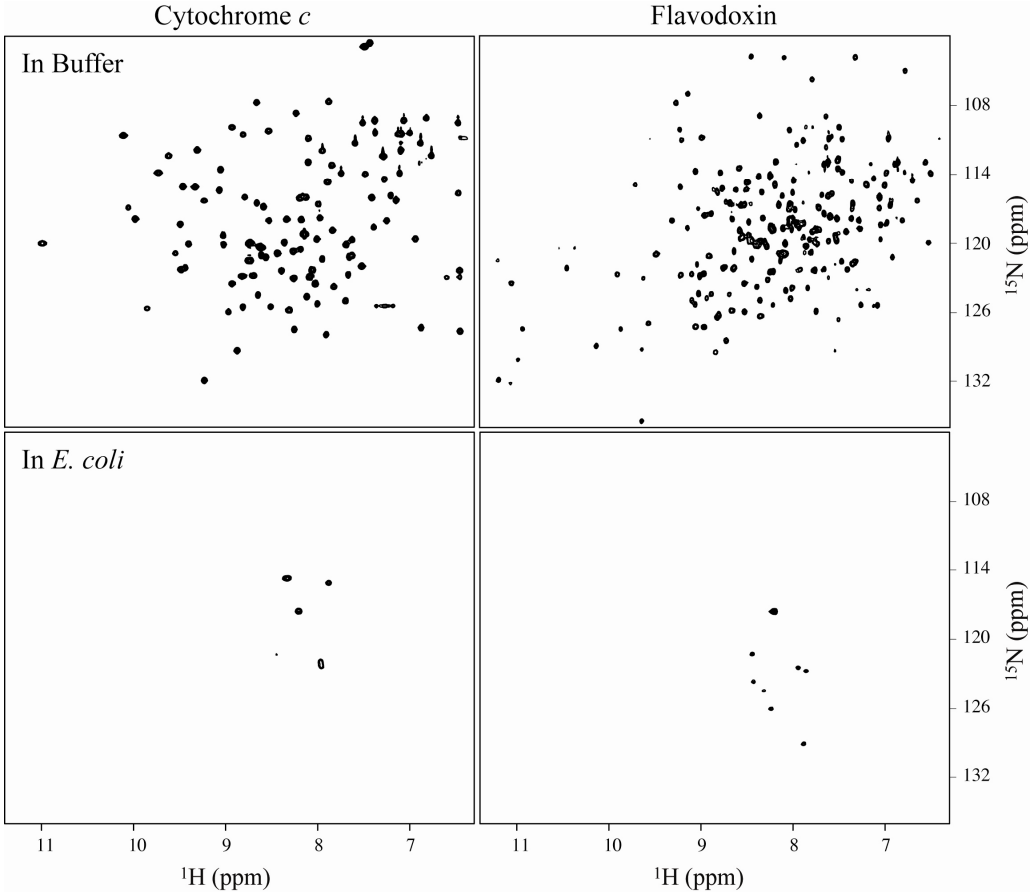



**Figure 1.** $^{1}H$, $^{15}N$ TROSY-HSQC spectra of cytochrome *c* or flavodoxin in buffer or in live *E. coli* cells. The cytochrome *c* spectra were published previously.[74] The spectra are contoured identically.

The NMR detectability of cytochrome *c* and flavodoxin was also assessed in dilute extracts (Figure 2). Two and fourfold diluted extracts were prepared by serial dilution, using 20 mM $Na_2HPO_4$ pH 7.0 (see Materials and Methods). With each dilution, the concentration of $D_2O$ and the $^{15}N$-labelled test protein was replenished to 10% and 0.3 mM, respectively. This approach facilitated a direct comparison of

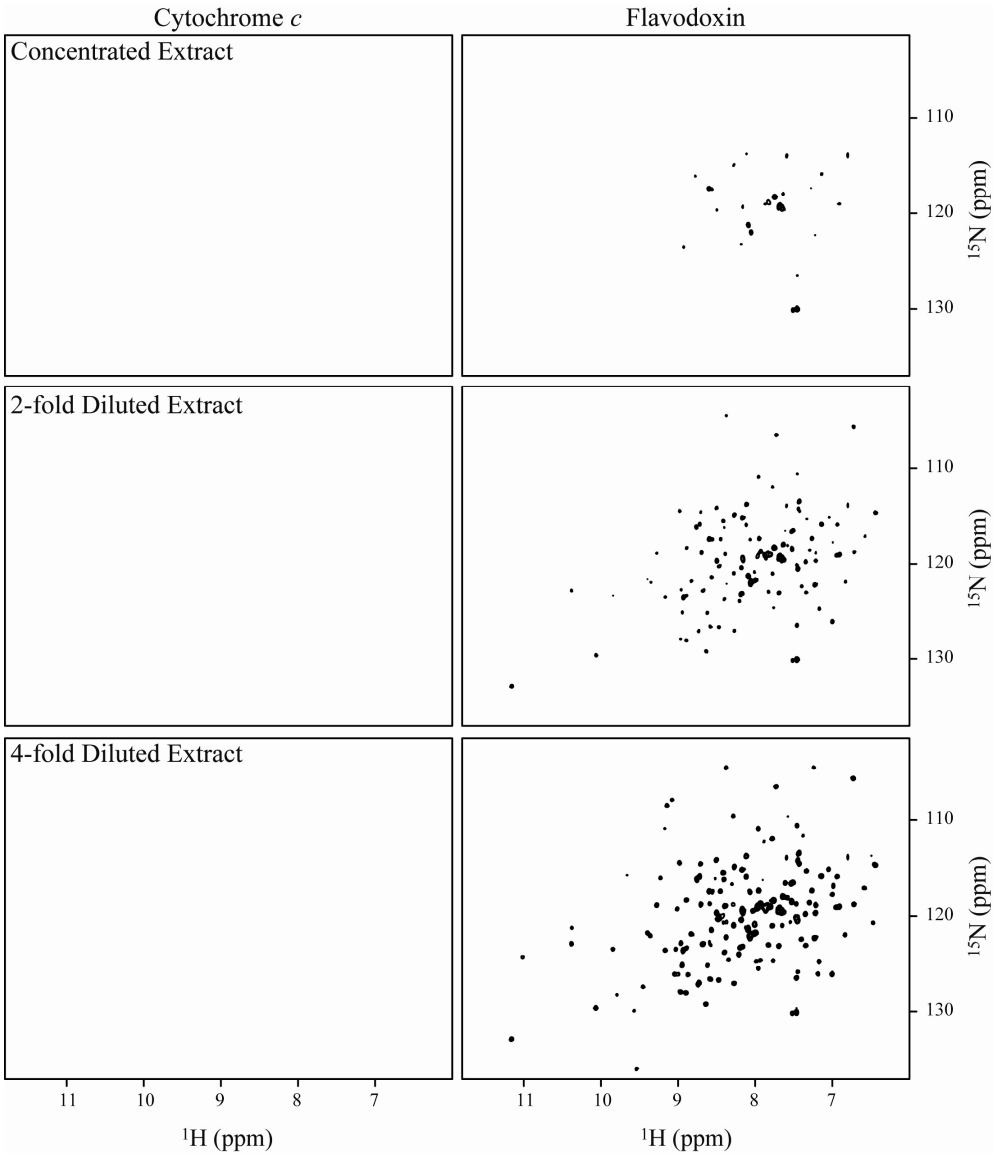



**Figure 2.** $^{1}H$, $^{15}N$ TROSY-HSQC spectra obtained from *E. coli* extracts containing 0.3 mM cytochrome *c* (left hand side) or flavodoxin (right hand side) before and during a two-stage, twofold serial dilution. The spectra are contoured identically.

the spectra. Cytochrome *c* was never detectable in the extracts suggesting that it interacts pervasively and perhaps tightly to *E. coli* macromolecules (Figure 2). This idea is supported by the fact that cytochrome *c* interactions with extract components are not disrupted upon cell lysis and passage through an SEC column (Chapter 3 and reference 74). The signal-to-noise ratio of the flavodoxin peaks increased with increasing dilution (Figure 2) suggesting that flavodoxin's interactions with *E. coli* macromolecules are weak and easily perturbed.[73,78,127] SDS-PAGE analysis of the extract samples taken directly after extract preparation and each NMR experiment indicate that no major changes in sample composition (*e.g.* aggregation) occur over time (Figure 3).

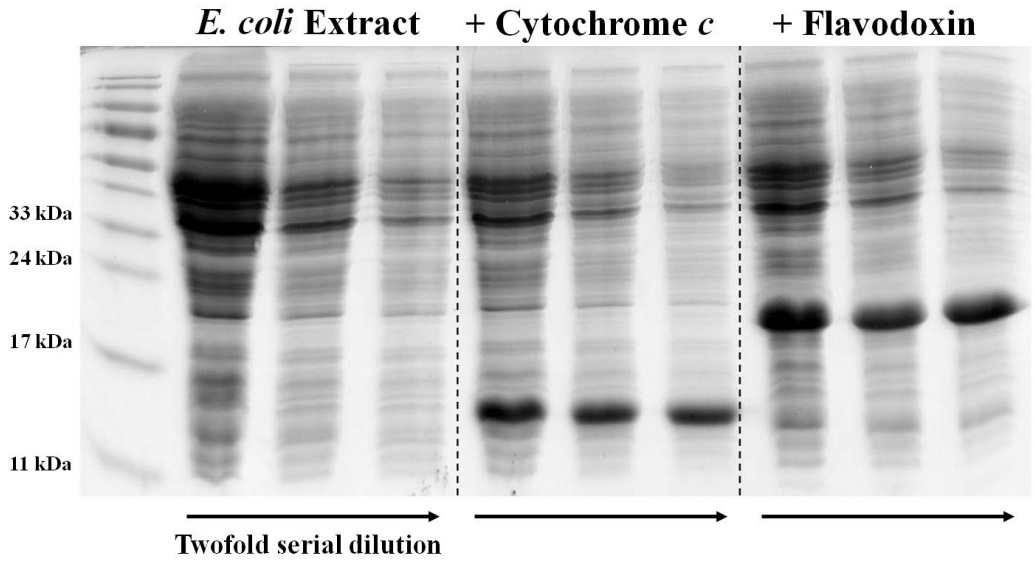



**Figure 3.** The total protein content of *E. coli* extracts used for NMR. Samples were loaded from left to right in order of increasing cell extract dilution (as per arrow). The left hand panel shows the composition of the freshly prepared extracts during a two-stage, twofold serial dilution (using 20 mM $Na_2HPO_4$, pH 7.0) without any added $^{15}N$-labelled protein. The middle and right hand panels show the extracts containing cytochrome *c* or flavodoxin during a two-stage, twofold serial dilution, taken directly after each NMR experiment.

**The Role of Charge in Protein Stickiness** Native gel electrophoresis was used to explore cytochrome *c* and flavodoxin charge in buffer and in *E. coli* extracts. Consistent with their calculated net charge pure flavodoxin (yellow) migrated toward the anode while pure cytochrome *c* (red) migrated toward the cathode in 2% agarose gels run in phosphate buffers ± 100 mM $KNO_3$ at pH 6.0 or 7.0 (Figure 4). The migration distance of cytochrome *c* was notably short compared to flavodoxin (Table 1). Indeed, membrane confined electrophoresis[129,271] indicated that the effective charge of cytochrome *c* (~+1) is up to 8 times lower than the calculated charge (+8.2) and counterion-dependent (Chapter 3). Cytochrome *c* has an effective charge of +1.04 in a buffer containing 100 mM $KNO_3$ which explains the poor migration of cytochrome *c* in agarose gels run under similar conditions. Given the physiological-relevance of the buffer ionic strengths,[61,393] it is likely that cytochrome *c* charge is comparably low *in vivo*. Flavodoxin carries more charge than cytochrome *c* and therefore migrates further (Figure 4; Table 1). The migration distance of cytochrome *c* and flavodoxin was typically twice as long in agarose gels run at low ionic strengths (*e.g.* 20 mM phosphate, pH 6.0 and 7.0) than at high ionic strengths (*e.g.* 20 mM phosphate + 100 mM $KNO_3$ Figure 4; Table 1). Moreover, in the case of cytochrome *c*, the migration distance was also approximately twice as long in the low ionic strength buffer at pH 6.0 compared to pH 7.0 (Figure 4; Table 1).

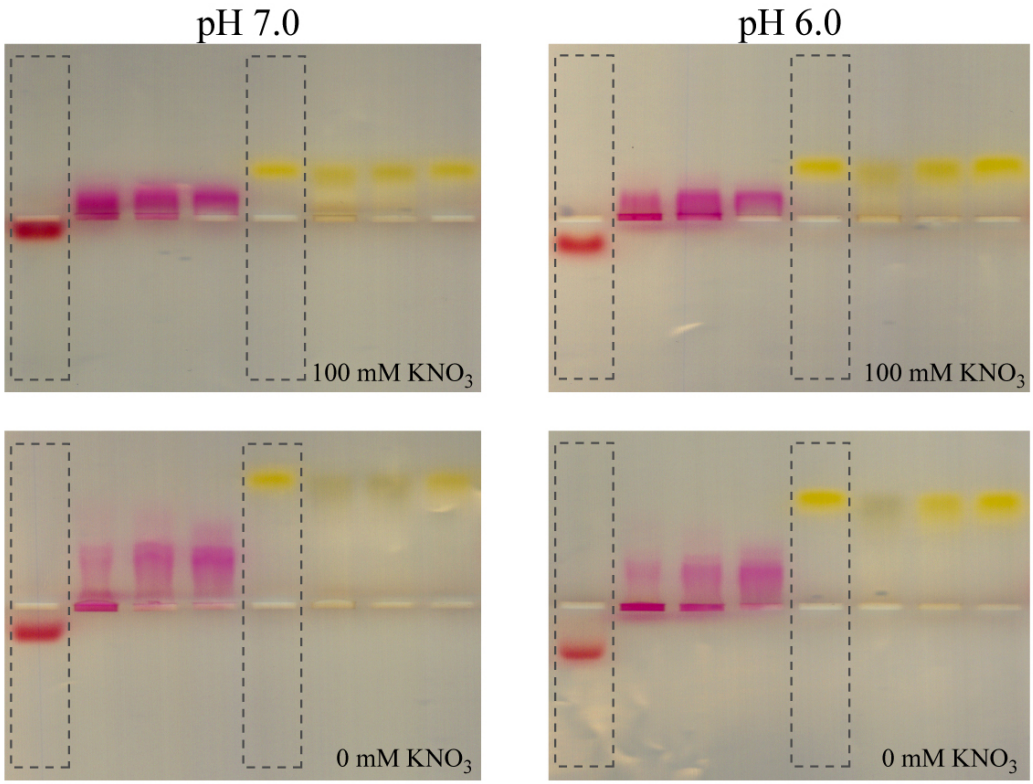



**Figure 4.** Agarose gel electrophoresis of cytochrome *c* (red/pink) and flavodoxin (yellow/green) in buffer or in *E. coli* extracts. Each sample (20 μL) contained 0.2 mM cytochrome *c* or flavodoxin and 10% glycerol. The grey boxes highlight the pure cytochrome *c* and flavodoxin samples. The remaining lanes contain extract samples which were loaded from left to right in order of decreasing total macromolecule concentration. Cytochrome *c* changes from red to pink in extracts due to reduction of the haem iron. Similarly, flavodoxin's flavin mononucleotide group is reduced to the semiquinone species. The electrophoresis buffers were 20 mM $KH_2PO_4$ or $Na_2HPO_4$ pH 6.0 or 7.0, respectively.

Accordingly, cytochrome *c* is more cationic at lower pH values. Similarly, pure flavodoxin migrates further in the low ionic strength buffers at pH 7.0 when it is more anionic (Figure 4; Table 1). These findings emphasise the role of ions in tuning cytochrome *c* and flavodoxin charge and this is supported by the MCE data for cytochrome *c* (Chapter 3). Agarose gels run at identical pH values show similar trends in cytochrome *c* and flavodoxin migration at both low and high ionic strengths. However, it is useful to consider the low ionic strength data when assessing the test protein interactions in cell extracts owing to the greater migration distances of the test proteins, which can be more readily interpreted/compared.

**Table 1.** Migration distance of pure cytochrome *c* and flavodoxin as a function of pH and ionic strength

| | Migration Distance at pH 6.0 /mm | | Migration Distance at pH 7.0 /mm | |
|---|---|---|---|---|
| **[$KNO_3$]/ mM** | 0 | 100 | 0 | 100 |
| **Cytochrome *c*** | 7.5 | 4.0 | 4.0 | 2.0 |
| **Flavodoxin** | 17.0 | 8.0 | 20.0 | 8.0 |

In *E. coli* extracts, cytochrome *c* migrated toward the anode at pH 6.0 and 7.0 (Figure 4). This indicates that quinary-like interactions between cytochrome *c* and *E. coli* macromolecules result in assemblies with a net negative charge. In the concentrated extracts a substantial proportion of cytochrome *c* was observed in the loading well suggesting that it participated in assemblies that are either charge-neutral or too large to migrate through the agarose gel pores (pore radius ~360 nm[394]). Despite containing identical cytochrome *c* concentrations (0.2 mM), the 2-fold serially diluted extracts had a decreased amount of cytochrome *c* in the loading well (Figure 4), indicating that dilution partially disrupted the large complexes. However, cytochrome *c* consistently migrated toward the anode, even in extracts diluted fourfold. This result, in agreement with the cell extract NMR studies (Figure 2), suggests that quinary-like cytochrome *c*-macromolecule interactions pervade under native-like conditions despite different degrees of macromolecular crowding and total protein concentration. Moreover, the resulting complexes were always anionic. In the 2fold diluted extracts, more cytochrome *c* was evident in the loading well at pH 6.0 than at pH 7.0. Thus cytochrome *c* appears to engage in high molecular weight complexes at pH 6 that are not completely disrupted by a 2fold dilution. At pH 7.0, the complexes are weaker. This finding coincides with recent work by Cohen *et al.* showing that quinary interactions were more extensive at lower pH values.[14] Flavodoxin migration was essentially unaltered by cell extracts. This suggests that attractive charge-charge interactions do not drive flavodoxin interactions in extracts as expected for such a highly anionic protein in a largely anionic environment.

**The Nature of Cytochrome *c* & Flavodoxin's Quinary-Like Interactions**

Cytochrome *c* and flavodoxin are oppositely charged proteins (Figure 5) that engage in pervasive interactions with cytoplasmic macromolecules in *E. coli* cells and extracts. Cytochrome *c* is lysine rich and was never NMR detectable in cell extracts during a two-stage, twofold serial dilution (Figure 2). Previously, the role of charge

in driving cytochrome *c* stickiness in extracts was demonstrated by SEC and NMR.[74] The greater the number of charge inverted mutations introduced to the cytochrome *c* surface, the less sticky the protein.[74] A triple cytochrome *c* mutant (R13E/K73E/K87E) was NMR detectable in concentrated extracts.[74] Native gel electrophoresis confirmed that the triple and quadruple mutants are charge-neutral and hydrated in extracts (Chapter 3). Together these data indicate the role of several lysine residues in governing cytochrome *c* interactions with anionic *E. coli* proteins. However, unlike arginine, lysine is not typically observed at specific protein-protein or crystal packing interfaces indicating that it is an unsticky residue.[83,365] The residues that are most frequently observed at protein-protein interfaces include tryptophan, tyrosine, histidine, phenylalanine, arginine, cysteine, methionine,

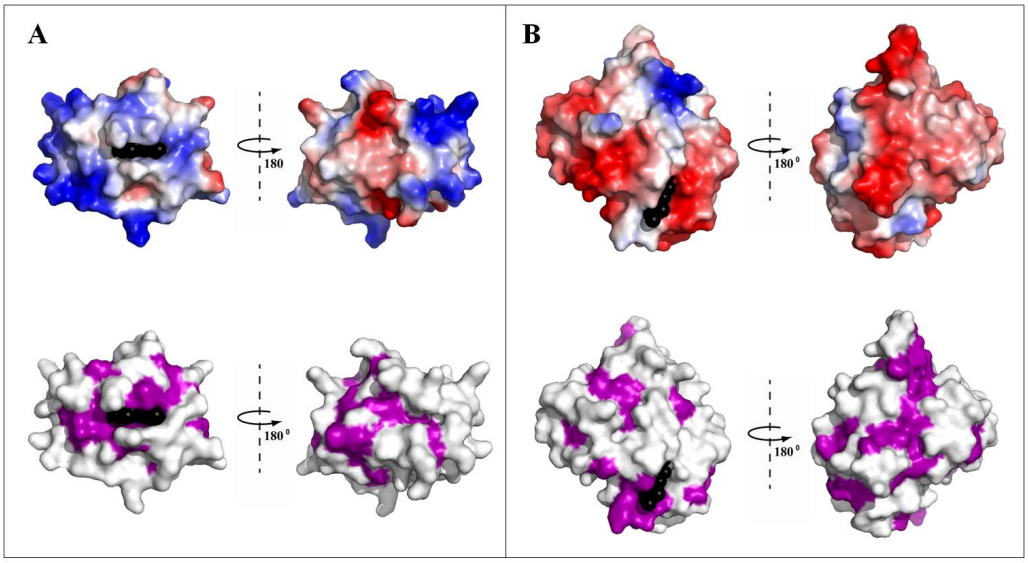



**Figure 5.** Electrostatic (top) and 'sticky' surface (bottom) representations of cytochrome *c* (A) and flavodoxin (B). In the sticky surface representations, residues are coloured according to the interface propensity scale devised by Janin and co-workers.[83] The residues most frequently observed at protein-protein interfaces are coloured magenta. The figure was generated in Pymol. The co-factors are shown as black spheres.

isoleucine, leucine and valine. Arginine, glutamine and the abovementioned hydrophobic residues most frequently comprise crystal packing interfaces.[83,365] Thus, surface representations of cytochrome *c* and flavodoxin prepared according to Janin's amino acid interface propensity scale[83] suggest that cytochrome is less sticky than flavodoxin (Figure 5). The cell extract studies presented herein and by Crowley and co-workers[74] indicate that cytochrome *c* is highly sticky. This finding emphasises the fact that protein stickiness is a system property that must be characterised in native-like environments. Under volume-occupied, physiological

conditions, lysine-rich surface clusters, although rare,[51,395] are apparently important mediators of quinary-like interactions. The quinary-like interactions of cytochrome *c* are relatively strong because the protein was never liberated from macromolecular assemblies in cell extracts during dilution experiments. The pervasive nature of cytochrome *c* interactions coincides with the expected role of small cationic proteins in maintaining the structure of protein-protein interaction networks and clusters.[396]

Like cytochrome *c*, flavodoxin engages in extensive attractive interactions with *E. coli* macromolecules and is weakly NMR detectable in concentrated extracts. This finding is counterintuitive, as flavodoxin's highly anionic surface might suggest that its quinary-like interactions would be dominated by charge-charge repulsion with the abundant macroanions. However, agarose gel analysis indicated that charge is not implicated in flavodoxin's attractive quinary interactions (Figure 4). This finding is also supported by numerous in-cell NMR studies of anionic and NMR undetectable test proteins.[73,184,397,398] Importantly, the flavodoxin surface contains clusters of 'sticky' residues that are frequently observed in protein-protein and crystal packing interfaces (Figure 5). These clusters likely mediate flavodoxin's interactions with macromolecules in the crowded cell extracts. Most of these clusters contain hydrophobic residues, some of which and are known to mediate flavodoxin's interactions with its anionic cofactor flavin mononucleotide (*e.g.* W57 and Y94).[125] The expected role of hydrophobicity and other short-range non-covalent interactions[307] in governing flavodoxin stickiness is supported by its increasing NMR detectability during serial dilution experiments. Serial dilutions decrease the total macromolecular concentrations of the extracts which leads to decreased levels of macromolecular crowding and increased intermolecular distances. Accordingly, short-range, attractive interactions are abolished upon diluting the extracts while long-range charge-charge repulsions between flavodoxin and other macroanions predominate. The liberated flavodoxin molecules then have sufficient rotational freedom for its spectrum to be observed in the dilute extracts.[78,127,384]

The modulation of flavodoxin interactions by different levels of macromolecular crowding suggests an important role for promiscuous anionic protein surfaces in mediating weak interactions in the responsive, abundantly anionic cytoplasm. Recently, Xu and co-workers explored the spatiotemporal dynamics of proteins and their interaction networks in a virtual cytoplasm to understand the

physicochemical basis for functional efficiency *in cellulo*.[396] The virtual cytoplasm contained over 200 essential protein types sourced from various organisms. This approach was expected to mimic the fundamental features of many proteomes.[396] Upon varying protein copy numbers relative to physiologically-accurate values determined for yeast, the authors used coarse-grained modelling to approximate the electrostatic and van der Waals interactions between the proteins (and ribosomes).[396] These simulations revealed that the highest number of interacting proteins was observed only when physiologically-accurate protein copy numbers were employed.[396] Moreover, the distributions of protein charge and molecular weight in the theoretical proteome were not random. Larger proteins were typically anionic and presented large collision cross-sections which likely mediate low-specificity quinary interactions.[396] Smaller proteins were usually cationic and more frequently collided with other macromolecules.[396] The flavodoxin data support the notion that anionic protein surfaces engage in weak macromolecular interactions that are readily perturbed (*i.e.* by a local decrease in macromolecular crowding levels) in cells replete with anionic macromolecules. The pervasive nature of cytochrome *c* interactions in extracts emphasises the role of small cationic proteins in maintaining the architecture of interaction networks. The lower abundance of cationic proteins in many proteomes,[51,64] is expected to reflect the need for highly dynamic clusters which is frequently afforded by anionic macromolecules.

**Conclusions**

Knowledge of quinary structure will advance our understanding of cellular phenomena. Insights into quinary structure are hampered, however, by a paucity of robust experimental techniques and computational models suited to studies of complex living systems.[90] A semi-reductionist approach was therefore employed for the exploration of the physicochemical origin of cytochrome *c* and flavodoxin's extensive quinary-like interactions with *E. coli* extracts. Different non-covalent interactions appear to drive the attractive quinary-like interactions of both test proteins in *E. coli* extracts. This finding corresponds with the different surface properties of cytochrome *c* and flavodoxin (Figure 5). The data support the expected role of cationic proteins in structuring cytoplasmic assemblies through pervasive, attractive charge-charge interactions. The role of anionic proteins in mediating co-operative short-range macromolecular interactions in the cytoplasm is also

suggested. The abundance of anionic macromolecules *in vivo* emphasises the importance of dynamic macromolecular interactions, which are easily perturbed *i.e.* by different levels of macromolecular crowding, and therefore essential for cytoplasmic responsivity.

# Chapter 6

# Effects of Crowding and Confinement on a Quinary-Like Protein-Protein Interaction

*Manuscript in preparation*

**Abstract**

*In vivo* studies are necessary to simultaneously capture the myriad, weak biomolecular interactions that govern quinary structure. However, investigations of weak interactions in physiologically-relevant solutions may reveal novel physicochemical aspects of quinary or cytoplasmic structure. It is well-appreciated from *in vitro* studies that electron transfer (ET) proteins interact weakly with several cognate partners, making them excellent models for studies of 'quinary-like' protein interactions. Herein, the quinary-like interaction of two oppositely-charged ET proteins, cytochrome *c* and flavodoxin was studied in environments with varying degrees of cytoplasmic mimicry. A combination of $^{1}H$, $^{15}N$ HSQC and $^{19}F$ NMR spectroscopy confirmed that cytochrome *c* and flavodoxin interact weakly in buffered solutions to produce a non-cognate complex. This quinary-like interaction was also investigated by NMR and electrophoresis in volume-occupied solutions. The data suggests that quinary-like interactions involving oppositely-charged partners are stabilised by moderate levels of macromolecular crowding and gel confinement. The biological implications of our findings are considered.

## Introduction

The exquisite precision of intracellular processes is not only dependent on the formation of highly specific, stable assemblies. Collections of weak, co-operative interactions result in the assembly of dynamic microenvironments and quinary structure that mediates most cellular events.[81,87,90,198] The disinclination to consider quinary structure has arisen mainly from the difficulties associated with the characterisation of dynamic microenvironments in the complex cell interior. Moreover, the weak nature of quinary interactions suggests that they may be disrupted by cell lysis[81] or skewed by test protein over expression, which is necessary for in-cell studies using low sensitivity techniques. Accordingly, quinary interactions are poorly understood.[90]

In the wake of post-reductionist biochemistry, the need for an increased understanding of quinary structure is apparent.[23,89] Such knowledge will bridge the gap between *in vitro* and *in vivo* biochemistry and enrich biologically-inspired fields such as supramolecular chemistry and chemical biology.[213,216,399] Several groups have therefore called for the development of new technologies suited to the study of weak interactions and short-lived microenvironments *in vivo*.[23,89,90,400] Despite the current absence of such technological advancements, key physicochemical mechanisms involved in quinary structure formation have been identified by in-cell and cell extract NMR studies.[74,79,87,100,198,384] For instance, many of these studies have indicated that (attractive and repulsive) charge-charge interactions are important in the regulation of quinary structure. Indeed, the role of charge-charge interactions in quinary structure formation was invoked during its original definition.[81] McConkey predicted that extensive, charge-based protein interactions lead to the formation of quinary structure which constrained protein evolution by only accommodating mutations at non-interactive sites.[81] New avenues related to quinary structure investigation have been opened by the exploration of the dynamic protein interactions that mediated liquid droplet formation in physiological environments.[24,46,49] These studies also indicate that attractive charge-charge interactions, typically involving low-complexity unfolded domains, drive droplet assembly and governs their unique compositions.[24,49] Indeed, the highly ionic nature of the cell interior[5,64] and the highly charged surfaces of hub proteins[401] supports the global importance of charge-charge interactions in regulating quinary structure.

A second defining feature of quinary interactions is their transient nature.[81,90] Structural analyses of the transient interactions of hub proteins suggests some of their key features and may be pertinent to quinary interactions.[401] Global protein flexibility and an abundance of charged residues were identified. Accordingly, the authors observed a high frequency of charge-charge, charge-polar and polar-polar contacts at the interface of transient hub complexes relative to stable hub complexes.[401] These features are expected to facilitate the competitive regulation of transient hub complexes with neighbouring partners and water.[401] Electron transfer (ET) complexes, such as those observed in photosynthesis and respiration, are the paradigm for transient protein-protein interactions. With flat, poorly packed interfaces,[83,266] ET complexes are considered the epitome of fast dissociation ($k_{off}$ > 1000 $s^{-1}$) and low-specificity.[268,402] These features enable dynamic and reversible interactions with multiple partners which is of obvious relevance to quinary structure. The similarity between ET and quinary (or quinary-like) interactions is also reflected in their similar $K_d$ values. Lantham and Kay performed NMR-monitored titrations of $^{13}C$ methyl-labelled calmodulin with concentrated *E. coli* extract aliquots to obtain a 'minimum average' effective $K_d$ for the calmodulin-extract interaction, which was ~0.2 mM.[388] ET complexes typically have $K_d$ values in the high μM-low mM range.[268]

*In vitro* studies have yielded important insights into the physicochemical basis for low-specificity protein-protein complexes that may be used as a springboard for the exploration of quinary-like interactions in physiologically-relevant environments. Confined and/or crowded samples such as polyacrylamide gels (PAGs)[403] and synthetic[404] or protein[237] crowders, are well-established proxies for the volume-occupied cell interior. However, reports describing the influence of confinement on protein-protein interactions are sparse, despite compelling evidence for the enhanced protein stability,[389,403,405] folding rates[406] and activity[407] effected in confined environments. It is generally accepted that the stability of high-affinity, heterodimeric protein-protein interactions remains closely similar in dilute and crowded solutions.[221] Some low affinity complexes appear to be significantly affected. For instance, aggregation was observed for the low-affinity CyPet and YPet FRET pair in samples crowded with > 250 g/L PEG 1000 or 8000.[404] By contrast, the weak interaction between the ε and θ subunits of the *E. coli* DNA polymerase III

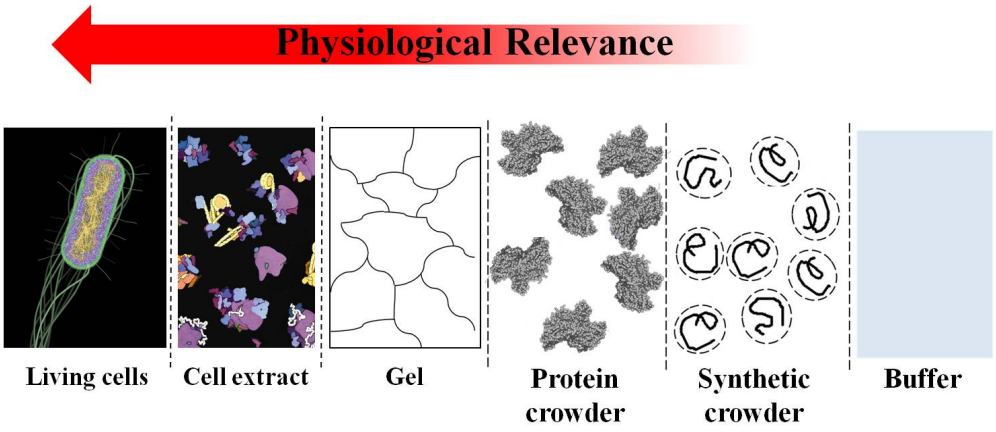



**Figure 1.** Knowledge of quinary-like interactions may be gained from studies of weak protein-protein complexes under conditions of increased physiological-relevance.

holoenzyme was stabilised by 1 kcal/mol in the presence of 100 g/L Ficoll or Dextran (both 70 kDa).[408] The stability of the weak, cognate cytochrome *f*-plastocyanin complex was unaffected in 50-70 g/L Ficoll70 and Dextran70.[409] Thus, within physiologically-relevant concentration ranges (50-400 g/L),[410] macromolecular crowding appears to differentially tune the stability of transient protein complexes. Although macromolecular crowding has a mild effect on the stability of specific complexes,[408,409] quinary interactions are weak and easily perturbed, indicating that they may be relatively non-specific.[81,87] Thus, knowledge of the effects of crowding and confinement on a low-specificity protein-protein interaction may reveal novel aspects of quinary structure.

Herein, the non-cognate ET pair, cytochrome *c* and flavodoxin were employed as models of a quinary-like interaction for studies in volume-occupied (*i.e.* crowded and confined) solutions (Figure 1). It is well established from standard solution studies that these oppositely charged proteins interact transiently with each other[286,411] and with various cognate partners[110,121]. $^{1}$H, $^{15}$N HSQC and $^{19}$F NMR was used to assess the cytochrome *c*-flavodoxin interaction in buffer and in volume-occupied solutions. Macromolecular crowders were used at concentrations precluding polymer overlap (*i.e.* 100 g/L),[237,412,390] such that the crowders behaved as mainly as individual particles and not confining networks. Polyacrylamide gels (PAGs) were used as a confining medium. This approach afforded the comparison of crowding and confinement effects on the cytochrome *c*-flavodoxin complex. Interestingly, the cytochrome *c*-flavodoxin interaction appeared to be stabilised in crowded or confined solutions. The stabilisation of a weak, low-specificity protein-

protein interaction in crowded or confined milieus supports the idea that charge-based quinary interactions are controlled by changes in intracellular volume occupancy. This mechanism may be used to stabilise quinary interactions involving oppositely charged proteins which are expected to be important in structuring macromolecular networks.[396]

**Experimental Design: Separating Crowding from Confinement.** Synthetic polymer solutions exhibit different phase behaviours within the defined concentration ranges: dilute, semidilute and concentrated.[237,390,412] In dilute regimes, polymer molecules behave mainly as independent, crowding entities.[390] In semi-

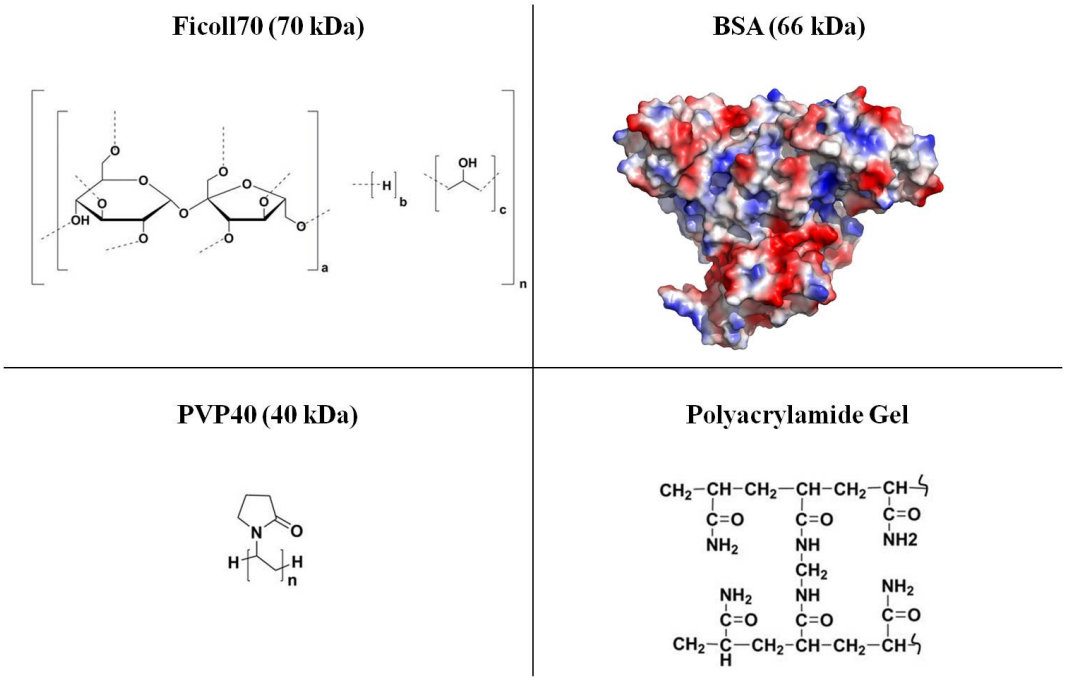



**Figure 2.** Structure of the synthetic crowders, BSA (electrostatic surface) and polyacrylamide gel (PAG).

dilute and concentrated solutions, the polymer chains entangle to produce porous networks with crowded (dynamically restricted) and confined (statically-restricted) characteristics.[412] Thus, to separate the effects of crowding and confinement on the cytochrome *c*-flavodoxin interaction, crowder solutions in the dilute regime (*i.e.* 100 g/L) were used.[390,413] The effects of three different crowders were compared (Figure 2): polyvinylpyrrolidone 40 (PVP40), Ficoll70 and bovine serum albumin (BSA). PVP40[414] (40 kDa) and Ficoll70[413] (70 kDa) are neutral, highly-soluble, synthetic polymers that are typically inert with respect to proteins. PVP40 is a random coil

polymer of *N*-vinylpyrrolidone. Ficoll70 is a branched sucrose polymer that is considered 'globular' relative to most other synthetic polymers. BSA is an anionic, globular protein (67 kDa, p*I* 4.7[237]) that interacts non-specifically with many proteins[237,73,79] and is therefore a more biologically-relevant crowder. The effect of confinement on the cytochrome *c*-flavodoxin interaction was explored in 8% PAGs which are neutral and have average pore dimensions of 8-12 nm[403] similar to the cytoskeleton.[5]

**Materials and Methods**

**Protein Expression and Purification** $^{15}$N-labelled and unlabelled *S. cerevisae* cytochrome *c*[371] and *E. coli* flavodoxin[121] were expressed in *E. coli* BL21 (DE3) and purified according to standard methods (Chapters 3-5).[74,371,119] $^{19}$F-tryptophan labelled flavodoxin was expressed in minimal medium containing 30 mg/L 5-fluoroindole (as per Chapter 4).[288]

**Crowder Stock Preparation** Ficoll70, polyvinylpyrrolidone 40 (PVP40) and bovine serum albumin (BSA) were purchased from Sigma Aldrich. 300 mg of each crowder was dissolved in ~1 mL Millipore water and dialysed against 2 L of Millipore water for 24 hours at 4 ˚C. The dialysate was concentrated to < 900 µL (miVac Duo Centrifugal Concentrator, Genevac), combined with 100 µL of 10 X NMR buffer and adjusted to pH 6.0 and a final volume of 1 mL.

**PAG Preparation** Polyacrylamide gels containing ($^{15}$N-labelled) test proteins were prepared directly in NMR tubes. Given the importance of pH in modulating protein interactions[271] and quinary structure,[87] PAG preparation was modified to ensure optimal pH control. A 30% acrylamide/bisacrylamide (SIGMA) mixture was dissolved to 8, 10 or 12% in NMR buffer containing 10% $D_2O$, ~10 mM TEMED and 0.001% riboflavin. The sample pH was adjusted to 6.0 prior to the addition of cytochrome *c* and/or flavodoxin. Acrylamide polymerisation was photoinitiated in 5 mm NMR tubes using 254 nm light for 40 minutes (Rayonet RMR 200 Photochemical Reactor). Flavodoxin-containing acrylamide mixtures were prepared and irradiated as described above but without riboflavin. PAGs did not form, confirming that the flavin mononucleotide group of flavodoxin is not implicated in acrylamide polymerisation, despite being chemically-related to the photoinitiator,

riboflavin. In keeping with earlier reports[389] higher concentration PAGs (> 10%) were difficult to prepare and fractured during polymerisation.

**NMR Spectroscopy** NMR samples typically contained 0.3 mM labelled protein and 10% $D_2O$ in NMR buffer at pH 6.0. The sample pH was monitored before and after each experiment. Notably, the pH remained stable and rarely required alteration. $^{1}H$, $^{15}N$ watergate[251] or TROSY-selected[292] HSQC spectra were acquired at 30 ˚C with 16 scans and 64 increments on a Varian 600 MHz Spectrometer equipped with a HCN coldprobe. The data were processed in Biopack (using linear prediction in the $^{15}N$ dimension) and NMRPipe.[386] The spectra were analysed using CCPN.[252] 1D $^{19}F$ spectra were acquired at 25 ˚C on a Varian 500 MHz Spectrometer equipped with OneNMR probe. Spectra were acquired with 1024 transients, 5435 k points and a spectral width of 231.1 ppm. A relaxation delay time of 0.7 s and an acquisition time of 0.05 s were used. Baseline correction was applied where necessary. The $^{19}F$ NMR data were analysed in MestReNova.[372]

**Phase separation** The hanging drop vapour diffusion method was used to observe phase separation in 0.3-1.2 mM samples of the test proteins at 21 ˚C. Drops were prepared by combining 1 µl of a cytochrome *c*-flavodoxin mixture with 1 µl of the reservoir solution containing 20 mM $KH_2PO_4$, 75 mM NaCl, and 100 g/L crowder (Ficoll70, PVP40 or BSA) at pH 6.0. Control drops containing either 0.3-1.2 mM cytochrome *c* or flavodoxin were prepared as above.

**Results and Discussion**

**Cytochrome *c* and Flavodoxin Interact Transiently in Buffer** Using NMR-monitored titrations the cytochrome *c*-flavodoxin interaction was explored in a buffer of 20 mM $KH_2PO_4$, 75 mM NaCl at pH 6.0. The total ionic strength of 100 mM represents the lower limit of physiological relevance.[61,393] Ideally, the complex should be characterised at physiological pH (7.4), however, amide proton exchange of cytochrome *c* resonances corresponding to binding site residues precludes characterisation at this pH (data not shown).[415] Additionally, flavodoxin is most stable at pH 6.2[119] so to optimise sample stability and NMR data collection, a pH of 6.0 was chosen. Figure 3 shows regions from overlaid $^{1}H$, $^{15}N$ HSQC cytochrome *c* spectra during a titration with unlabelled flavodoxin. Several cytochrome *c* resonances showed gradual chemical shift changes with increasing flavodoxin

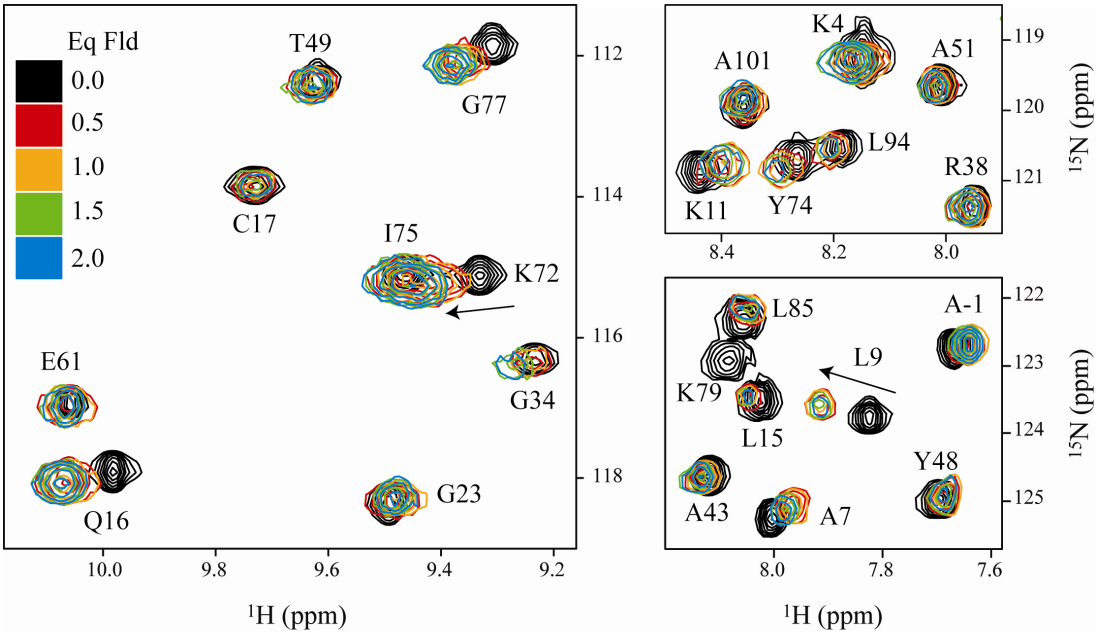



**Figure 3.** Regions from the overlaid $^{1}$H, $^{15}$N HSQC spectra of $^{15}$N-labelled cytochrome *c* during a titration with unlabelled flavodoxin.

concentration (Figure 3). This feature is characteristic of a fast exchange process on the NMR time scale[80,268,383] and, in agreement with earlier studies,[286,411,112] indicates that the cytochrome *c*-flavodoxin complex is transient. A general broadening of the resonances was also observed (*i.e.* average $^{1}H^{N}$ linewidth increase of 75%; Table 1), which is consistent with the reduced tumbling rates associated with complex formation (cytochrome *c* = 12.6 kDa; flavodoxin = 20.1 kDa; cytochrome *c* flavodoxin complex = 32.7 kDa).[268,112] The binding sites of cytochrome *c* were mapped according to the significant chemical shift changes effected during the NMR-monitored titrations (Figure 4). The chemical-shift map defines the cytochrome *c* binding site as a contiguous, charged patch surrounding the solvent-exposed haem co-factor (Figure 4). This is the known lysine-rich cytochrome *c* binding site.

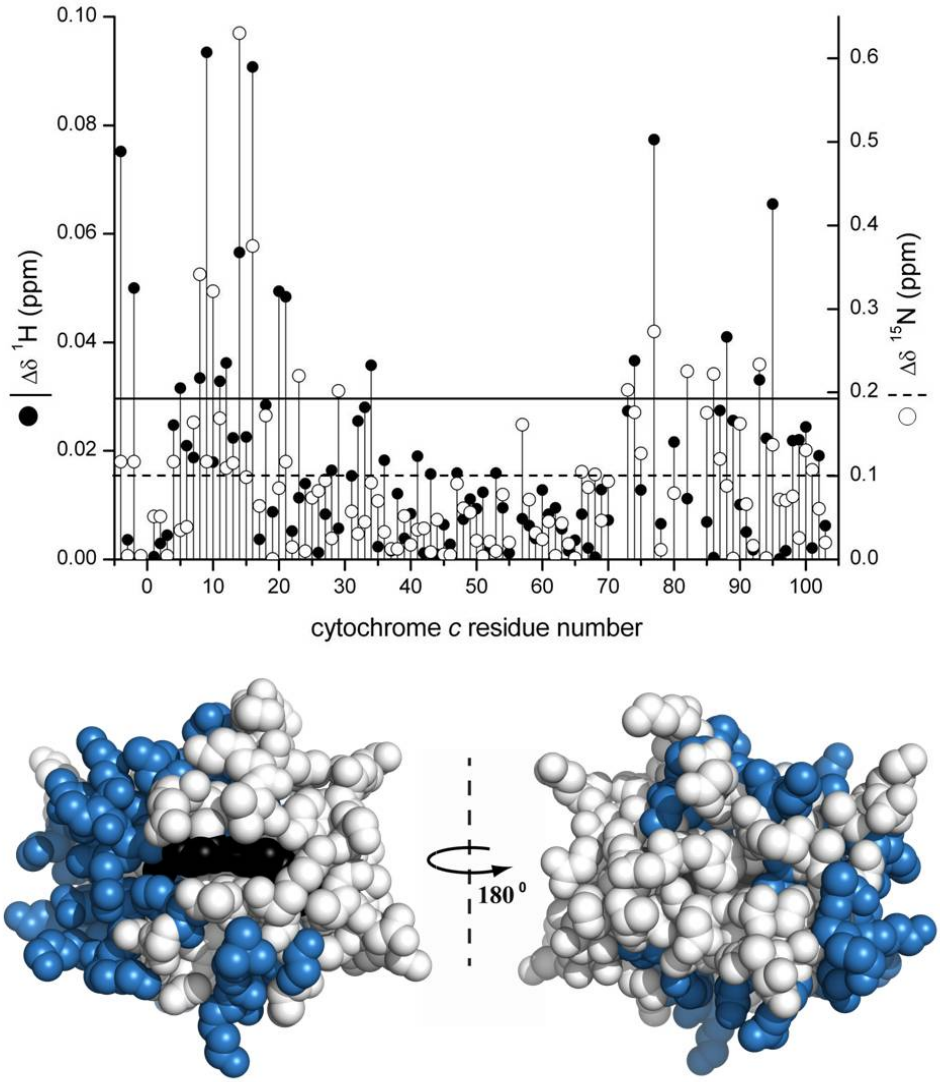



**Figure 4. Upper:** Plot of chemical shift perturbations measured for cytochrome *c* backbone amides in the presence of 2 equivalents of flavodoxin. Cytochrome *c* residues are numbered from –5 to 103. Blanks correspond to proline residues 25, 30, 71 and 76 and glycines 83 and 84, which were undetectable. During the titration, lysines 72 and 79 broadened beyond detection. The dashed and solid lines are the average chemical shift changes observed in the $^{1}H$ and $^{15}N$ dimensions, respectively. **Lower:** Space-filling representation of cytochrome *c* highlighting the flavodoxin binding patch. Residues for which the amide resonance had a significant chemical shift perturbation ($\Delta\delta\ ^{1}H^{N} \geq 0.03$ or $^{15}N \geq 0.1$ ppm) are in blue. The haem edge is black.

**Table 1.** Line-widths (Hz) of 22 non-overlapping cytochrome *c* $^{1}H^{N}$ resonances in 20 mM $KH_2PO_4$, 75 mM NaCl, pH 6.0 in the free form or in complex with 2 equivalents of flavodoxin.

| | $^{1}H^{N}$ line-widths (Hz) | |
|---|---|---|
| **Residue** | **Pure cytochrome c** | **Cytochrome *c* + flavodoxin** |
| F-3 | 13.7 | 22.2 |
| K-2 | 20.0 | 29.3 |
| A-1 | 12.3 | 21.4 |
| L9 | 15.2 | 32.7 |
| T12 | 17.8 | 36.4 |
| Q16 | 14.4 | 36.0 |
| R38 | 19.7 | 27.3 |
| G45 | 19.4 | 27.8 |
| Y46 | 20.2 | 27.8 |
| S47 | 19.4 | 30.8 |
| A51 | 13.6 | 25.9 |
| N52 | 17.2 | 35.5 |
| K55 | 13.7 | 24.8 |
| V57 | 18.5 | 25.1 |
| L58 | 17.6 | 26.7 |
| E61 | 14.7 | 35.0 |
| G77 | 21.4 | 28.5 |
| L85 | 21.1 | 40.1 |
| E90 | 15.7 | 31.4 |
| Y97 | 15.7 | 35.5 |
| K100 | 16.4 | 27.3 |
| E103 | 15.5 | 24.4 |
| Average line-width (Hz)* | Cytochrome *c* = 17.0 ± 2.7<br>Cytochrome *c* + flavodoxin = 29.7 ± 5.1 | |

* Error represents standard deviation of the mean

$^{19}F$ tryptophan-labelled flavodoxin was also studied by NMR during a titration with unlabelled cytochrome *c*. Four well-resolved peaks were observed in the $^{19}F$-flavodoxin NMR spectrum (Figure 5). Despite their simplicity, the $^{19}F$-flavodoxin spectra capture several features evident in the corresponding 2D NMR-monitored titration experiment. Broadening of all $^{19}F$-flavodoxin peaks was observed upon binding to cytochrome *c* (average $^{19}F$ linewidth increase of 56%; Table 2), and one peak shifted significantly during the titration (Figure 5, Table 2). Crystal structure[118] analysis revealed that two of the tryptophans, W57 and W120, are solvent exposed and that only W57 is proximal to the exposed flavin mononucleotide edge (Chapter 4, Figure 7). The fluorine atom of 5-fluorotryptophan 57 is expected to be within van der Waals distance of the solvent exposed flavin methyl groups and likely reports on cytochrome *c* binding.

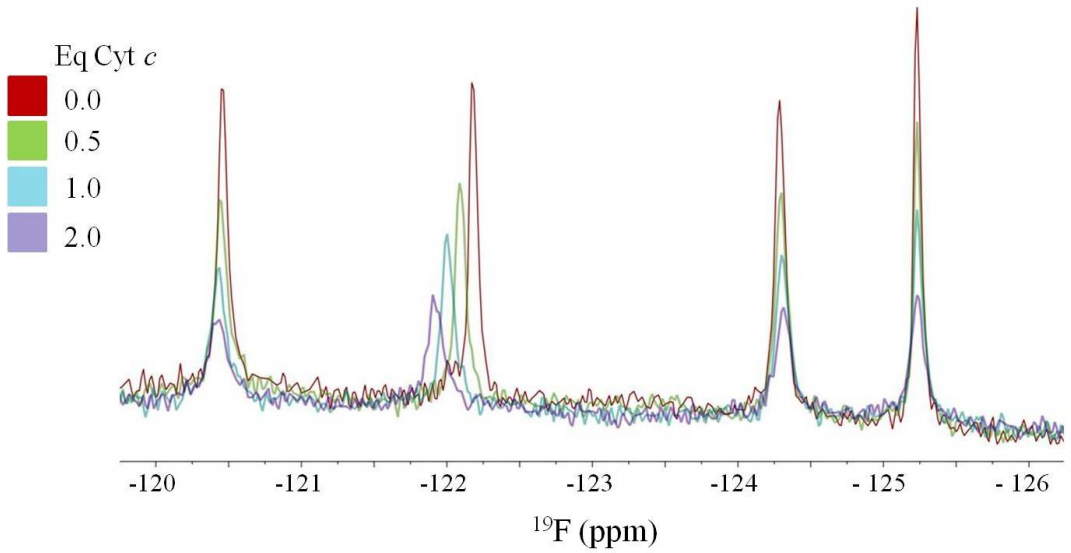



**Figure 5.** Overlaid $^{19}F$ spectra of pure 5-fluorotryptophan flavodoxin in buffer and in the presence of 0.5, 1.0 or 2.0.

**Table 2.** δ (ppm) and line-widths (Hz) of the $^{19}F$ peaks of flavodoxin in 20 mM $KH_2PO_4$, 75 mM NaCl, pH 6.0 in the free form or in complex with 2 equivalents of cytochrome *c*.

| | **Pure flavodoxin** | | **Flavodoxin + cytochrome *c*** | |
|---|---|---|---|---|
| **Peak** | **δ (ppm)** | **$^{19}F$ line-width (Hz)** | **δ (ppm)** | **$^{19}F$ line-width (Hz)** |
| 1 | -120.46 | 36.4 | -120.43 | 75.5 |
| 2 | -122.18 | 31.0 | -121.91 | 59.7 |
| 3 | -124.29 | 37.1 | -124.32 | 56.3 |
| 4 | -125.23 | 29.1 | -125.23 | 49.3 |
| Average line-width (Hz)* | | Pure flavodoxin = 33.4 ± 4.0<br>Flavodoxin + cytochrome *c* = 60.2 ± 11.1 | | |

* Error represents standard deviation of the mean

The cytochrome *c*-flavodoxin interaction was also explored in 2% agarose gels. Notably, Pastore and co-workers previously determined that the pore sizes of 1.5-4% agarose gels were too large to serve as a confining medium for smaller proteins (*e.g.* hen egg white lysozyme) and peptides.[416] Owing to the relatively small sizes of cytochrome *c* and flavodoxin, the 2% agarose electrophoresis gels are not expected to induce physiologically-relevant confinement effects but, instead, likely reveal aspects of the cytochrome *c*-flavodoxin interaction pertinent to dilute solutions. Figure 6 shows the migration of cytochrome *c*, flavodoxin and the complex at different protein ratios. The band due to free, cationic cytochrome *c* diminishes as a function of increasing flavodoxin concentration. Similarly, the band due to free flavodoxin becomes less intense with increasing cytochrome *c* concentration (Figure 6). The anionic orange band at 6.25 mm corresponds to the mixture of cytochrome *c* and flavodoxin in the complex. By adjusting the colour

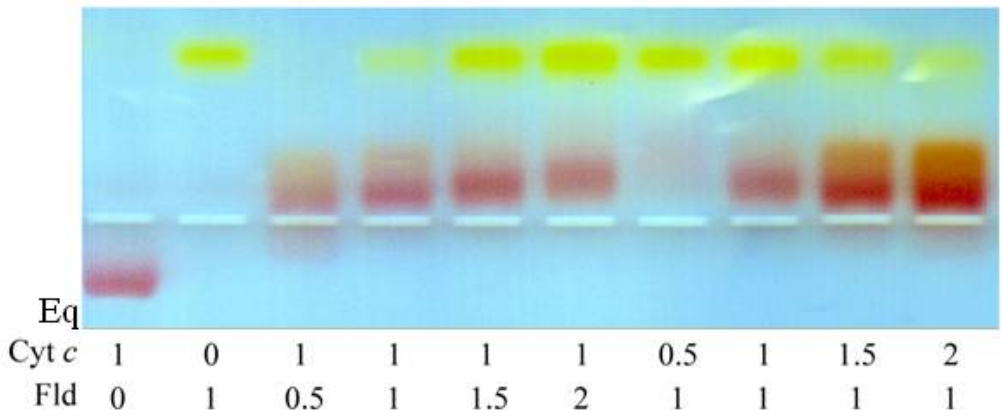



**Figure 6.** Cytochrome *c* and flavodoxin migration as a function of complex formation. 1 equivalent (1 Eq) corresponds to a concentration of 0.2 mM. The electrophoresis buffer was 20 mM $KH_2PO_4$, pH 6.0. Each sample (volume = 20 µL) contained 10% glycerol. The colour balance of the entire gel image has been adjusted to highlight blue, green and cyan hues.

balance of the gel image to highlight cyan, blue and green hues, the complex band separates into its red (cytochrome *c*) and orange (complex) components (Figure 6). The transient nature of the complex is confirmed by the streaking and separation of the 'complex' band due to cytochrome *c* dissociation. Taken together the NMR and electrophoresis data indicate that, in buffer, the cytochrome *c*-flavodoxin complex proceeds transiently via their known electron transfer sites.[74,268,109,121]

**Control NMR Experiments.** Control $^{1}H$, $^{15}N$ HSQC spectra were acquired for cytochrome *c* in the presence of 100 g/L crowders and 8% PAG (Figure 7). The spectra obtained in crowded or confined media were closely similar to the corresponding spectra obtained in buffer alone (Figure 7). Preliminary studies indicate that $^{15}N$-flavodoxin spectra were comparable in dilute solution and in crowded and confined conditions also (data not shown). These results indicate that the structures of cytochrome *c* and flavodoxin remain unperturbed in crowded or confined environments. Moreover, the absence of significant chemical shift changes for cytochrome *c* in the crowded and confined samples suggests that cytochrome *c*-crowder and cytochrome *c*-gel interactions were minimal under these conditions (Figures 7 and 8). The average cytochrome *c* $^{1}H^{N}$ line widths were calculated from 22 non-overlapping resonances in each control spectrum (Table 3). Cytochrome *c* becomes reduced during PAG preparation and a slightly different set of resonances were used for the extraction of $^{1}H^{N}$ data from the reduced cytochrome *c* HSQC. In crowded and confined solutions the average line widths for cytochrome were greater by 1.1-7.2 Hz than in buffer alone (Tables 1 and 3). In the case of the synthetic

crowder solutions and PAGs, average linewidth increases of 1.1-1.6 Hz were observed (Table 3). Previously, it was found that cytochrome *c* experienced significant chemical shift perturbations (0.1 ppm $^{1}$H and 0.3 ppm in $^{15}$N dimension) and a general linewidth increase of up to 35% in the presence of 300 g/L polyethylene glycol (PEG; Mn = 8 kDa). The insignificant chemical shift changes observed for cytochrome *c* in the presence of 100 g/L PVP40, Ficoll70 or the PAG as well as the ~5-10% average increase in $^{1}H^{N}$ linewidth suggest that cytochrome *c* does not interact with the neutral, highly soluble synthetic polymers or the gel at the concentrations employed herein. The linewidth increases likely arise from the increased viscosity of the crowded solutions. The average line width for cytochrome *c* resonances increased by ~40% in the presence of 100 g/L BSA compared to in buffer. At equal g/L concentrations however, solutions crowded with BSA are less viscous than solutions crowded with PVP40 or Ficoll70.[417] Thus, the increased rotational correlation time of cytochrome *c* in the presence of 100 g/L BSA arises from weak cytochrome *c*-BSA interactions. This idea is supported by the observation that the average cytochrome *c* line width decreased by ~10% when the ionic strength of the 100 g/L BSA solution was increased by 50 mM (Table 3). This suggests that screening of the charge-charge interactions weakens the BSA-cytochrome *c* interaction. Comparison of the chemical shift perturbation plots for cytochrome *c* in the presence of BSA (Figure 8) or flavodoxin (Figure 4) shows that the magnitude of cytochrome *c* chemical shift changes were insignificant in the BSA solution compared to those observed at 2 molar equivalents of flavodoxin. The small chemical shift changes observed in the presence of a large excess of BSA ($\Delta\delta$ $^{1}HN_{Avg}$ = 0.0032 or $^{15}N_{Avg}$ = 0.018 ppm) indicate that the cytochrome *c*-BSA interaction is substantially weaker than the cytochrome *c*-flavodoxin interaction.[383] This finding is in agreement with earlier work by Li and Pielak, who determined that the BSA-CI2 interaction has a $K_d$ of 35 mM.[418] Comparably small chemical shift changes were observed for cytochrome *c* in the presence of 100 g/L PVP40 and Ficoll70 (data not shown).

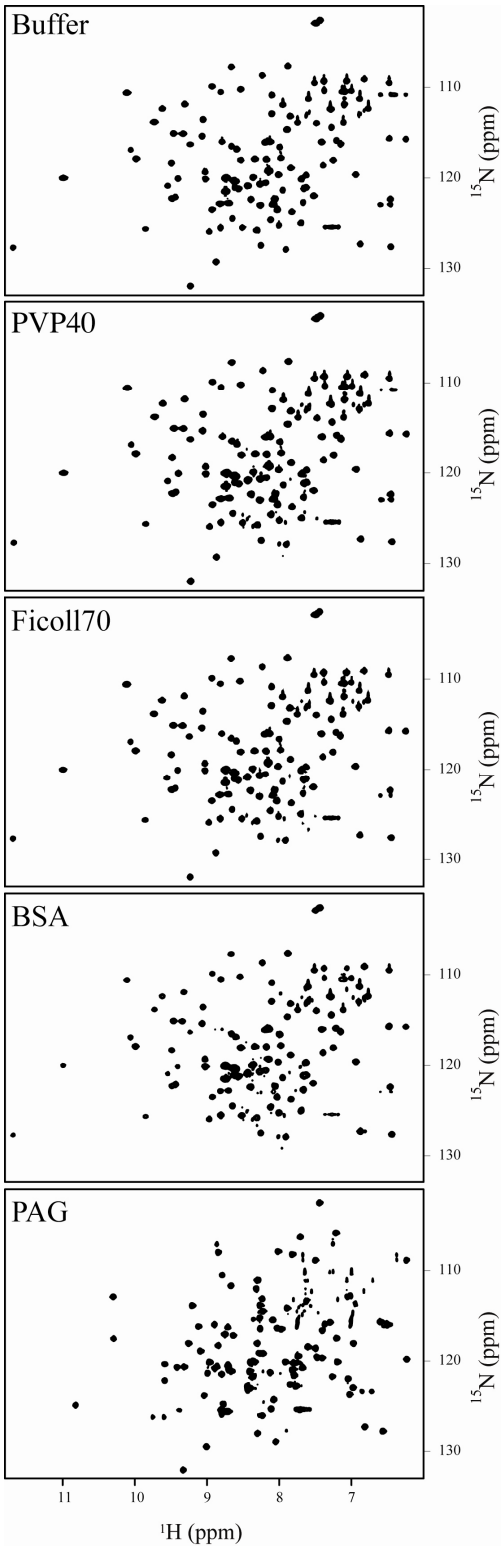



**Figure 7.** $^{1}$H, $^{15}$N HSQC spectra for cytochrome *c* in NMR buffer ± crowders (at 100 g/L) or 8% PAG. The spectra are contoured identically. Cytochrome *c* was oxidized prior to each experiment but becomes reduced during PAG preparation, producing a different HSQC. To eliminate background signals from PAG amides, $NH_2$-deselected HSQC experiments are used for data collection of PAG samples. Accordingly, side chain amides are not detected in PAGs.

**Table 3.** Line-widths (Hz) of 22 non-overlapping cytochrome *c* $^{1}H^{N}$ resonances in crowded or confined solutions containing 75 mM NaCl, 20 mM $KH_2PO_4$, pH 6.0

| $^{1}H^{N}$ cytochrome *c* line-widths | | | | | | |
|---|---|---|---|---|---|---|
| **Residue** | **PVP40** | **Ficoll70** | **BSA** | **BSA‡** | **Residue** | **PAG** |
| F-3 | 14.7 | 14.2 | 17.5 | 17.6 | K-2 | 19.9 |
| K-2 | 20.8 | 20.8 | 24.4 | 23.3 | G6 | 19.3 |
| A-1 | 13.7 | 12.4 | 18.1 | 16.1 | A7 | 15.2 |
| L9 | 17.7 | 16.9 | 28.9 | 23.4 | T12 | 19.0 |
| T12 | 19.9 | 19.0 | 26.4 | 24.2 | R13 | 22.6 |
| Q16 | 16.3 | 16.9 | 23.1 | 20.1 | Q16 | 13.5 |
| R38 | 20.3 | 19.7 | 22.7 | 21.4 | T19 | 24.8 |
| G45 | 19.4 | 20.6 | 25.7 | 23.5 | G23 | 20.7 |
| Y46 | 21.0 | 20.7 | 23.6 | 22.1 | G24 | 19.1 |
| S47 | 21.5 | 21.0 | 25.2 | 23.9 | K27 | 20.6 |
| A51 | 15.3 | 14.3 | 19.5 | 17.6 | V28 | 16.1 |
| N52 | 19.3 | 17.8 | 26.2 | 21.8 | S40 | 16.1 |
| K55 | 15.9 | 14.5 | 19.8 | 18.3 | G45 | 18.9 |
| V57 | 20.8 | 19.9 | 21.7 | 22.0 | Y46 | 19.1 |
| L58 | 18.6 | 18.4 | 22.5 | 22.1 | S47 | 19.5 |
| E61 | 16.0 | 16.3 | 22.9 | 19.4 | L58 | 19.2 |
| G77 | 23.8 | 22.6 | 29.0 | 27.4 | W59 | 16.4 |
| L85 | 22.0 | 21.8 | 30.4 | 25.6 | N63 | 17.3 |
| E90 | 17.8 | 16.8 | 32.1 | 26.0 | N70 | 14.9 |
| Y97 | 17.7 | 18.1 | 26.67 | 21.9 | K73 | 17.7 |
| K100 | 18.2 | 18.2 | 23.5 | 20.8 | F82 | 17.2 |
| E103 | 17.6 | 16.1 | 21.3 | 18.3 | E103 | 16.1 |
| Average $^{1}H^{N}$ line-width (Hz) | | | | PVP40 = 18.6 ± 2.6<br>Ficoll70 = 18.1 ± 2.7<br>BSA = 24.1 ± 3.9<br>BSA‡ = 21.7 ± 3.0<br>PAG = 18.3 ± 2.6 | | |

* Error represents standard deviation of the mean

‡ Solutions contained 100 g/L BSA, 125 mM NaCl, 20 mM $KH_2PO_4$, pH 6.0

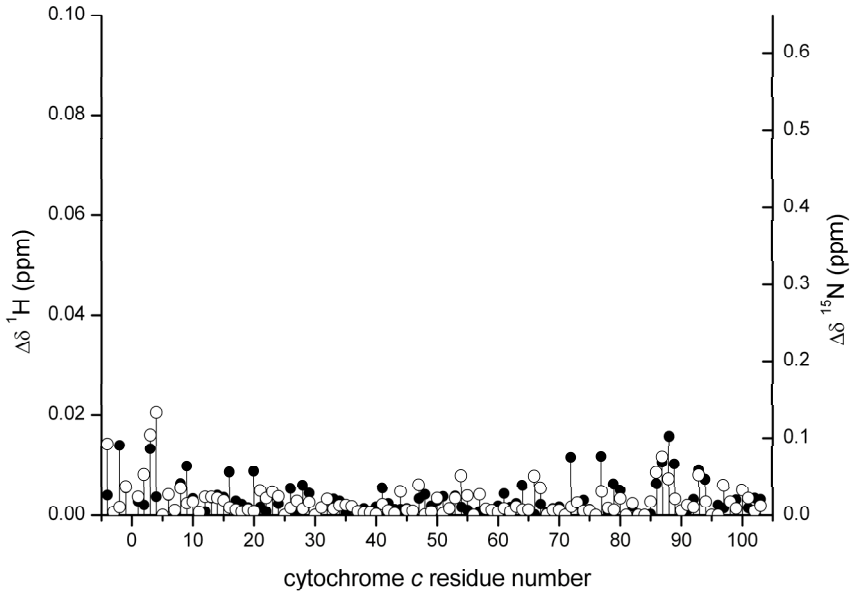



**Figure 8.** Plot of the chemical shift perturbations measured for cytochrome *c* backbone amides in the presence of 100 g/L BSA. The data are plotted at scales identical to those used in Figure 4.

**Effects of Crowding and Confinement on the Cytochrome *c*-Flavodoxin complex.** $^{1}H$, $^{15}N$ HSQC spectra were acquired for $^{15}N$ cytochrome *c* in the presence of 1 equivalent of unlabelled flavodoxin in dilute, crowded and confined conditions (Figure 9). Notably, severe broadening was observed for most of the cytochrome *c* resonances in volume-occupied solutions containing flavodoxin (Table 4) with 19-24 peaks broadened beyond detection in each case. Such acute broadening is characteristic of high affinity complexes in slow exchange on the NMR timescale.[80,392,308,419] Given that cytochrome *c* and flavodoxin interact more strongly with each other than with the crowders or the gel (Figures 3-8), it is likely that the spectral deterioration arises from the enhanced stability of the cytochrome *c*-flavodoxin complex in crowded and confined environments. The cytochrome *c* spectra were partially restored when the ionic strength of the solutions was increased by 50 mM (*i.e.* total buffer composition = 20 mM $KH_2PO_4$, 125 mM NaCl, pH 6.0) and the cytochrome *c*-flavodoxin interaction was disrupted (Figure 9 and Table 4).

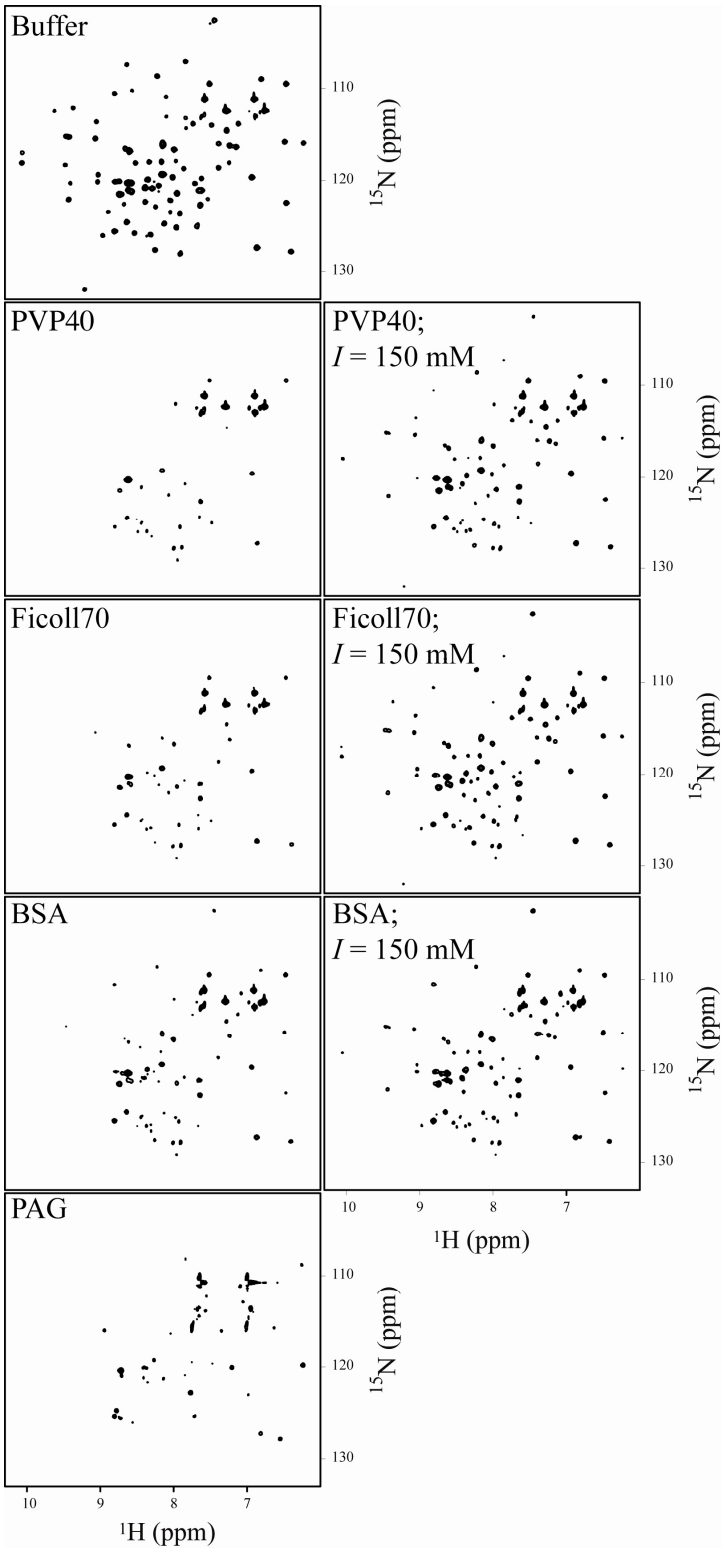



**Figure 9.** $^{1}H$, $^{15}N$ HSQC (identically contoured) spectra for $^{15}N$ cytochrome *c* in the presence of 1 equivalent of flavodoxin in dilute, crowded or confined (8% PAG) conditions, as indicated. The crowder concentration was always 100 g/L. The samples used to generate the spectra in the left hand column had a total ionic strength of 100 mM (20 mM $KH_2PO_4$, 75 mM NaCl, pH 6.0). The spectra in the right hand column were acquired from the same samples after spiking with an additional 50 mM NaCl.

**Table 4.** Linewidths (Hz) of cytochrome *c* $^{1}H^{N}$ resonances in samples containing 1 molar equivalent of flavodoxin + 100 g/L crowder in moderate and increased‡ ionic strengths

| $^{1}H^{N}$ cytochrome *c* line-widths | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Residue** | **PVP40** | **PVP40‡** | **Ficoll70** | **Ficoll70‡** | **BSA** | **BSA‡** | **Residue** | **PAG** |
| F-3 | 26.0 | 25.3 | 27.9 | 23.5 | 27.8 | 25.2 | K-2 | 29.0 |
| K-2 | 41.5 | 31.3 | 35.1 | 31.9 | 37.9 | 29.1 | G6 | 16.2 |
| A-1 | 29.1 | 25.1 | 27.4 | 21.5 | 25.1 | 23.5 | A7 | 27.6 |
| L9 | 38.7 | 37.6 | 44.3 | 36.5 | 26.6 | 37.1 | T12 | 42.6 |
| T12 | 33.7 | 46.7 | 34.9 | 28.6 | 37.9 | 34.0 | R13 | 35.0 |
| Q16 | 43.3 | 45.3 | 46.7 | 40.7 | 42.3 | 33.4 | Q16 | 27.6 |
| R38 | 45.3 | 32.3 | 36.5 | 29.2 | 30.3 | 28.2 | T19 | 20.1 |
| G45 | 47.1 | 29.9 | 39.5 | 36.1 | 34.9 | 29.2 | G23 | 29.9 |
| Y46 | 41.1 | 32.3 | 31.6 | 25.7 | 31.0 | 26.1 | G24 | 35.4 |
| S47 | 36.9 | 33.2 | 37.1 | 31.8 | 37.2 | 31.5 | K27 | 37.4 |
| A51 | 40.2 | 29.9 | 35.7 | 25.8 | 39.0 | 24.6 | V28 | 28.5 |
| N52 | 26.5 | 37.6 | 44.6 | 27.6 | 41.0 | 32.0 | S40 | 22.5 |
| K55 | 39.1 | 32.6 | 29.8 | 32.0 | 35.9 | 25.5 | G45 | 33.4 |
| V57 | 37.7 | 32.5 | 34.5 | 30.6 | 39.2 | 28.9 | Y46 | 25.9 |
| L58 | 32.0 | 33.6 | 32.2 | 28.6 | 31.4 | 30.5 | S47 | 33.9 |
| E61 | 43.2 | 41.8 | 33.7 | 40.3 | 64.1 | 30.9 | L58 | 30.5 |
| G77 | 49.2 | 37.7 | 33.0 | 28.0 | 48.8 | 30.0 | W59 | 37.4 |
| L85 | 15.7 | 17.4 | 16.7 | 34.6 | 39.1 | 40.9 | N63 | 36.2 |
| E90 | 39.5 | 34.2 | 36.9 | 34.2 | 34.1 | 25.5 | N70 | 30.0 |
| Y97 | 32.7 | 47.5 | 39.5 | 30.7 | 54.3 | 42.0 | K73 | 32.8 |
| K100 | 31.2 | 28.8 | 32.0 | 30.8 | 35.9 | 34.5 | F82 | 29.5 |
| E103 | 37.0 | 29.2 | 30.6 | 27.0 | 35.6 | 28.1 | E103 | 33.3 |
| Average $^{1}H^{N}$ line-width (Hz) | | | | | PVP40 = 37.0 ± 7.8<br>PVP40‡ = 33.7 ± 7.2<br>Ficoll70 = 35.0 ± 6.5<br>Ficoll70‡ = 30.7 ± 5.0<br>BSA = 37.7 ± 9.0<br>BSA‡ = 30.5 ± 5.0<br>PAG = 30.7 ± 6.1 | | | |

* Error represents standard deviation of the mean

‡ Solutions contained 100 g/L crowder, 125 mM NaCl, 20 mM $KH_2PO_4$, pH 6.0

**Crowding Effects Corroborated by Native Gels and Phase Separation.** Figure 10 shows the control electrophoresis gels of cytochrome *c* or flavodoxin in buffer and in 100 and 400 g/L PVP40, Ficoll 70 and BSA at pH 6.0. Flavodoxin migration was essentially identical in buffer and in solutions containing crowders at 100 g/L. The migration distance of cytochrome *c* decreased by ~25 % in the presence of 100 g/L BSA, probably due to weak cytochrome *c*-BSA interactions. Interestingly, interactions with BSA do not alter the electrophoretic mobility of cytochrome *c* to the same degree as interactions with flavodoxin. This data is in keeping with the NMR results, indicating that the cytochrome *c*-BSA interaction is significantly weaker than the quinary-like cytochrome *c*-flavodoxin interaction. Cytochrome *c* migration was also impeded in the presence of 100 g/L PVP40, the most viscous crowder solution.[417] The differential effect of 100 g/L PVP40 on cytochrome *c* and flavodoxin mobility in agarose gels results from the different net charges on the

protein. The electrophoretic mobility of a protein depends directly on its effective charge and solution viscosity ($\mu = z_{eff}.6\pi\eta R_s$). Since flavodoxin has a higher effective charge than cytochrome *c*, it will migrate further in viscous solutions compared to the weakly charged cytochrome *c*. Indeed, flavodoxin migration was unaffected in solutions containing 420 g/L glycerol, which has a higher viscosity that 100 g/L samples of Ficoll70 and BSA.[417] By comparsion, cytochrome *c* mobility was hampered at 350 and 420 g/L glycerol (Figure 11). Importantly, at a macromolecular crowder concentration of 400 g/L, both cytochrome *c* and flavodoxin mobility were substantially impeded due to crowding, confinement and viscosity with the largest effects imposed by BSA (Figure 10). This recapitulates the importance of studying crowding effects below the overlap concentration (~100 g/L) of synthetic polymers.

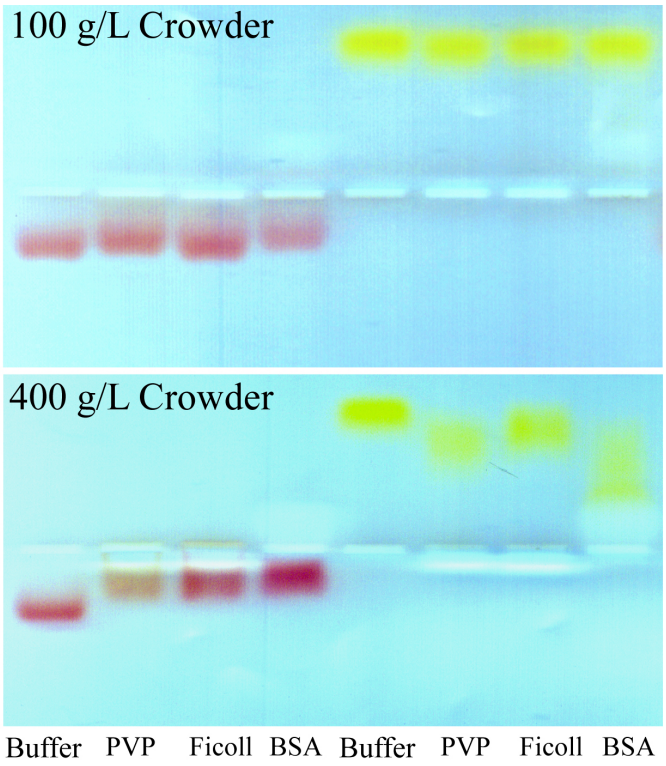



**Figure 10.** 2% Agarose gel electrophoresis of cytochrome *c* or flavodoxin in dilute and crowded samples containing 20 mM $KH_2PO_4$ pH 6.0 (sample volume = 20 µL). The concentration of the crowder samples is indicated. The electrophoresis buffer was 20 mM $KH_2PO_4$ pH 6.0. Note that the lanes labelled PVP and Ficoll contain the crowders PVP40 and Ficoll70, respectively. Lanes labelled "Buffer" refer to dilute samples.

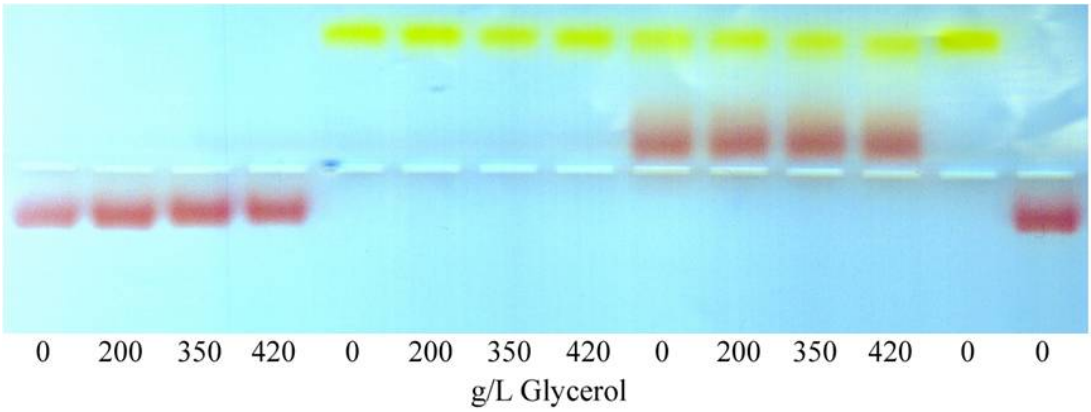



**Figure 11.** 2% Agarose gel electrophoresis of cytochrome *c* and/or flavodoxin at different g/L concentrations of glycerol. The concentration of glycerol in each sample is indicated below each lane. The sample volume was always 20 μL and the concentration of cytochrome *c* and/or flavodoxin is 0.2 mM. The sample and electrophoresis buffer was 20 mM $KH_2PO_4$ pH 6.0.

The cytochrome *c*-flavodoxin complex was also explored under dilute and crowded conditions by native gel electrophoresis (Figure 12). The rendered gel image dissects the 'complex' band into the red and orange components due to free and flavodoxin-bound forms of cytochrome *c*, respectively. It is difficult to observe differences in the relative proportion of free and bound cytochrome *c* in buffer and in solutions containing 100 g/L PVP40 or Ficoll70. This result is likely due to the fact that the synthetic crowders are charge neutral and will not migrate in the applied electric field. The effect of the crowders is therefore lost as the complex migrates into the gel. A significant increase in the intensity of the orange, complex band can be observed in solutions containing 100 g/L BSA (Figure 12). This suggests that the greatest complex stabilisation is effected by the most physiologically-relevant crowder and that this finding is pertinent to cytoplasmic interactions. Phase separation experiments were used to explore the effects of 100 g/L PVP40, Ficoll70 and BSA on cytochrome *c* and flavodoxin solubility (Figure 13). Intriguingly, drops that contained a synthetic crowder plus cytochrome *c* and flavodoxin showed orange liquid phase separations or amorphous precipitates between 2 and 4 days after setup (Figure 13). The control drops (one protein only) remained clear during this time. Precipitates were notably slower to form in the presence of 100 g/L BSA, suggesting that, although the cytochrome *c*-flavodoxin complex is favoured in solutions crowded with BSA, precipitation is attenuated.

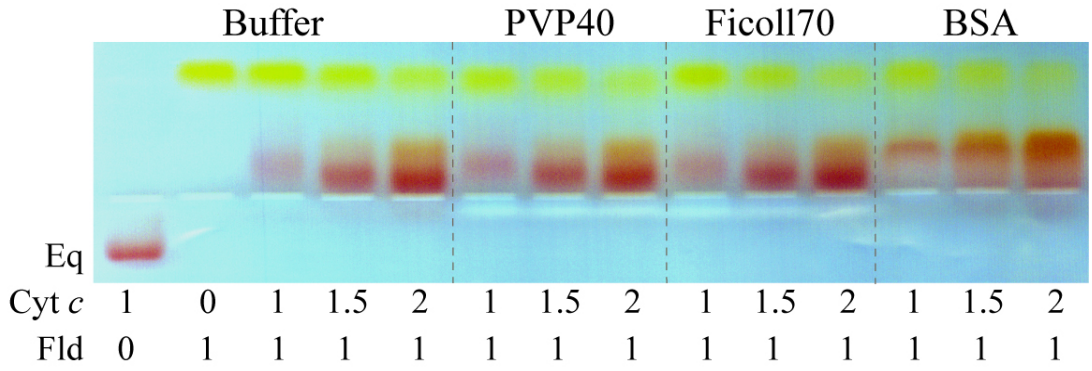



**Figure 12.** 2% Agarose gel electrophoresis of cytochrome *c* - flavodoxin migration in the absence or presence of crowders at 100 g/L. 1 equivalent of protein corresponds to a concentration of 0.2 mM. The sample and electrophoresis buffer was 20 mM $KH_2PO_4$ pH 6.0. The sample volume was always 20 µL.

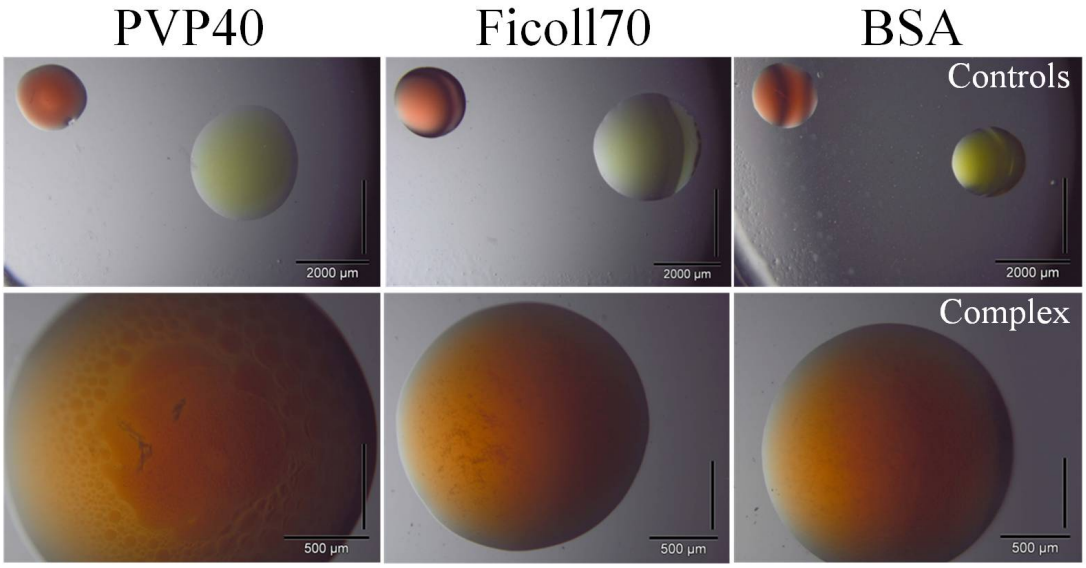



**Figure 13.** Hanging drops containing100 g/L PVP40, ficoll70 or BSA in 20 mM $KH_2PO_4$, 75 mM NaCl, pH 6.0 plus 0.6 mM cytochrome *c* and/or 0.6 mM flavodoxin. The control drops (upper row) contain one of the test proteins and the test drops contain both proteins.

**Effect of Volume Occupancy on the Quinary-Like Interaction of Oppositely Charged Partners** Although studies of quinary structure are in their infancy, early investigations have indicated the critical role of charge in regulating quinary structure.[81,87,384] Similarly, cell extract NMR, SEC and native gel electrophoresis revealed that the attractive quinary-like interactions involving cytochrome *c* and *E. coli* macromolecules were governed by charge (Chapters 3 and 5; reference 74). Intriguingly, these pervasive interactions were not disrupted in fourfold diluted extracts (Chapter 5). This finding suggests that charge-based quinary-like interactions are stabilised by physiological environments. This idea contrasts with

the expected ubiquity of weak interactions in forming quinary structure. This hypothesis prompted an investigation of the interaction between the oppositely charged cytochrome *c*-flavodoxin pair in volume-occupied conditions. As ET proteins, these partners are considered the paradigm for weak interactions.

NMR-monitored titration studies in dilute, buffered solutions indicated that the cytochrome *c*-flavodoxin interaction is transient. Native gel electrophoresis showed that the complex is anionic in nature which was reminiscent of the anionic cytochrome *c*-containing complexes identified in *E. coli* extracts (Chapters 3 and 5). This finding recapitulated the suitability of the cytochrome *c*-flavodoxin interaction as a mimic of those governing quinary structure. Notably, the $^{1}H$, $^{15}N$ spectrum for cytochrome *c* showed dramatically-increased line-widths in the presence of 1 molar equivalent of flavodoxin and 100 g/L crowders or in an 8% PAG (Table 4). This spectral feature is frequently observed for static complexes,[80,392,419] indicating that the cytochrome *c*-flavodoxin complex is stabilised in volume-occupied solutions. This stabilisation was partially mitigated by an ionic strength increase of 50 mM (total $I$ = 150 mM). Thus, ionic strength changes may counteract the stabilising effects of macromolecular crowding. However, it is likely that the disruptive effect of the physiologically-abundant anion, glutamate, is less pronounced than that of chloride.[74,100,245] The enhanced stability of the cytochrome *c*-flavodoxin complex effected in volume-occupied solutions may therefore represent an important mechanism for driving charge-based quinary interactions. Indeed, the interactions of small basic proteins are expected to be of critical importance in structuring macromolecular networks.[396] Thus, the stabilisation of charge-based complexes may act to preserve formative elements of quinary structure.

**Conclusions**

A major goal of biochemistry is to develop molecular-level models of the cell. Although achieving this aim ultimately relies on sophisticated *in vivo* measurements, valuable information may also be gleaned from studies of model proteins/systems in increasingly native-like environments. This approach may also help to bridge the gap between *in vitro* and *in vivo* biochemistry. Herein, the complex between cytochrome *c* and flavodoxin was explored in dilute and volume-occupied solutions. We propose that this interaction, involving oppositely charged ET proteins, is a model for a

quinary-like interaction owing to the dynamic nature and low specificity of many ET complexes. The data suggest that the complex is stabilised in crowded and confined conditions. This stabilisation is especially apparent in the presence of BSA, a physiologically-relevant, anionic protein crowder. This finding challenges the notion that all quinary structure-forming interactions are weak and dynamic, emphasising the complex nature of cytoplasmic structure and its effect on predicted 'weak' interactions.

# Discussion

**Overview**

Protein interactions are a crucial component of all cellular phenomena. The vast majority of knowledge relating to protein interactions has come from studies performed in dilute, simple solutions. However, it is now appreciated that protein behaviour *in vivo* may be modified by low-specificity interactions with the complex, volume-occupied cytoplasm. Studies of these pervasive 'quinary'[81] interactions are in their infancy but it is expected that quinary interactions organise the cytoplasm into spatiotemporal functional zones. Knowledge of the physicochemical basis of quinary interactions will therefore provide unprecedented insight into the molecular underpinnings of intracellular processes. The aim of this work was to advance the current understanding of quinary interactions and cytoplasmic structure by exploring the interaction propensities of three test proteins (ΔTat-GB1, cytochrome *c* and flavodoxin) in the *E. coli* cytoplasm and other physiological environments. Owing to its suitability for studies of transient protein interactions at atomic-level resolution,[80,383] NMR spectroscopy was used. Indeed, in-cell NMR spectroscopy has emerged as a useful indicator of test protein 'stickiness' whereby sticky proteins tend to engage in high molecular weight assemblies that tumble too slowly to be detected by HSQC NMR and related experiments.[71,73,74,76,78,230,232] The test proteins employed herein differ in size, p*I* and surface charge density (Introduction) and were therefore expected to exhibit different degrees of stickiness and NMR detectability in *E. coli* and extracts. However, as observed previously for cytochrome *c*,[74] ΔTat-GB1 and flavodoxin were completely undetectable by in-cell NMR (Chapter 2 and 5). This finding indicates that all three test proteins engage extensively in attractive quinary interactions with *E. coli* macromolecules.

ΔTat-GB1's quinary-like interactions were assessed in concentrated *E. coli* extracts using size exclusion chromatography (SEC; Chapter 2).[100] SEC revealed that ΔTat-GB1 (7.7 kDa) interacted pervasively with *E. coli* macromolecules in DNase I-pretreated extracts, despite the absence of physiological partners for the Tat motif. The resulting high molecular weight assemblies (> 44 kDa) were not disrupted during SEC in buffers containing (i) 50-200 mM NaCl or 100 mM of (ii) the physiologically abundant salt KGlu or (iii) most laboratory ions. ΔTat-GB1 interactions were disrupted when the extracts were pretreated with RNase A or

during SEC in a buffer containing 35 mM $MgCl_2$ ($I$ = 100 mM). These findings indicate that ΔTat-GB1 interacted preferentially with *E. coli* RNA. Given the affinity of $Mg^{2+}$ for phosphate oxyanions, the data suggest that $Mg^{2+}$ disrupts ΔTat-GB1 interactions with *E. coli* RNA by displacing arginine-phosphate salt bridges. ΔTat-GB1 was rendered NMR detectable in *E. coli* extracts pretreated with RNase A and spiked with > 10 mM $MgCl_2$. The need for both RNA digestion and $Mg^{2+}$ indicates the role of the chelate effect[264] in driving numerous arginine-phosphate salt bridge interactions between the disordered arginine-rich motif of ΔTat-GB1 and the RNA backbone. This finding coincides with the role of arginine/lysine-phosphate interactions in driving non-specific protein-RNA interactions[258,420] and also agrees with the idea that some RNA-binding proteins interact at sites lacking defined recognition elements.[263] The identification of ΔTat-GB1 interactions with RNA is also consistent with recent evidence indicating that protein-RNA interactions are an important component of quinary structure.[79] Moreover, similar arginine-rich peptide/RNA systems have been used to explore the roles of charge, multivalency and structural plasticity in facilitating supramacromolecular system assembly (*i.e.* reversible compartmentalisation).[56] Studies of such processes in composite, native-like environments may reveal the effects of quinary interactions on transitions facilitating micrometer-scale architecture formation from Ångström-scale structures.

Previously, cell extract NMR and SEC were used to identify the role of charge in driving non-specific interactions between cytochrome *c* and *E. coli* macromolecules.[74] Chapter 3 describes how electrophoresis was used to explore cytochrome *c* charge in buffer and in extracts. Membrane confined electrophoresis (MCE) revealed that, in a buffer containing 100 mM $KNO_3$ at pH 7.0, the effective valence of cytochrome *c* is + 1.04. This value is ~8 times lower than the calculated charge (based on the primary structure). Moreover, cytochrome *c* valence is dependent on the buffer composition. For instance, cytochrome *c* was charge neutral in the presence of 35 mM $K_2SO_4$ ($I$ = 100 mM), indicating that $SO_4^{2-}$ binds tightly to cytochrome *c*, neutralising its charge. In agreement with the MCE data, native gel electrophoresis showed that cytochrome *c* migrated toward the cathode (Chapters 3, 5 and 6). The migration distance of cytochrome *c* was notably short in native gels (compared to flavodoxin) and was dependent on buffer, ionic strength and pH. In *E. coli* extracts, cytochrome *c* migrated toward the anode, indicating that it interacted

with macroanions to produce anionic assemblies. Indeed, cytochrome *c* was identified in anionic assemblies at pH 6.0 or 7.0 in two- and fourfold serially diluted samples. Similarly, cytochrome *c* was never NMR detectable in extracts despite dilution of the cellular components. These data indicate that cytochrome *c* interactions with *E. coli* macroanions are relatively strong. SEC was used to identify whether cytochrome *c* interactions in extracts could be disrupted by DNase I or RNase A pretreatment or by charge neutralisation in buffers containing 35 mM $K_2SO_4$. In each case, cytochrome *c* eluted in the high molecular weight fractions, suggesting that it interacts preferentially with *E. coli* proteins. Cytochrome *c* interactions were disrupted during SEC in buffers containing of 35 mM $MgCl_2$ but the protein was not NMR detectable in extracts spiked with $MgCl_2$.

Using SEC, Crowley identified that single (R13E), double (R13E/K73E) and triple (R13E/K73E/K87E) cytochrome *c* mutants showed progressively fewer interactions in *E. coli* extracts.[74] The role of charge-charge interactions in governing cytochrome *c* stickiness was suggested. These mutants, and a quadruple (R13E/K73E/K87E/K100E) mutant, were characterised by MCE in a buffer containing 10 mM BIS-TRIS propane, 100 mM KCl pH 7.0 (Chapter 3). The effective valences of WT cytochrome *c* and the single, double, triple and quadruple mutants were 1.08, 0.71, 0.46, 0 and 0, respectively. In keeping with Crowley's SEC findings,[74] native gel electrophoresis showed that cytochrome *c*'s interactions with *E. coli* proteins diminished further with each additional K/R→E mutation. Despite their net charge of zero and expected low solubilities, the triple and quadruple mutants were not preferentially solvated[307] by *E. coli* macromolecules. They are preferentially hydrated. This finding coincides with the fact that glutamate is one of the most strongly hydrated side chains[421] and that the replacement of surface residues to glutamate can enhance protein solubility.[310]

Although undetectable in living *E. coli* cells, flavodoxin was weakly detectable in *E. coli* extracts and substantial improvements in peak linewidths and signal-to-noise ratios were effected in two- and fourfold serially diluted extracts (Chapter 5). Native gel electrophoresis revealed that flavodoxin migration was unaltered by cell extracts indicating that the role of charge was minimal in driving non-specific flavodoxin interactions in extracts as expected for such a highly anionic protein.

Native gel electrophoresis showed that WT cytochrome *c* interacts with flavodoxin to produce an anionic complex reminiscent of the anionic, cytochrome *c*-containing assemblies identified in *E. coli* extracts (Chapters 5 and 6). This finding suggests that the low-specificity cytochrome *c*-flavodoxin interaction is a useful model for 'quinary-like interactions'. Although it is well-established that the cytochrome *c*-flavodoxin complex is transient in dilute solution, NMR studies revealed that the complex is stabilised by crowding or gel confinement. Notably, a low degree of crowding (100 g/L) was sufficient to produce a tight complex. Native gel electrophoresis suggested that the greatest complex stabilisation was effected by the anionic protein BSA. This latter result, in particular, highlights the importance of using physiologically-relevant sample compositions.

The exploration of cell extracts was presented herein as a means of overcoming the debilitating effects of intracellular complexity on biophysical methods routinely used to study protein interactions. The combined use of SEC, NMR and native gel electrophoresis facilitated the dissection of ΔTat-GB1, cytochrome *c* and flavodoxin interaction propensities in physiological solutions. A novel, inexpensive strategy for protein labelling with fluorotryptophan was also presented (Chapter 4),[288] that involves supplementing minimal medium with 30-60 mg/L of the fluorinated tryptophan precursor 5-fluoroindole and no other additives. This method is suited to systems that over-express proteins and will extend the utility of $^{19}$F NMR spectroscopy for the investigation of biophysical protein properties under native-like conditions.

**Comparison of Test Protein Surfaces: Insights into Quinary-Like Interactions**

Biomolecular interactions are an inevitable consequence of the high surface-to-volume ratios imposed by macromolecular crowding. Consideration of protein surfaces and their known interaction propensities in physiological media may help to elaborate the principles of quinary-like interactions. The three test proteins studied herein displayed different degrees of stickiness and NMR detectability in concentrated *E. coli* extracts. They can be ranked cytochrome *c* > ΔTat-GB1 > flavodoxin in order of decreasing stickiness. This rank also tallies with the isoelectric point (Table 1), suggesting an over-arching role of protein charge in governing stickiness. This idea is consistent with the fact that cationic surface patches are a

major contributor to *E. coli* protein insolubility during over-expression in cell free expression systems.[422,423] Cytochrome *c* was never NMR detectable in extracts, despite spiking with $Mg^{2+}$ or extract dilution. This indicates that cytochrome *c* interacts most pervasively with *E. coli* macromolecules, probably due to its overall high surface charge density (12.6 kDa, up to 21 cationic residues at physiological pH). The stickiness of ΔTat-GB1 is conferred solely by the 11-residue arginine-rich Tat motif which can adopt α-helical or random coil conformations, depending on its environment.[424,425] ΔTat-GB1 was detectable in extracts pretreated with RNase A and spiked with $MgCl_2$. Thus, despite the high local charge density of the short Tat-motif, ΔTat-GB1 is probably less sticky than cytochrome *c* due to its lower total number of cationic residues. Although highly anionic, flavodoxin was only weakly detectable in extracts. GB1, by comparison, is highly detectable by NMR in *E. coli* cells and extracts[74,76,230] despite having a higher p*I* than flavodoxin. GB1 never formed a boundary during MCE experiments, indicating that it is charge neutral at a physiologically-relevant ionic strength and pH. This challenges the idea that protein stickiness correlates solely with net charge. Surface properties other than protein charge must therefore tune stickiness. For instance, flavodoxin is larger than GB1 (Table 1). Thus, flavodoxin has a larger surface area with which to engage in co-operative, attractive interactions. Interestingly, ~1/3 of ionisable residues in both flavodoxin and GB1 are cationic at physiological pH. However, GB1 lacks arginine residues, while flavodoxin has 4 arginines. Given arginine's versatility in its contribution toward protein interactions[25,26], its frequent role as a hot spot residue[257] and in the cationic surfaces patches of proteins that are poorly soluble during cell-free expression,[422,423] the difference in arginine content may partly explain the difference in stickiness observed for flavodoxin and GB1.

**Table 1.** Ionizable residues, p*I* and accessible surface areas of GB1 and the test proteins.

| Number of residues | GB1-QDD | ΔTat-GB1 | Cytochrome *c* | Flavodoxin |
|---|---|---|---|---|
| Arginine | 0 | 6 | 3 | 4 |
| Lysine | 6 | 8 | 16 | 9 |
| Histidine | 0 | 0 | 4 | 5 |
| Aspartate | 7 | 7 | 4 | 20 |
| Glutamate | 5 | 6 | 7 | 17 |
| Isoelectric point | 4.5 | 8.2* | 9.6 | 3.5 |
| Surface Area ($Å^2$) | 3700 | ~ 6500 | 6100 | 8000 |

* p*I* calculated using ProtParam (ExPASy)

Majumder and co-workers recently used in-cell NMR spectroscopy to identify the binding sites of four sticky proteins in *E. coli* and HeLa cells.[79] They found that protein quinary interactions involved their known binding sites,[79] suggesting that trends in protein recognition identified *in vitro* also govern quinary structure. For instance, it is likely that quinary structure may be specified by electrostatic interactions between hydrophilic groups but stabilised by hydrophobic interactions.[83,428–431] Evidence for the importance of co-operative, low-specificity hydrophobic interactions comes from in-cell NMR studies indicating that a large hydrophobic surface patch of ubiquitin mediates its quinary interactions.[79] Weaker ubiquitin-cytoplasm interactions resulted from mutating hydrophobic residues in the 'sticky' ubiquitin surface.[79] A cursory glance at the hydrophobicity of the test protein surfaces indicates that flavodoxin has the greatest number of hydrophobic patches followed by cytochrome *c* and then GB1 (Figure 1). Non-specific interactions between flavodoxin and *E. coli* macromolecules may therefore involve its hydrophobic surface patches, providing another explanation for the increased stickiness of flavodoxin relative to GB1. Flavodoxin has 13 solvent exposed hydrophobic residues, many of which are clustered (*e.g.* W56, Y57 and Y58) and involve residues most frequently observed at protein-protein interfaces: valine, leucine, isoleucine, phenylalanine, tryptophan and tyrosine.[83,428] GB1 contains only 4 surface exposed hydrophobic residues: I6, L7, V29 and V54. Cytochrome *c* contains 7 solvent exposed hydrophobic residues amongst those frequently observed at protein-protein interfaces. Four of these residues (V28, Y48, Y74 and F82) line the haem exposed edge in a non-contiguous manner. The role of cationic side chains in governing cytochrome *c* stickiness indicates that its non-clustered hydrophobic residues are not implicated in quinary interactions. Clustered hydrophobic surface residues such as those observed on flavodoxin likely drive hydrophobic quinary interactions. This idea relates to a computational simulation of the *E. coli* cytoplasm by McGuffee and Elcock showing that the known translational and rotational diffusion coefficients for GFP could not be reproduced by modelling the electrostatic and steric interactions between the 50 most abundant *E. coli* proteins.[432] Realistic GFP diffusion co-efficients were obtained only when short-range attractions between exposed hydrophobic atoms were included in the model.[432]

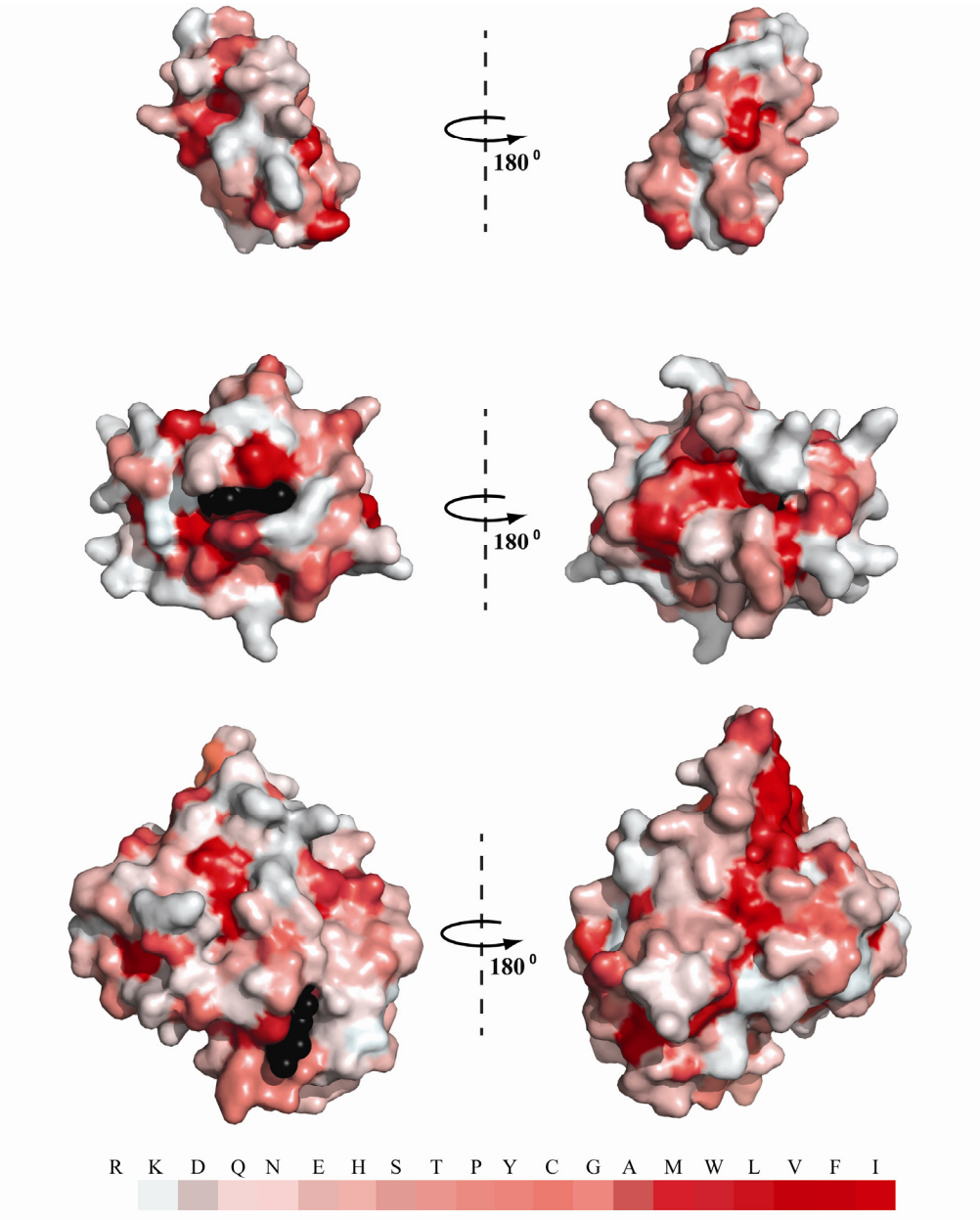



**Figure 1.** Surface representations of GB1 (top), cytochrome *c* (middle) and flavodoxin (bottom) coloured from white to red in order of increasing hydrophobicity according to the Eisenberg hydrophobicity scale.[433] The figure was generated in Pymol. The co-factors are shown as black spheres.

Knowledge of the effective valence and cell extract interactions of cytochrome *c* and its charge inverted mutants emphasise the need to consider protein solubility and amino acid composition to understand quinary-like interactions. Pure cytochrome *c* is soluble up to 5.5 mM in low ionic strength buffer (20 mM $KH_2PO_4$, 50 mM NaCl, pH 6.0).[74] However, cytochrome *c* precipitation occurs readily (*i.e.* after 40-45 minutes at 30 ˚C) in concentrated *E. coli* extracts, even those spiked with 35 mM $MgCl_2$. Although strongly hydrated in buffer, cytochrome *c* is preferentially solvated

by anionic proteins in *E. coli* extracts even after twofold serial dilutions with a low ionic strength buffer (20 mM $Na_2HPO_4$, pH 7.0). Native gel electrophoresis showed that cytochrome *c* stickiness (*i.e.* solvation), and charge decreased with each charge inverted mutation. The calculated p*I* of cytochrome *c* also decreased with each mutation (Table 2). Although it is generally accepted that protein solubility is lowest at its p*I* and tends to increase with increasing net charge[434] the quadruple cytochrome *c* mutant was preferentially hydrated (*i.e.* highly soluble) and unsticky in extracts at a pH close to its calculated p*I* (Table 2). Similarly, the triple cytochrome *c* mutant was charge neutral and NMR detectable in extracts. Thus, the replacement of cationic residues with glutamate changes the interaction propensity of cytochrome *c* such that interactions with water are favoured in the triple and quadruple mutants. This finding correlates with those of Kramer and co-workers who demonstrated that the solubility of seven proteins correlates weakly with net protein charge but strongly with the degree of negative charge.[434] No strong correlation was observed between protein solubility and the degree of positive protein charge, polarity or hydrophobicity.[434] Therefore, the kosmotropic aspartate and glutamate residues contribute most favourably to protein solubility, possibly by facilitating water clustering and structure formation (*i.e.* electrostriction) at protein surfaces.[247,435] Together, these findings suggest that anionic surface charges disfavour protein-macromolecule interactions in extracts through a combination of charge-charge repulsion with *E. coli* macroanions and preferential hydration.[63,225] This idea is supported by the results of crystal structure analyses indicating that interfacial water molecules frequently interact with the carboxylate groups of glutamate and aspartate side chains and that these interactions are enriched in low-specificity interfaces (*i.e.* crystal packing interfaces), which are the most strongly hydrated of all protein-macromolecule interfaces.[429] Importantly, crystallographic analyses also revealed that bridging water molecules (*i.e.* those mediating contacts between polar residues of protein partners) are a minority in most protein-protein complexes.[429] The role of carboxylate-interfacial water interactions in mediating low-specificity, quinary interactions is invoked. The cytochrome *c* and mutant studies also emphasise the role of side chain carboxylate groups in hydrating proteins and opposing aggregation. The abundance of anionic proteins in many proteomes[51,64] indicates that this is a global evolutionary trend that shapes cytoplasmic physicochemistry. Indeed, this trend is not unique to macromolecules. Prominent correlations have also been

reported between the measured concentration of metabolites and their water solubility such that the most abundant metabolites are highly water-soluble and often contain carboxylate or phosphate groups.[3,436]

**Table 2.** Cationic and anionic residues and calculated p*I*s for cytochrome *c* and R/K→E mutants

| Cytochrome *c* Variant | Calculated p*I* | # Cationic Residues | # Anionic Residues |
|---|---|---|---|
| WT | 9.5 | 23 | 11 |
| SM | 9.32 | 22 | 12 |
| DM | 9.04 | 21 | 13 |
| TM | 8.49 | 20 | 14 |
| QM | 6.84 | 19 | 15 |

## Implications for Cytoplasmic Structure

The physicochemical mechanisms that govern macromolecular assembly in cells are likely similar to those in extracts. It is therefore instructive to consider the test protein interaction propensities identified in cell extracts in light of Spitzer and Poolman's model for intracellular assembly.[63,65] This model emphasises the importance of screened electrostatic interactions and macromolecular crowding in driving the formation of cytoplasmic domains such as supercrowded 'cytogels' and the dilute 'cytosol'.[63,65] Cytogels are composed of functional macromolecular assemblies or 'clusters' that are expected to have a net negative charge owing to the preponderance of anionic macromolecules in *E. coli* (Introduction).[63–65] The inability to detect ΔTat-GB1, cytochrome *c* and flavodoxin, as well as most other globular proteins, by in-cell NMR supports the idea that cytoplasmic macromolecules interact extensively in the cytoplasm. SEC of concentrated *E. coli* extracts containing ΔTat-GB1 or cytochrome *c* confirmed the abundance of high molecular weight assemblies which were not disrupted at physiological concentrations of the abundant cellular ions $K^+$ and $Glu^-$. Native gel electrophoresis of *E. coli* extracts indicated that both cytochrome *c* and flavodoxin-containing assemblies were anionic in nature. This finding provides fresh evidence in support of abundant anionic clusters in native-like environments. Moreover, these anionic clusters pervaded at low and physiological ionic strengths suggesting the significance of screened charge-charge interactions in driving cluster formation in volume-occupied solutions. Indeed, the non-specific complexes between cationic proteins (*e.g.* ΔTat-GB1 or cytochrome *c*) and macroanions (*e.g.* RNA or proteins) resisted cell lysis and exposure to the SEC column. This finding suggests that low-specificity interactions, governed

predominantly by multiple charge-charge and/or salt bridge interactions, were notably strong (*i.e.* $K_d$ < 10 μM[109,437]). Similarly, the cytochrome *c*-flavodoxin interaction was apparently stabilised under volume-occupied conditions. The interactions of flavodoxin in *E. coli* extracts were relatively easy to disrupt *i.e.* by cell lysis and dilutions. This result suggests that flavodoxin's attractive quinary interactions involve weak, short range contacts (*e.g.* hydrogen bonds, dispersion forces) which diminish significantly with increasing intermolecular distances effected during dilution. Charge-charge interactions apply over comparably longer distances (Table 3)[307] and, thus, the increased NMR detectability of flavodoxin in dilute extracts suggests that charge-charge repulsion between flavodoxin and other macroanions predominate in the increasingly dilute extract samples. This finding supports Spitzer and Poolman's notion that anionic cytoplasmic surfaces remain 'watery' in the cytoplasm and indicates that, at increased levels of macromolecular crowding, favourable intermolecular contacts can overcome charge-charge repulsion to extend cytoplasmic clusters.

**Table 3.** Potential energies and their distance dependence

| Type | n[a] | Dependence[b] |
|---|---|---|
| Charge-charge | 1 | *I* |
| Charge-dipole | 2 | *I* and Φ |
| Dipole-dipole | 3 | *I* and Φ |
| Charge-induced dipole | 4 | Φ |
| Hydrogen bond | 4 | Φ |
| Dipole-induced dipole | 5 | Φ |
| Dispersion | 6 | Not D |
| van der Waals | 12 | Not D |

[a]Distance dependence of the potential energy which is proportional to $1/r^n$ where r is the distance between two proteins; [b] *I*, ionic strength–dependent; Θ, orientation-dependent; Not D, not dependent on solvent dielectric constant. Table derived from reference 23.

Spitzer and Poolman surmise that the supercrowded 'clusters' are bathed in dilute pools and interconnecting channels (*i.e.* the cytosol).[63,65] Pools and channel boundaries are defined by the charged, heterogeneous surfaces of the plasma membrane and macromolecular clusters.[63] Electrolytes (*e.g.* simple ions and metabolites) diffuse rapidly throughout the cytosol in response to electrochemical gradients arising between compositionally unique pools and cluster surface properties (*e.g.* charge and inter-cluster distance).[63] Not all intracellular ions are found in the cytosol, however. Ions that preferentially interact with macromolecules

are expected to reside within clusters. SEC studies of extracts containing ΔTat-GB1 or cytochrome *c* revealed differences between the shielding effects of biologically relevant ions which were consistent with their expected location in the cytoplasm. For instance, $Mg^{2+}$ is likely sequestered by nucleic acid phosphates or anionic protein surface patches. This model is supported by the shielding effect of $Mg^{2+}$ identified by the cell extract studies and the low concentration of free $Mg^{2+}$ (~ 1 mM) in *E. coli*.[7] Other ions (*e.g.* $K^+$, $Glu^-$, $Ac^-$, $HPO_4^{2-}$, and $Cl^-$) are expected to reside in electrolyte reservoirs and channels to assist with metabolite transfer by establishing electrochemical gradients.[63] The inability of these ions to disrupt ΔTat-GB1 interactions invokes their role in gradient formation. The TCA cycle intermediates $Succ^{2-}$ and $Cit^{3-}$ are expected to reside in the electrolyte system for transport to the TCA machinery. These anions disrupted some ΔTat-GB1 interactions. This may reflect the predicted role for metabolites in the regulation and rearrangement of the 'metabolon'.[438,439] Future work could explore the effects of other potential 'disruptors' such as polyamines (*e.g.* spermidine and putrescine) which are present in *E. coli* at concentrations up to 40 mM and interact predominantly with RNA,[440,441] a major component of quinary structure.[79]

# Bibliography

(1) Cronan, J. E. *Escherichia coli* as an Experimental Organism. In *eLS*; John Wiley & Sons, Ltd, **2001**.

(2) Sundararaj, S.; Guo, A.; Habibi-Nazhad, B.; Rouani, M.; Stothard, P.; Ellison, M.; Wishart, D. S. The CyberCell Database (CCDB): A Comprehensive, Self-Updating, Relational Database to Coordinate and Facilitate *in Silico* Modeling of *Escherichia coli*. *Nucleic Acids Res.* **2004**, *32*, D293–D295.

(3) Bennett, B. D.; Kimball, E. H.; Gao, M.; Osterhout, R.; Van Dien, S. J.; Rabinowitz, J. D. Absolute Metabolite Concentrations and Implied Enzyme Active Site Occupancy in *Escherichia coli*. *Nat. Chem. Biol.* **2009**, *5*, 593–599.

(4) Richey, B.; Cayley, D. S.; Mossing, M. C.; Kolka, C.; Anderson, C. F.; Farrar, T. C.; Record, M. T. J. Variability of the Intracellular Ionic Environment of *Escherichia coli*. Differences between *in Vitro* and *in Vivo* Effects of Ion Concentrations on Protein-DNA Interactions and Gene Expression. *J. Biol. Chem.* **1987**, *262*, 7157–7164.

(5) Theillet, F.-X.; Binolfi, A.; Frembgen-Kesner, T.; Hingorani, K.; Sarkar, M.; Kyne, C.; Li, C.; Crowley, P. B.; Gierasch, L.; Pielak, G. J.; Elcock, A. H.; Gershenson, A.; Selenko, P. Physicochemical Properties of Cells and their Effects on Intrinsically Disordered Proteins (IDPs). *Chem. Rev.* **2014**, *114*, 6661–6714.

(6) Moncany, M. L.; Kellenberger, E. High Magnesium Content of *Escherichia coli* B. *Experientia* **1981**, *37*, 846–847.

(7) Tyrrell, J.; McGinnis, J. L.; Weeks, K. M.; Pielak, G. J. The Cellular Environment Stabilizes Adenine Riboswitch RNA Structure. *Biochemistry* **2013**, *52*, 8777–8785.

(8) Outten, C. E.; O'Halloran, T. V. Femtomolar Sensitivity of Metalloregulatory Proteins Controlling Zinc Homeostasis. *Science* **2001**, *292*, 2488–2492.

(9) Finney, L. A.; O'Halloran, T. V. Transition Metal Speciation in the Cell: Insights from the Chemistry of Metal Ion Receptors. *Science* **2003**, *300*, 931–936.

(10) Kurihara, S.; Tsuboi, Y.; Oda, S.; Kim, H. G.; Kumagai, H.; Suzuki, H. The Putrescine Importer PuuP of *Escherichia coli* K-12 . *J. Bacteriol.* **2009**, *191*, 2776–2782.

(11) Wood, J. M. Osmosensing by Bacteria: Signals and Membrane-Based Sensors. *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 230–262.

(12) Navon, G.; Ogawa, S.; Shulman, R. G.; Yamane, T. High-Resolution $^{31}$P Nuclear Magnetic Resonance Studies of Metabolism in Aerobic *Escherichia coli* Cells. *Proc. Natl. Acad. Sci.* **1977**, *74*, 888–891.

(13) Asami, K.; Hanai, T.; Koizumi, N. Dielectric Analysis of *Escherichia coli* Suspensions in the Light of the Theory of Interfacial Polarization. *Biophys. J.* **1980**, *31*, 215–228.

(14) 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.

(15) Ellis, R. J. Macromolecular Crowding: Obvious but Underappreciated. *Trends Biochem. Sci.* **2001**, *26*, 597–604.

(16) Cayley, S.; Lewis, B. A.; Guttman, H. J.; Record, M. T. J. Characterization of the Cytoplasm of *Escherichia coli* K-12 as a Function of External Osmolarity. Implications for Protein-DNA Interactions *in Vivo*. *J. Mol. Biol.* **1991**, *222*, 281–300.

(17) Blattner, F. R.; Plunkett, G.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao Y. The Complete Genome Sequence of *Escherichia coli* K-12. *Science* **1997**, *277*, 1453–1462.

(18) Dorman, C. J. Genome Architecture and Global Gene Regulation in Bacteria: Making Progress towards a Unified Model? *Nat. Rev. Micro.* **2013**, *11*, 349–355.

(19) Cayley, S.; Record, M. T. J. Large Changes in Cytoplasmic Biopolymer Concentration with Osmolality Indicate That Macromolecular Crowding May Regulate Protein-DNA Interactions and Growth Rate in Osmotically Stressed *Escherichia coli* K-12. *J. Mol. Recognit.* **2004**, *17*, 488–496.

(20) Van den Bogaart, G.; Hermans, N.; Krasnikov, V.; Poolman, B. Protein Mobility and Diffusive Barriers in *Escherichia coli*: Consequences of Osmotic Stress. *Mol. Microbiol.* **2007**, *64*, 858–871.

(21) Mika, J. T.; van den Bogaart, G.; Veenhoff, L.; Krasnikov, V.; Poolman, B. Molecular Sieving Properties of the Cytoplasm of *Escherichia coli* and Consequences of Osmotic Stress. *Mol. Microbiol.* **2010**, *77*, 200–207.

(22) Walter, H.; Srere, P. A.; Brooks, D. E. *Microcompartmentation and Phase Separation in Cytoplasm*, 1st ed.; Academic Press, **1999**.

(23) Laue, T.; Demeler, B. A Postreductionist Framework for Protein Biochemistry. *Nat. Chem. Biol.* **2011**, *7*, 331–334.

(24) Molliex, A.; Temirov, J.; Lee, J.; Coughlin, M.; Kanagaraj, A. P.; Kim, H. J.; Mittag, T.; Taylor, J. P. Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **2015**, *163*, 123–133.

(25) Tolstoguzov, V. Compositions and Phase Diagrams for Aqueous Systems Based on Proteins and Polysaccharides. In *Microcompartmentation and Phase Separation in Cytoplasm*; Walter, H.; Brookes, D. E. Academic Press, **1999**.

(26) Walter, H.; Brooks, D. E. Phase Separation in Cytoplasm, due to Macromolecular Crowding, is the Basis for Microcompartmentation. *FEBS Lett.* **1995**, *361*, 135–139.

(27) Luisi, P. L. Toward the Engineering of Minimal Living Cells. *Anat. Rec.* **2002**, *268*, 208–214.

(28) Roodbeen, R.; van Hest, J. C. M. Synthetic Cells and Organelles: Compartmentalization Strategies. *Bioessays* **2009**, *31*, 1299–1308.

(29) Dzieciol, A. J.; Mann, S. Designs for Life: Protocell Models in the Laboratory. *Chem. Soc. Rev.* **2012**, *41*, 79–85.

(30) Miyata, H.; Hotani, H. Morphological Changes in Liposomes Caused by Polymerization of Encapsulated Actin and Spontaneous Formation of Actin Bundles. *Proc. Natl. Acad. Sci.* **1992**, *89*, 11547–11551.

(31) Cans, A.-S.; Wittenberg, N.; Karlsson, R.; Sombers, L.; Karlsson, M.; Orwar, O.; Ewing, A. Artificial Cells: Unique Insights into Exocytosis Using Liposomes and Lipid Nanotubes. *Proc. Natl. Acad. Sci.* **2003**, *100*, 400–404.

(32) Nomura, S. M.; Tsumoto, K.; Hamada, T.; Akiyoshi, K.; Nakatani, Y.; Yoshikawa, K. Gene Expression within Cell-Sized Lipid Vesicles. *ChemBioChem* **2003**, *4*, 1172–1175.

(33) Noireaux, V.; Libchaber, A. A Vesicle Bioreactor as a Step toward an Artificial Cell Assembly. *Proc. Natl. Acad. Sci.* **2004**, *101*, 17669–17674.

(34) Hanczyc, M. M.; Fujikawa, S. M.; Szostak, J. W. Experimental Models of Primitive Cellular Compartments: Encapsulation, Growth, and Division. *Science* **2003**, *302*, 618–622.

(35) Long, M. S.; Jones, C. D.; Helfrich, M. R.; Mangeney-Slavin, L. K.; Keating, C. D. Dynamic Microcompartmentation in Synthetic Cells. *Proc. Natl. Acad. Sci.* **2005**, *102*, 5920–5925.

(36) Marguet, M.; Sandre, O.; Lecommandoux, S. Polymersomes in “Gelly” Polymersomes: Toward Structural Cell Mimicry. *Langmuir* **2012**, *28*, 2035–2043.

(37) Peters, R. J. R. W.; Marguet, M.; Marais, S.; Fraaije, M. W.; van Hest, J. C. M.; Lecommandoux, S. Cascade Reactions in Multicompartmentalized Polymersomes. *Angew. Chemie Int. Ed.* **2014**, *53*, 146–150.

(38) Peters, R. J. R. W.; Nijemeisland, M.; van Hest, J. C. M. Reversibly Triggered Protein–Ligand Assemblies in Giant Vesicles. *Angew. Chemie Int. Ed.* **2015**, *54*, 9614–9617.

(39) Uversky, V. N.; Kuznetsova, I. M.; Turoverov, K. K.; Zaslavsky, B. Intrinsically Disordered Proteins as Crucial Constituents of Cellular Aqueous Two Phase Systems and Coacervates. *FEBS Lett.* **2015**, *589*, 15–22.

(40) Urry, D. W. Free Energy Transduction in Polypeptides and Proteins Based on Inverse Temperature Transitions. *Prog. Biophys. Mol. Biol.* **1992**, *57*, 23–57.

(41) Meyer, D. E.; Chilkoti, A. Quantification of the Effects of Chain Length and Concentration on the Thermal Behavior of Elastin-Like Polypeptides. *Biomacromolecules* **2004**, *5*, 846–851.

(42) Ge, X.; Conley, A. J.; Brandle, J. E.; Truant, R.; Filipe, C. D. M. *In Vivo* Formation of Protein Based Aqueous Microcompartments. *J. Am. Chem. Soc.* **2009**, *131*, 9094–9099.

(43) Montgomery, T. S. H. Comparative Cytological Studies, with Especial Regard to the Morphology of the Nucleolus. *J. Morphol.* **1898**, *15*, 265–582.

(44) Cajal, S. R. Un Sencillo Método de Coloración Selectiva Del Retículo Protoplásmico y Sus Efectos en Los Diversos Órganos Nerviosos de Vertebrados e Invertebrados. *Tra. Lab. Invest. Biol.* **1903**, *2*, 129–221.

(45) Wippich, F.; Bodenmiller, B.; Trajkovska, M. G.; Wanka, S.; Aebersold, R.; Pelkmans, L. Dual Specificity Kinase DYRK3 Couples Stress Granule Condensation/Dissolution to mTORC1 Signaling. *Cell* **2013**, *152*, 791–805.

(46) Brangwynne, C. P.; Eckmann, C. R.; Courson, D. S.; Rybarska, A.; Hoege, C.; Gharakhani, J.; Julicher, F.; Hyman, A. A. Germline P Granules are Liquid Droplets that Localize by Controlled Dissolution/Condensation. *Science* **2009**, *324*, 1729–1732.

(47) Brangwynne, C. P.; Mitchison, T. J.; Hyman, A. A. Active Liquid-like Behavior of Nucleoli Determines their Size and Shape in *Xenopus Laevis* Oocytes. *Proc. Natl. Acad. Sci.* **2011**, *108*, 4334–4339.

(48) Malinovska, L.; Kroschwald, S.; Alberti, S. Protein Disorder, Prion Propensities and Self-Organizing Macromolecular Collectives. *Biochim. Biophys. Acta* **2013**, *1834*, 918–931.

(49) Nott, T. J.; Petsalaki, E.; Farber, P.; Jervis, D.; Fussner, E.; Plochowietz, A.; Craggs, T. D.; Bazett-Jones, D. P.; Pawson, T.; Forman-Kay, J. D.; Baldwin, A. J. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell* **2015**, *57*, 936–947.

(50) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Why are "Natively Unfolded" Proteins Unstructured under Physiologic Conditions? *Proteins* **2000**, *41*, 415–427.

(51) Choura, M.; Rebaï, A. Exploring Charged Biased Regions in the Human Proteome. *Gene* **2013**, *515*, 277–280.

(52) Koga, S.; Williams, D. S.; Perriman, A. W.; Mann, S. Peptide–Nucleotide Microdroplets as a Step Towards a Membrane-Free Protocell Model. *Nat. Chem.* **2011**, *3*, 720–724.

(53) Crosby, J.; Treadwell, T.; Hammerton, M.; Vasilakis, K.; Crump, M. P.; Williams, D. S.; Mann, S. Stabilization and Enhanced Reactivity of Actinorhodin Polyketide Synthase Minimal Complex in Polymer-Nucleotide Coacervate Droplets. *Chem. Commun.* **2012**, *48*, 11832–11834.

(54) Kaibara, K.; Okazaki, T.; Bohidar, H. B.; Dubin, P. L. pH-Induced Coacervation in Complexes of Bovine Serum Albumin and Cationic Polyelectrolytes. *Biomacromolecules* **2000**, *1*, 100–107.

(55) Semenov, S. N.; Wong, A. S. Y.; van der Made, R. M.; Postma, S. G. J.; Groen, J.; van Roekel, H. W. H.; de Greef, T. F. A.; Huck, W. T. S. Rational Design of Functional and Tunable Oscillating Enzymatic Networks. *Nat. Chem.* **2015**, *7*, 160–165.

(56) Aumiller Jr, W. M.; Keating, C. D. Phosphorylation-Mediated RNA/Peptide Complex Coacervation as a Model for Intracellular Liquid Organelles. *Nat. Chem.* **2016**, *8*, 129–137.

(57) Rigaud, J.-L.; Pitard, B.; Levy, D. Reconstitution of Membrane Proteins into Liposomes: Application to Energy-Transducing Membrane Proteins. *Biochim. Biophys. Acta - Bioenerg.* **1995**, *1231*, 223–246.

(58) Van der Heide, T.; Stuart, M. C.; Poolman, B. On the Osmotic Signal and Osmosensing Mechanism of an ABC Transport System for Glycine Betaine. *EMBO J.* **2001**, *20*, 7022–7032.

(59) Karasawa, A.; Erkens, G. B.; Berntsson, R. P.-A.; Otten, R.; Schuurman-Wolters, G. K.; Mulder, F. A. A.; Poolman, B. Cystathionine Beta-Synthase (CBS) Domains 1 and 2 Fulfill Different Roles in Ionic Strength Sensing of the ATP-Binding Cassette (ABC) Transporter OpuA. *J. Biol. Chem.* **2011**, *286*, 37280–37291.

(60) Mahmood, N. A. B. N.; Biemans-Oldehinkel, E.; Patzlaff, J. S.; Schuurman-Wolters, G. K.; Poolman, B. Ion Specificity and Ionic Strength Dependence of the Osmoregulatory ABC Transporter OpuA. *J. Biol. Chem.* **2006**, *281*, 29830–29839.

(61) Poolman, B.; Spitzer, J. J.; Wood, J. M. Bacterial Osmosensing: Roles of Membrane Structure and Electrostatics in Lipid-Protein and Protein-Protein Interactions. *Biochim. Biophys. Acta - Biomembr.* **2004**, *1666*, 88–104.

(62) Spitzer, J. J. Maxwellian Double Layer Forces: From Infinity to Contact. *Langmuir* **2003**, *19*, 7099–7111.

(63) Spitzer, J. J.; Poolman, B. Electrochemical Structure of the Crowded Cytoplasm. *Trends Biochem. Sci.* **2005**, *30*, 536–541.

(64) 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.

(65) Spitzer, J.; Poolman, B. How Crowded Is the Prokaryotic Cytoplasm? *FEBS Lett.* **2013**, *587*, 2094–2098.

(66) Jones, L. S.; Yazzie, B.; Middaugh, C. R. Polyanions and the Proteome. *Mol. Cell. Proteomics* **2004**, *3*, 746–769.

(67) Weiller, G. F.; Caraux, G.; Sylvester, N. The Modal Distribution of Protein Isoelectric Points Reflects Amino Acid Properties Rather than Sequence Evolution. *Proteomics* **2004**, *4*, 943–949.

(68) Link, A. J.; Robison, K.; Church, G. M. Comparing the Predicted and Observed Properties of Proteins Encoded in the Genome of *Escherichia coli* K-12. *Electrophoresis* **1997**, *18*, 1259–1313.

(69) Lee, K. K.; Fitch, C. A.; Garcia-Moreno E, B. Distance Dependence and Salt Sensitivity of Pairwise, Coulombic Interactions in a Protein. *Protein Sci.* **2002**, *11*, 1004–1016.

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

(71) Serber, Z.; Keatinge-Clay, A. T.; Ledwidge, R.; Kelly, A. E.; Miller, S. M.; Dötsch, V. High-Resolution Macromolecular NMR Spectroscopy inside Living Cells. *J. Am. Chem. Soc.* **2001**, *123*, 2446–2447.

(72) Augustus, A. M.; Reardon, P. N.; Spicer, L. D. MetJ Repressor Interactions with DNA Probed by In-Cell NMR. *Proc. Natl. Acad. Sci.* **2009**, *106*, 5065–5069.

(73) 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.

(74) Crowley, P. B.; Chow, E.; Papkovskaia, T. Protein Interactions in the *Escherichia coli* Cytosol: An Impediment to In-Cell NMR Spectroscopy. *ChemBioChem* **2011**, *12*, 1043–1048.

(75) Li, C.; Charlton, L. M.; Lakkavaram, A.; Seagle, C.; Wang, G.; Young, G. B.; Macdonald, J. M.; Pielak, G. J. Differential Dynamical Effects of Macromolecular Crowding on an Intrinsically Disordered Protein and a Globular Protein: Implications for In-Cell NMR Spectroscopy. *J. Am. Chem. Soc.* **2008**, *130*, 6310–6311.

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

(77) Arnesano, F.; Banci, L.; Bertini, I.; Felli, I. C.; Losacco, M.; Natile, G. Probing the Interaction of Cisplatin with the Human Copper Chaperone Atox1 by Solution and In-Cell NMR Spectroscopy. *J. Am. Chem. Soc.* **2011**, *133*, 18361–18369.

(78) Barnes, C. O.; Monteith, W. B.; Pielak, G. J. Internal and Global Protein Motion Assessed with a Fusion Construct and In-Cell NMR Spectroscopy. *ChemBioChem* **2011**, *12* (3), 390–391.

(79) Majumder, S.; Xue, J.; DeMott, C. M.; Reverdatto, S.; Burz, D. S.; Shekhtman, A. Probing Protein Quinary Interactions by In-Cell Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **2015**, *54*, 2727–2738.

(80) Williamson, M. P. Progress in Nuclear Magnetic Resonance Spectroscopy Using Chemical Shift Perturbation to Characterise Ligand Binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16.

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

(82) Jones, S.; Daley, D. T. A.; Luscombe, N. M.; Berman, H. M.; Thornton, J. M. Protein-RNA Interactions: A Structural Analysis. *Nucleic Acids Res.* **2001**, *29*, 943–954.

(83) Janin, J.; Bahadur, R. P.; Chakrabarti, P. Protein-Protein Interactions and Quaternary Structure. *Q. Rev. Biophys.* **2008**, *41*, 133–180.

(84) Jones, S.; van Heyningen, P.; Berman, H. M.; Thornton, J. M. Protein-DNA Interactions: A Structural Analysis. *J. Mol. Biol.* **1999**, *287*, 877–896.

(85) Rudner, D. Z.; Pan, Q.; Losick, R. M. Evidence that Subcellular Localization of a Bacterial Membrane Protein is Achieved by Diffusion and Capture. *Proc. Natl. Acad. Sci.* **2002**, *99*, 8701–8706.

(86) Cayley, S.; Record, M. T. Roles of Cytoplasmic Osmolytes, Water, and Crowding in the Response of Escherichia Coli to Osmotic Stress: Biophysical Basis of Osmoprotection by Glycine Betaine. *Biochemistry* **2003**, *42* (43), 12596–12609.

(87) Cohen, R. D.; Guseman, A. J.; Pielak, G. J. Intracellular pH Modulates Quinary Structure. *Protein Sci.* **2015**.

(88) Alberts, B. The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists. *Cell* **1998**, *92*, 291–294.

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

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

(91) Srere, P. A. Macromolecular Interactions: Tracing the Roots. *Trends Biochem. Sci.* **2000**, *25*, 150–153.

(92) Gronenborn, A. M.; Clore, G. M. Identification of the Contact Surface of a *Streptococcal* Protein G Domain Complexed with a Human Fc Fragment. *J. Mol. Biol.* **1993**, *233*, 331–335.

(93) Gronenborn, A. M.; Frank, M. K.; Clore, G. M. Core Mutants of the Immunoglobulin Binding Domain of *Streptococcal* Protein G: Stability and Structural Integrity. *FEBS Lett.* **1996**, *398*, 312–316.

(94) Malakauskas, S. M.; Mayo, S. L. Design, Structure and Stability of a Hyperthermophilic Protein Variant. *Nat. Struct. Biol.* **1998**, *5*, 470–475.

(95) Klemba, M.; Gardner, K. H.; Marino, S.; Clarke, N. D.; Regan, L. Novel Metal-Binding Proteins by Design. *Nat. Struct. Mol. Biol.* **1995**, *2*, 368–373.

(96) Lindman, S.; Xue, W.-F.; Szczepankiewicz, O.; Bauer, M. C.; Nilsson, H.; Linse, S. Salting the Charged Surface: pH and Salt Dependence of Protein G B1 Stability. *Biophys. J.* **2006**, *90*, 2911–2921.

(97) Zhou, P.; Wagner, G. Overcoming the Solubility Limit with Solubility-Enhancement Tags: Successful Applications in Biomolecular NMR Studies. *J. Biomol. NMR* **2010**, *46*, 23–31.

(98) Gallagher, T.; Alexander, P.; Bryan, P.; Gilliland, G. L. Two Crystal Structures of the B1 Immunoglobulin-Binding Domain of *Streptococcal* Protein G and Comparison with NMR. *Biochemistry* **1994**, *33* (15), 4721–4729.

(99) 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.

(100) Kyne, C.; Ruhle, B.; Gautier, V. W.; Crowley, P. B. Specific Ion Effects on Macromolecular Interactions in *Escherichia coli* Extracts. *Protein Sci.* **2015**, *24*, 310–318.

(101) Narita, K.; Titani, K. The Complete Amino Acid Sequence in Baker's Yeast Cytochrome *c*. *J. Biochem.* **1969**, *65*, 259–267.

(102) Saraste, M. Oxidative Phosphorylation at the Fin de Siecle. *Science* **1999**, *283*, 1488–1493.

(103) Kagan, V. E.; Tyurin, V. A.; Jiang, J.; Tyurina, Y. Y.; Ritov, V. B.; Amoscato, A. A.; Osipov, A. N.; Belikova, N. A.; Kapralov, A. A.; Kini, V.; Vlasova, I. I.; Zhao, Q.; Zou, M.; Di, P.; Svistunenko, D.; Kurnikov, I. V.; Borisenko, G. G. Cytochrome *c* Acts as a Cardiolipin Oxygenase Required for Release of Proapoptotic Factors. *Nat. Chem. Biol.* **2005**, *1*, 223–232.

(104) Louie, G. V; Brayer, G. D. High-Resolution Refinement of Yeast Iso-1-Cytochrome *c* and Comparisons with Other Eukaryotic Cytochromes *c*. *J. Mol. Biol.* **1990**, *214*, 527–555.

(105) Park, K.; Frost, B.; Lee, H.; Kim, S.; Paik, W. Effect of Enzymatic Methylation of Proteins on their Isoelectric Points. *Arch. Pharm. Res.* **1989**, *12*, 79–87.

(106) McGovern, R. E.; Fernandes, H.; Khan, A. R.; Power, N. P.; Crowley, P. B. Protein Camouflage in Cytochrome *c*-Calixarene Complexes. *Nat. Chem.* **2012**, *4*, 527–533.

(107) Clark-Ferris, K. K.; Fisher, J. Topographical Mimicry of the Enzyme Binding Domain of Cytochrome *c*. *J. Am. Chem. Soc.* **1985**, *107*, 5007–5008.

(108) Crowley, P. B.; Ganji, P.; Ibrahim, H. Protein Surface Recognition: Structural Characterisation of Cytochrome *c*-Porphyrin Complexes. *ChemBioChem* **2008**, *9*, 1029–1033.

(109) Wettstein, C.; Kyne, C.; M. Doolan, A.; Mohwald, H.; Crowley, P. B.; Lisdat, F. Study of Cytochrome *c*-DNA Interaction-Evaluation of Binding Sites on the Redox Protein. *Nanoscale* **2014**, *6*, 13779–13786.

(110) Volkov, A. N.; van Nuland, N. a J. Electron Transfer Interactome of Cytochrome *c*. *PLoS Comput. Biol.* **2012**, *8*, e1002807.

(111) Boswell, A. P.; McClune, G. J.; Moore, G. R.; Williams, R. J.; Pettigrew, G. W.; Inubishi, T.; Yonetani, T.; Harris, D. E. Nuclear Magnetic Resonance Study of the Interaction of Cytochrome *c* with Cytochrome *c* Peroxidase. *Biochim. Biophys. Acta - Bioenerg.* **1973**, *292*, 502–508.

(112) Hazzard, J. T.; Tollin, G. Proton NMR Study of the Cytochrome *c*:Flavodoxin Electron Transfer Complex. *Biochem. Biophys. Res. Commun.* **1985**, *130*, 1281–1286.

(113) Worrall, J. A.; Kolczak, U.; Canters, G. W.; Ubbink, M. Interaction of Yeast Iso-1-Cytochrome *c* with Cytochrome *c* Peroxidase Investigated by [$^{15}$N, $^{1}$H] Heteronuclear NMR Spectroscopy. *Biochemistry* **2001**, *40*, 7069–7076.

(114) Van de Water, K.; Sterckx, Y. G. J.; Volkov, A. N. The Low-Affinity Complex of Cytochrome *c* and its Peroxidase. *Nat. Commun.* **2015**, *6*. 7073–7084.

(115) Dronov, R.; Kurth, D. G.; Mohwald, H.; Scheller, F. W.; Lisdat, F. Communication in a Protein Stack: Electron Transfer Between Cytochrome *c* and Bilirubin Oxidase within a Polyelectrolyte Multilayer. *Angew. Chem. Int. Ed.* **2008**, *47*, 3000–3003.

(116) Beissenhirtz, M. K.; Scheller, F. W.; Lisdat, F. A Superoxide Sensor Based on a Multilayer Cytochrome *c* Electrode. *Anal. Chem.* **2004**, *76*, 4665–4671.

(117) McGovern, R. E.; Feifel, S. C.; Lisdat, F.; Crowley, P. B. Microscale Crystals of Cytochrome *c* and Calixarene on Electrodes: Interprotein Electron Transfer Between Defined Sites. *Angew. Chemie Int. Ed.* **2015**, *54*, 6356–6359.

(118) Hoover, D. M.; Ludwig, M. L. A Flavodoxin that is Required for Enzyme Activation: The Structure of Oxidized Flavodoxin from *Escherichia coli* at 1.8 Å Resolution. *Protein Sci.* **1997**, *6*, 2525–2537.

(119) Ponstingl, H.; Otting, G. NMR Assignments, Secondary Structure and Hydration of Oxidized *Escherichia coli* Flavodoxin. *Eur. J. Biochem.* **1997**, *244*, 384–399.

(120) McIver, L.; Leadbeater, C.; Campopiano, D. J.; Baxter, R. L.; Daff, S. N.; Chapman, S. K.; Munro, A. W. Characterisation of Flavodoxin NADP+ Oxidoreductase and Flavodoxin; Key Components of Electron Transfer in *Escherichia coli. Eur. J. Biochem.* **1998**, *257*, 577–585.

(121) Hall, D. A.; Vander Kooi, C. W.; Stasik, C. N.; Stevens, S. Y.; Zuiderweg, E. R. P.; Matthews, R. G. Mapping the Interactions Between Flavodoxin and its Physiological Partners Flavodoxin Reductase and Cobalamin-Dependent Methionine Synthase. *Proc. Natl. Acad. Sci.* **2001**, *98*, 9521–9526.

(122) Genzor, C. G.; Beldarrain, A.; Gomez-Moreno, C.; Lopez-Lacomba, J. L.; Cortijo, M.; Sancho, J. Conformational Stability of Apoflavodoxin. *Protein Sci.* **1996**, *5*, 1376–1388.

(123) Van Mierlo, C. P.; van den Oever, J. M.; Steensma, E. Apoflavodoxin (un)folding Followed at the Residue Level by NMR. *Protein Sci.* **2000**, *9*, 145–157.

(124) Cremades, N.; Sancho, J. Molten Globule and Native State Ensemble of *Helicobacter Pylori* Flavodoxin: Can Crowding, Osmolytes or Cofactors Stabilize the Native Conformation Relative to the Molten Globule? *Biophys. J.* **2008**, *95*, 1913–1927.

(125) Lostao, A.; Daoudi, F.; Irún, M. P.; Ramón, Á.; Fernández-Cabrera, C.; Romero, A.; Sancho, J. How FMN Binds to Anabaena Apoflavodoxin: A Hydrophobic Encounter at an Open Binding Site . *J. Biol. Chem.* **2003**, *278*, 24053–24061.

(126) Tollin, G.; Edmondson, D. E. Flavoprotein Chemistry. II. Chemical and Physical Characterization of the Shethna Flavoprotein and Apoprotein and Kinetics and Thermodynamics of Flavine Analog Binding to the Apoprotein. *Biochemistry* **1971**, *10*, 124–132.

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

(128) Selenko, P.; Frueh, D. P.; Elsaesser, S. J.; Haas, W.; Gygi, S. P.; Wagner, G. *In Situ* Observation of Protein Phosphorylation by High-Resolution NMR Spectroscopy. *Nat. Struct. Mol. Biol.* **2008**, *15*, 321–329.

(129) Laue, T. M.; Shepard, H. K.; Ridgeway, T. M.; Moody, T. P.; Wilson, T. J. Membrane-Confined Analytical Electrophoresis. *Methods Enzymol.* **1998**, *295*, 494–518.

(130) Li, C.; Wang, G.-F.; Wang, Y.; Creager-Allen, R.; Lutz, E. A.; Scronce, H.; Slade, K. M.; Ruf, R. A. S.; Mehl, R. A.; Pielak, G. J. Protein $^{19}$F NMR in *Escherichia coli*. *J. Am. Chem. Soc.* **2010**, *132* (1), 321.

(131) Ye, Y.; Liu, X.; Zhang, Z.; Wu, Q.; Jiang, B.; Jiang, L.; Zhang, X.; Liu, M.; Pielak, G. J.; Li, C. $^{19}$F NMR Spectroscopy as a Probe of Cytoplasmic Viscosity and Weak Protein Interactions in Living Cells. *Chem. – A Eur. J.* **2013**, *19*, 12705–12710.

(132) Troland, L. T. Biological Enigmas and the Theory of Enzyme Action. *Am. Nat.* **1917**, *51*, 321–350.

(133) Bedau, M. A.; Cleland, C. E. *The Nature of Life: Classical and Contemporary Perspectives from Philosophy and Science*; Cambridge University Press: New York, **2010**.

(134) Piccolino, M. Marcello Malpighi and the Difficult Birth of Modern Life Sciences. *Endeavour* **1999**, *23*, 175–179.

(135) Muller, J. *Elements of Physiology*; Taylor and Walton: London, **1842**.

(136) Lipman, T. O. Wohler's Preparation of Urea and the Fate of Vitalism. *J. Chem. Educ.* **1964**, *41*, 452–458.

(137) Ramberg, P. J. The Death of Vitalism and the Birth of Organic Chemistry: Wohler's Urea Synthesis and the Disciplinary Identity of Organic Chemistry. *Ambix* **2000**, *47*, 170–195.

(138) Rocke, A. J. *The Quiet Revolution Hermann Kolbe and the Science of Organic Chemistry*; University of California Press: Berkeley, **1993**.

(139) Nurse, P. The Great Ideas of Biology. *Clin. Med.* **2003**, *3*, 560–568.

(140) Mazzarello, P. A Unifying Concept: The History of Cell Theory. *Nat. Cell Biol.* **1999**, *1*, E13–E15.

(141) Goodale, G. L. Protoplasm and its History. *Bot. Gaz.* **1889**, *14*, 235–240.

(142) Welch, G. R.; Clegg, J. S. Cell versus Protoplasm: Revisionist History. *Cell Biol. Int.* **2012**, *36*, 643–647.

(143) Loeb, J. *Proteins and the Theory of Colloidal Behaviour*; McGraw-Hill Book Company: New York, **1922**.

(144) Tanford, C.; Reynolds, J. *Nature's Robots*; Oxford University Press: Oxford, **2001**.

(145) Allen, G. E. Mechanism, Vitalism and Organicism in Late Nineteenth and Twentieth-Century Biology: The Importance of Historical Context. *Stud. Hist. Philos. Biol. Biomed. Sci.* **2005**, *36*, 261–283.

(146) Loeb, J. *The Mechanistic Conception of Life*; University of Chicago Press: Chicago, **1912**.

(147) Czapek, F. Recent Investigations on the Protoplasm of Plant Cells and its Colloidal Properties. *Ann. Missouri Bot. Gard.* **1915**, *2*, 241–252.

(148) Vale, R. D. The Molecular Motor Toolbox for Intracellular Transport. *Cell* **2003**, *112*, 467–480.

(149) Kohler, R. The Reception of Eduard Buchner's Discovery of Cell-Free Fermentation. *J. Hist. Biol.* **1972**, *5*, 327–353.

(150) Loeb, J.; Chamberlain, M. M. An Attempt at a Physico-Chemical Explanation of Certain Groups of Fluctuating Variation. *J. Exp. Zool.* **1915**, *19*, 559–568.

(151) Michaelis, L.; Menten, M. L.; Johnson, K. A.; Goody, R. S. The Original Michaelis Constant: Translation of the 1913 Michaelis-Menten Paper. *Biochemistry* **2011**, *50*, 8264–8269.

(152) Ritter, W. E. *The Unity of the Organism; Or, The Organismal Conception of Life*; Badger, R. G. : Boston, **1919**.

(153) Peters, R. A. Surface Structure in the Integration of Cell Activity. *Trans Faraday Soc* **1930**, *26*, 797–807.

(154) Krebs Henseleit, K, H. A. Untersuchungen Über Die Harnsstoffbildung in Tierkörper. *Z. Physiol. Chem.* **1932**, *210*, 33–66.

(155) Kopac, M. J. The Deveaux Effect at Oil–protoplasm Interfaces. *Biol. Bull.* **1938**, *75*, 372–383.

(156) Cori, C. F.; Schmidt, G.; Cori, G. T. The Synthesis of a Polysaccharide from Glucose-1-Phosphate in Muscle Extract. *Science* **1939**, *89*, 464–465.

(157) Zalokar, M. Cytochemistry of Centrifuged Hyphae of *Neurospora*. *Exp. Cell Res.* **1960**, *19*, 114–132.

(158) Kempner, E. S.; Miller, J. H. The Molecular Biology of *Euglena Gracilis*. V. Enzyme Localization. *Exp. Cell Res.* **1968**, *51*, 150–156.

(159) Yanofsky, C.; Rachmeler, M. The Exclusion of Free Indole as an Intermediate in the Biosynthesis of Tryptophan in *Neurospora Crassa*. *Biochim. Biophys. Acta* **1958**, *28*, 640–641.

(160) Srere, P. A. Enzyme Concentrations in Tissues. *Science* **1967**, *158*, 936–937.

(161) Srere, P. A. An Eclectic View of Metabolic Regulation: Control of Citrate Synthase Activity. *Adv. Enzyme Regul.* **1971**, *9*, 221–233.

(162) Srere, P. A. *Is There an Organization of Krebs Cycle Enzymes in the Mitochondrial Matrix?* Hanson, M. A. M. Academic Press, **1972**.

(163) Srere, P. A.; Mattiasson, B.; Mosbach, K. An Immobilized Three Enzyme System: A Model for Microenvironmental Compartmentation in Mitochondria. *Proc. Natl. Acad. Sci.* **1973**, *70*, 2534–2538.

(164) Alberts, B.; Miake-Lye, R. Unscrambling the Puzzle of Biological Machines: The Importance of the Details. *Cell* **1992**, *68*, 415–420.

(165) Osborne, T. B. Crystallized Vegetable Proteids. *Am. Chem. J.* **1892**, *14*, 662–689.

(166) Giege, R. A Historical Perspective on Protein Crystallization from 1840 to the Present Day. *FEBS J.* **2013**, *280*, 6456–6497.

(167) Sumner, J. B. The Isolation and Crystallization of the Enzyme Urease: Preliminary Paper *J. Biol. Chem.* **1926**, *69*, 435–441.

(168) Bernal, J. D.; Crowfoot, D. X-Ray Photographs of Crystalline Pepsin. *Nature* **1934**, *133*, 794–795.

(169) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **1958**, *181*, 662–666.

(170) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(171) Thornton, J. M.; Singh, J.; Campbell, S.; Blundell, T. O. M. L. Protein-Protein Recognition via Side-Chain Interactions. *Biochem. Soc. Trans.* **1988**, *16*, 927–930.

(172) Caetano-Anolles, G.; Wang, M.; Caetano-Anolles, D.; Mittenthal, J. E. The Origin, Evolution and Structure of the Protein World. *Biochem. J.* **2009**, *417*, 621–637.

(173) Nooren, I. M.; Thornton, J. M. Diversity of Protein-Protein Interactions. *EMBO J.* **2003**, *22*, 3486–3492.

(174) Fersht, A. R. From the First Protein Structures to Our Current Knowledge of Protein Folding: Delights and Scepticisms. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 650–654.

(175) Englander, S. W.; Mayne, L.; Bai, Y.; Sosnick, T. R. Hydrogen Exchange: The Modern Legacy of Linderstrom-Lang. *Protein Sci.* **1997**, *6*, 1101–1109.

(176) Perutz, M. F.; Muirhead, H.; Cox, J. M.; Goaman, L. C. Three-Dimensional Fourier Synthesis of Horse Oxyhaemoglobin at 2.8 Å Resolution: The Atomic Model. *Nature* **1968**, *219*, 131–139.

(177) Wüthrich, K. NMR Studies of Structure and Function of Biological Macromolecules (Nobel Lecture). *Angew. Chemie Int. Ed.* **2003**, *42*, 3340–3363.

(178) Otting, G.; Qian, Y. Q.; Billeter, M.; Muller, M.; Affolter, M.; Gehring, W. J.; Wuthrich, K. Protein-DNA Contacts in the Structure of a Homeodomain-DNA Complex Determined by Nuclear Magnetic Resonance Spectroscopy in Solution. *EMBO J.* **1990**, *9*, 3085–3092.

(179) Chouard, T. Structural Biology: Breaking the Protein Rules. *Nature* **2011**, 471,151–153.

(180) Uversky, V. N. The Multifaceted Roles of Intrinsic Disorder in Protein Complexes. *FEBS Lett.* **2015**, *589*, 2498–2506.

(181) Alberts, B. M. The DNA Enzymology of Protein Machines. *Cold Spring Harb. Symp. Quant. Biol.* **1984**, *49*, 1–12.

(182) Gavin, A.-C.; Aloy, P.; Grandi, P.; Krause, R.; Boesche, M.; Marzioch, M.; Rau, C.; Jensen, L. J.; Bastuck, S.; Dumpelfeld, B.; Edelmann, A.; Heurtier, M.; Hoffman, V; Hoefert, C.; Klein, K.; Hudak, M.; Michon, M.; Schelder, M.; Schirle, M.; Remor, M.; Rudi, T.; Hooper, S.; Bauer, A.; Bouwmeester, T.; Casari, G.; Drewes, G.; Neubauer, G.; Rick, J. M.; Kuster, G; Bork, P.; Russell, R. B.; Superti-Furga, G. Proteome Survey Reveals Modularity of the Yeast Cell Machinery. *Nature* **2006**, *440*, 631–636.

(183) Andres, C.; Agne, B.; Kessler, F. Preparation of Multiprotein Complexes from Arabidopsis Chloroplasts Using Tandem Affinity Purification. *Methods Mol. Biol.* **2011**, *775*, 31–49.

(184) Xu, G.; Ye, Y.; Liu, X.; Cao, S.; Wu, Q.; Cheng, K.; Liu, M.; Pielak, G. J.; Li, C. Strategies for Protein NMR in *Escherichia coli*. *Biochemistry* **2014**, *53*, 1971–1981.

(185) Kroe, R. R.; Laue, T. M. NUTS and BOLTS: Applications of Fluorescence-Detected Sedimentation. *Anal. Biochem.* **2009**, *390*, 1–13.

(186) Dedmon, M. M.; Patel, C. N.; Young, G. B.; Pielak, G. J. FlgM Gains Structure in Living Cells. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12681–12684.

(187) An, S.; Kumar, R.; Sheets, E. D.; Benkovic, S. J. Reversible Compartmentalization of *de Novo* Purine Biosynthetic Complexes in Living Cells. *Science* **2008**, *320*, 103–106.

(188) Lander, G. C.; Saibil, H. R.; Nogales, E. Go Hybrid: EM, Crystallography, and Beyond. *Curr. Opin. Struct. Biol.* **2012**, *22*, 627–635.

(189) Blanchet, C. E.; Svergun, D. I. Small-Angle X-Ray Scattering on Biological Macromolecules and Nanocomposites in Solution. *Annu. Rev. Phys. Chem.* **2013**, *64*, 37–54.

(190) Torres, T.; Levitus, M. Measuring Conformational Dynamics: A New FCS-FRET Approach. *J. Phys. Chem. B* **2007**, *111*, 7392–7400.

(191) Larson, J. D.; Rodgers, M. L.; Hoskins, A. A. Visualizing Cellular Machines with Colocalization Single Molecule Microscopy. *Chem. Soc. Rev.* **2014**, *43*, 1189–1200.

(192) Nickell, S.; Kofler, C.; Leis, A. P.; Baumeister, W. A Visual Approach to Proteomics. *Nat Rev Mol Cell Biol* **2006**, *7*, 225–230.

(193) Von Hippel, P. H.; Delagoutte, E. A General Model for Nucleic Acid Helicases and their "Coupling" within Macromolecular Machines. *Cell* **2001**, *104*, 177–190.

(194) Sali, A.; Berman, H. M.; Schwede, T.; Trewhella, J.; Kleywegt, G.; Burley, S. K.; Markley, J.; Nakamura, H.; Adams, P.; Bonvin, A. M.; Chiu. W.; Peraro, M. D.; Di Maio, F.; Ferrin T. E.; Grünewald, K.; Gutmanas, A.; Henderson, R.; Hummer, G.; Iwasaki, K.; Johnson, G.; Lawson, C. L.; Meiler, J.; Marti-Renom, M. A.; Montelione, G. T.; Nilges, M.; Nussinov, R.; Patwardhan, A.; Rappsilber, J.; Read, R. J.; Saibil, H.; Schröder, G. F.; Schwieters, C. D.; Seidel, C. A.; Svergun, D.; Topf, M.; Ulrich, E. L.; Velankar, S.; Westbrook, J. D. Outcome of the First wwPDB Hybrid/Integrative Methods Task Force Workshop. *Structure* **2015**, *23*, 1156–1167.

(195) Politis, A.; Borysik, A. J. Assembling the Pieces of Macromolecular Complexes: Hybrid Structural Biology Approaches. *Proteomics* **2015**, *15*, 2792–2803.

(196) Norris, V.; den Blaauwen, T.; Doi, R. H.; Harshey, R. M.; Janniere, L.; Jimenez-Sanchez, A.; Jin, D. J.; Levin, P. A.; Mileykovskaya, E.; Minsky, A.; Misevic, G.; Ripoll, C.; Saier, M. Jr.; Skarstad, K. Toward a Hyperstructure Taxonomy. *Annu. Rev. Microbiol.* **2007**, *61*, 309–329.

(197) Sear, R. P.; Pagonabarraga, I.; Flaus, A. Life at the Mesoscale: The Self-Organised Cytoplasm and Nucleoplasm. *BMC Biophys.* **2015**, *8*, 4.

(198) 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.* **2015**, *112*, 1739–1742.

(199) Hoenger, A. High-Resolution Cryo-Electron Microscopy on Macromolecular Complexes and Cell Organelles. *Protoplasma* **2014**, *251*, 417–427.

(200) Lucic, V.; Forster, F.; Baumeister, W. Structural Studies by Electron Tomography: From Cells to Molecules. *Annu. Rev. Biochem.* **2005**, *74*, 833–865.

(201) Irobalieva, R. N.; Martins, B.; Medalia, O. Cellular Structural Biology as Revealed by Cryo-Electron Tomography. *J. Cell Sci.* **2016**, *129*, 469–476.

(202) Asano, S.; Fukuda, Y.; Beck, F.; Aufderheide, A.; Förster, F.; Danev, R.; Baumeister, W. A Molecular Census of 26S Proteasomes in Intact Neurons. *Science* **2015**, *347*, 439–442.

(203) Lin, J.; Yin, W.; Smith, M. C.; Song, K.; Leigh, M. W.; Zariwala, M. A.; Knowles, M. R.; Ostrowski, L. E.; Nicastro, D. Cryo-Electron Tomography Reveals Ciliary Defects Underlying Human RSPH1 Primary Ciliary Dyskinesia. *Nat. Commun.* **2014**, *5*, 5727-5746.

(204) McGuffee, S. R.; Elcock, A. H. Diffusion, Crowding & Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm. *PLoS Comput Biol* **2010**, *6*, 1000694.

(205) Hasnain, S.; McClendon, C. L.; Hsu, M. T.; Jacobson, M. P.; Bandyopadhyay, P. A New Coarse-Grained Model for *E. coli* Cytoplasm: Accurate Calculation of the Diffusion Coefficient of Proteins and Observation of Anomalous Diffusion. *PLoS One* **2014**, *9*, e106466.

(206) 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.

(207) Bucci, M.; Goodman, C.; Sheppard, T. L. A Decade of Chemical Biology. *Nat. Chem. Biol.* **2010**, *6*, 847–854.

(208) Novina, C. D.; Sharp, P. A. The RNAi Revolution. *Nature* **2004**, *430*, 161–164.

(209) Lautru, S.; Deeth, R. J.; Bailey, L. M.; Challis, G. L. Discovery of a New Peptide Natural Product by *Streptomyces Coelicolor* Genome Mining. *Nat Chem. Biol.* **2005**, *1*, 265–269.

(210) McKay, C. S.; Finn, M. G. Click Chemistry in Complex Mixtures: Bioorthogonal Bioconjugation. *Chem. Biol.* **2015**, *21*, 1075–1101.

(211) Hideshima, T.; Bradner, J. E.; Wong, J.; Chauhan, D.; Richardson, P.; Schreiber, S. L.; Anderson, K. C. Small-Molecule Inhibition of Proteasome and Aggresome Function Induces Synergistic Antitumor Activity in Multiple Myeloma. *Proc. Natl. Acad. Sci.* **2005**, *102*, 8567–8572.

(212) Haggarty, S. J.; Koeller, K. M.; Wong, J. C.; Grozinger, C. M.; Schreiber, S. L. Domain-Selective Small-Molecule Inhibitor of Histone Deacetylase 6 (HDAC6)-Mediated Tubulin Deacetylation. *Proc. Natl. Acad. Sci.* **2003**, *100*, 4389–4394.

(213) Dandapani, S.; Marcaurelle, L. A. Grand Challenge Commentary: Accessing New Chemical Space for "Undruggable" Targets. *Nat. Chem. Biol.* **2010**, *6*, 861–863.

(214) Chinai, J. M.; Taylor, A. B.; Ryno, L. M.; Hargreaves, N. D.; Morris, C. A.; Hart, P. J.; Urbach, A. R. Molecular Recognition of Insulin by a Synthetic Receptor. *J. Am. Chem. Soc.* **2011**, *133*, 8810–8813.

(215) Bier, D.; Rose, R.; Bravo-Rodriguez, K.; Bartel, M.; Ramirez-Anguita, J. M.; Dutt, S.; Wilch, C.; Klärner, F. G.; Sanchez-Garcia, E.; Schrader, T.; Ottmann, C. Molecular Tweezers Modulate 14-3-3 Protein–protein Interactions. *Nat. Chem.* **2013**, *5*, 234–239.

(216) Mattia, E.; Otto, S. Supramolecular Systems Chemistry. *Nat. Nano.* **2015**, *10*, 111–119.

(217) Boekhoven, J.; Brizard, A. M.; Kowlgi, K. N. K.; Koper, G. J. M.; Eelkema, R.; van Esch, J. H. Dissipative Self-Assembly of a Molecular Gelator by Using a Chemical Fuel. *Angew. Chemie Int. Ed.* **2010**, *49*, 4825–4828.

(218) Sprung, R.; Chen, Y.; Zhang, K.; Cheng, D.; Zhang, T.; Peng, J.; Zhao, Y. Identification and Validation of Eukaryotic Aspartate and Glutamate Methylation in Proteins. *J. Proteome Res.* **2008**, *7*, 1001–1006.

(219) Zhou, H.-X.; Rivas, G.; Minton, A. P. Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences. *Annu. Rev. Biophys.* **2008**, *37*, 375–397.

(220) Sarkar, M.; Li, C.; Pielak, G. J. Soft Interactions and Crowding. *Biophys. Rev.* **2013**, *5*, 187–194.

(221) Phillip, Y.; Schreiber, G. Formation of Protein Complexes in Crowded Environments – From *In Vitro* to *In Vivo*. *FEBS Lett.* **2013**, *587*, 1046–1052.

(222) Groen, J.; Foschepoth, D.; te Brinke, E.; Boersma, A. J.; Imamura, H.; Rivas, G.; Heus, H. A.; Huck, W. T. S. Associative Interactions in Crowded Solutions of Biopolymers Counteract Depletion Effects. *J. Am. Chem. Soc.* **2015**, *137*, 13041–13048.

(223) Keating, C. D. Aqueous Phase Separation as a Possible Route to Compartmentalization of Biological Molecules. *Acc. Chem. Res.* **2012**, *45*, 2114–2124.

(224) TangT-Y., D.; HakC., R. C.; Thompson, A. J.; Kuimova, M. K.; S., W.; Perriman, A. W.; Mann, S. Fatty Acid Membrane Assembly on Coacervate Microdroplets as a Step towards a Hybrid Protocell Model. *Nat. Chem.* **2014**, *6*, 527–533.

(225) Spitzer, J. From Water and Ions to Crowded Biomacromolecules: *In Vivo* Structuring of a Prokaryotic Cell. *Microbiol. Mol. Biol. Rev.* **2011**, *75*, 491–506.

(226) Serber, Z.; Dotsch, V. In-Cell NMR Spectroscopy. *Biochemistry* **2001**, *40*, 14317–14323.

(227) Serber, Z.; Straub, W.; Corsini, L.; Nomura, A. M.; Shimba, N.; Craik, C. S.; Ortiz de Montellano, P.; Dotsch, V. Methyl Groups as Probes for Proteins and Complexes in In-Cell NMR Experiments. *J. Am. Chem. Soc.* **2004**, *126*, 7119–7125.

(228) Reckel, S.; Lohr, F.; Dotsch, V. In-Cell NMR Spectroscopy. *ChemBioChem* **2005**, *6*, 1601–1606.

(229) Reardon, P. N.; Spicer, L. D. Multidimensional NMR Spectroscopy for Protein Characterization and Assignment Inside Cells. *J. Am. Chem. Soc.* **2005**, *127*, 10848–10849.

(230) Serber, Z.; Selenko, P.; Hansel, R.; Reckel, S.; Lohr, F.; Ferrell, J. E.; Wagner, G.; Dotsch, V. Investigating Macromolecules Inside Cultured and Injected Cells by In-Cell NMR Spectroscopy. *Nat. Protoc.* **2007**, *1*, 2701–2709.

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

(232) Pielak, G. J.; Li, C.; Miklos, A. C.; Schlesinger, A. P.; Slade, K. M.; Wang, G.-F.; Zigoneanu, I. G. Protein Nuclear Magnetic Resonance Under Physiological Conditions. *Biochemistry* **2009**, *48*, 226–234.

(233) Luh, L. M.; Hänsel, R.; Löhr, F.; Kirchner, D. K.; Krauskopf, K.; Pitzius, S.; Schäfer, B.; Tufar, P.; Corbeski, I.; Güntert, P.; Dötsch, V. Molecular Crowding Drives Active Pin1 into Nonspecific Complexes with Endogenous Proteins Prior to Substrate Recognition. *J. Am. Chem. Soc.* **2013**, *135*, 13796–13803.

(234) Amata, I.; Maffei, M.; Igea, A.; Gay, M.; Vilaseca, M.; Nebreda, A. R.; Pons, M. Multi-Phosphorylation of the Intrinsically Disordered Unique Domain of c-Src Studied by In-Cell and Real-Time NMR Spectroscopy. *ChemBioChem* **2013**, *14*, 1820–1827.

(235) Banci, L.; Barbieri, L.; Bertini, I.; Luchinat, E.; Secci, E.; Zhao, Y.; Aricescu, A. R. Atomic-Resolution Monitoring of Protein Maturation in Live Human Cells by NMR. *Nat. Chem. Biol.* **2013**, *9*, 297–299.

(236) Bertini, I.; Felli, I. C.; Gonnelli, L.; Kumar M., V.; Pierattelli, R. $^{13}$C Direct-Detection Biomolecular NMR Spectroscopy in Living Cells. *Angew. Chemie Int. Ed.* **2011**, *50*, 2339–2341.

(237) Wang, Y.; Li, C.; Pielak, G. J. Effects of Proteins on Protein Diffusion. *J. Am. Chem. Soc.* **2010**, *16*, 9392–9397.

(238) Sarkar, M.; Pielak, G. J. An Osmolyte Mitigates the Destabilizing Effect of Protein Crowding. *Protein Sci.* **2014**, *23*, 1161–1164.

(239) Liokatis, S.; Dose, A.; Schwarzer, D.; Selenko, P. Simultaneous Detection of Protein Phosphorylation and Acetylation by High-Resolution NMR Spectroscopy. *J. Am. Chem. Soc.* **2010**, *132*, 14704–14705.

(240) Liu, X.; Yang, W.; Gao, Q.; Regnier, F. Toward Chromatographic Analysis of Interacting Protein Networks. *J. Chromatogr. A.* **2008**, *1178*, 24–32.

(241) Ruben, S.; Perkins, A.; Purcell, R.; Joung, K.; Sia, R.; Burghoff, R.; Haseltine, W. A.; Rosen, C. A. Structural and Functional Characterization of Human Immunodeficiency Virus Tat Protein. *J. Virol.* **1989**, *63*, 1–8.

(242) André, I.; Kesvatera, T.; Jönsson, B.; Linse, S. Salt Enhances Calmodulin-Target Interaction. *Biophys. J.* **2006**, *90*, 2903–2910.

(243) Raghuraman, H.; Ganguly, S.; Chattopadhyay, A. Effect of Ionic Strength on the Organization and Dynamics of Membrane-Bound Melittin. *Biophys. Chem.* **2006**, *124*, 115–124.

(244) Buell, A. K.; Hung, P.; Salvatella, X.; Welland, M. E.; Dobson, C. M.; Knowles, T. P. J. Electrostatic Effects in Filamentous Protein Aggregation. *Biophys. J.* **2013**, *104*, 1116–1126.

(245) Kontur, W. S.; Capp, M. W.; Gries, T. J.; Saecker, R. M.; Record, M. T. J. Probing DNA Binding, DNA Opening, and Assembly of a Downstream Clamp/Jaw in *Escherichia coli* RNA Polymerase-lambdaP(R) Promoter

Complexes Using Salt and the Physiological Anion Glutamate. *Biochemistry* **2010**, *49*, 4361–4373.

(246) Gokarn, Y. R.; Fesinmeyer, R. M.; Saluja, A.; Razinkov, V.; Chase, S. F.; Laue, T. M.; Brems, D. N. Effective Charge Measurements Reveal Selective and Preferential Accumulation of Anions, But Not Cations, at the Protein Surface in Dilute Salt Solutions. *Protein Sci.* **2011**, *20*, 580–587.

(247) Collins, K. D. Ion Hydration: Implications for Cellular Function, Polyelectrolytes, and Protein Crystallization. *Biophys. Chem.* **2006**, *119*, 271–281.

(248) Collins, K. D. Sticky Ions in Biological Systems. *Proc. Natl. Acad. Sci.* **1995**, *92*, 5553–5557.

(249) Zhang, Y.; Cremer, P. S. The Inverse and Direct Hofmeister Series for Lysozyme. *Proc. Natl. Acad. Sci.* **2009**, *106*, 15249–15253.

(250) Studier, F. W. Protein Production by Auto-Induction in High Density Shaking Cultures. *Protein Expr. Purif.* **2005**, *41*, 207–234.

(251) Mori, S.; Abeygunawardana, C.; Johnson, M. O.; Vanzijl, P. C. M. Improved Sensitivity of HSQC Spectra of Exchanging Protons at Short Interscan Delays Using a New Fast HSQC (FHSQC) Detection Scheme that Avoids Water Saturation. *J. Magn. Reson. Ser. B* **1995**, *108*, 94–98.

(252) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins* **2005**, *59*, 687–696.

(253) Davidson, A.; Patora-Komisarska, K.; Robinson, J. A.; Varani, G. Essential Structural Requirements for Specific Recognition of HIV TAR RNA by Peptide Mimetics of Tat Protein. *Nucleic Acids Res.* **2011**, *39*, 248–256.

(254) Gautier, V. W.; Gu, L.; O'Donoghue, N.; Pennington, S.; Sheehy, N.; Hall, W. W. *In Vitro* Nuclear Interactome of the HIV-1 Tat Protein. *Retrovirology* **2009**, *6*, 47–64.

(255) Austin, R. J.; Xia, T.; Ren, J.; Takahashi, T. T.; Roberts, R. W. Designed Arginine-Rich RNA-Binding Peptides with Picomolar Affinity. *J. Am. Chem. Soc.* **2002**, *124*, 10966–10967.

(256) Crowley, P. B.; Golovin, A. Cation-Pi Interactions in Protein-Protein Interfaces. *Proteins* **2005**, *59*, 231–239.

(257) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Hot Spots-a Review of the Protein-Protein Interface Determinant Amino-Acid Residues. *Proteins* **2007**, *68*, 803–812.

(258) Bahadur, R. P.; Zacharias, M.; Janin, J. Dissecting Protein-RNA Recognition Sites. *Nucleic Acids Res.* **2008**, *36*, 2705–2716.

(259) Klein, D. J.; Schmeing, T. M.; Moore, P. B.; Steitz, T. A. The Kink-Turn: A New RNA Secondary Structure Motif. *EMBO J.* **2001**, *20*, 4214–4221.

(260) Groninger-Poe, F. P.; Bouvier, J. T.; Vetting, M. W.; Kalyanaraman, C.; Kumar, R.; Almo, S. C.; Jacobson, M. P.; Gerlt, J. A. Evolution of Enzymatic Activities in the Enolase Superfamily: Galactarate Dehydratase III from *Agrobacterium Tumefaciens C58*. *Biochemistry* **2014**, *53*, 4192–4203.

(261) Dudev, M.; Lim, C. Discovering Structural Motifs Using a Structural Alphabet: Application to Magnesium-Binding Sites. *BMC Bioinformatics* **2007**, *8*, 1–12.

(262) Shojania, S.; O'Neil, J. D. HIV-1 Tat is a Natively Unfolded Protein: The Solution Conformation and Dynamics of Reduced HIV-1 Tat-(1-72) by NMR Spectroscopy. *J. Biol. Chem.* **2006**, *281*, 8347–8356.

(263) Guenther, U.-P.; Yandek, L. E.; Niland, C. N.; Campbell, F. E.; Anderson, D.; Anderson, V. E.; Harris, M. E.; Jankowsky, E. Hidden Specificity in an Apparently Nonspecific RNA-Binding Protein. *Nature* **2013**, *502*, 385–388.

(264) Page, M. I.; Jencks, W. P. Entropic Contributions to Rate Accelerations in Enzymic and Intramolecular Reactions and the Chelate Effect. *Proc. Natl. Acad. Sci.* **1971**, *68*, 1678–1683.

(265) Lo Conte, L.; Chothia, C.; Janin, J. The Atomic Structure of Protein-Protein Recognition Sites. *J. Mol. Biol.* **1999**, *285*, 2177–2198.

(266) Crowley, P. B.; Carrondo, M. A. The Architecture of the Binding Site in Redox Protein Complexes: Implications for Fast Dissociation. *Proteins* **2004**, *55*, 603–612.

(267) Sheinerman, F. B.; Norel, R.; Honig, B. Electrostatic Aspects of Protein–Protein Interactions. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153–159.

(268) Crowley, P. B.; Ubbink, M. Close Encounters of the Transient Kind: Protein Interactions in the Photosynthetic Redox Chain Investigated by NMR Spectroscopy. *Acc. Chem. Res.* **2003**, *36* , 723–730.

(269) Bi, E. F.; Lutkenhaus, J. FtsZ Ring Structure Associated with Division in *Escherichia coli*. *Nature* **1991**, *354*, 161–164.

(270) Nel, A. E.; Madler, L.; Velegol, D.; Xia, T.; Hoek, E. M. V; Somasundaran, P.; Klaessig, F.; Castranova, V.; Thompson, M. Understanding Biophysicochemical Interactions at the Nano-Bio Interface. *Nat. Mater.* **2009**, *8*, 543–557.

(271) Filoti, D. I.; Shire, S. J.; Yadav, S.; Laue, T. M. Comparative Study of Analytical Techniques for Determining Protein Charge. *J. Pharm. Sci.* **2015**, *104*, 2123–2331.

(272) Menon, M. K.; Zydney, A. L. Measurement of Protein Charge and Ion Binding Using Capillary Electrophoresis. *Anal. Chem.* **1998**, *70*, 1581–1584.

(273) Jordan, K. *Real-Time Electrophoretic Mobility in Membrane Confined Electrophoresis*; University of New Hampshire, Durham New Hampshire: MSc Thesis, **2014**.

(274) Moody, T. P.; Shepard, H. K. Nonequilibrium Thermodynamics of Membrane-Confined Electrophoresis. *Biophys. Chem.* **2004**, *108*, 51–76.

(275) Babu, Y. S.; Bugg, C. E.; Cook, W. J. Structure of Calmodulin Refined at 2.2 Å Resolution. *J. Mol. Biol.* **1988**, *204*, 191–204.

(276) Moody, T. P.; Kingsbury, J. S.; Durant, J. A.; Wilson, T. J.; Chase, S. F.; Laue, T. M. Valence and Anion Binding of Bovine Ribonuclease A Between pH 6 and 8. *Anal. Biochem.* **2005**, *336*, 243–252.

(277) Henry, D. C. The Cataphoresis of Suspended Particles. Part I. The Equation of Cataphoresis. *Proc. R. Soc. London A Math. Phys. Eng. Sci.* **1931**, *133*, 106–129.

(278) Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A. Calmodulin Structure Refined at 1.7 Å Resolution. *J. Mol. Biol.* **1992**, *228*, 1177–1192.

(279) Li, L.; Shi, X.; Guo, X.; Li, H.; Xu, C. Ionic Protein-Lipid Interaction at the Plasma Membrane: What Can the Charge Do? *Trends Biochem. Sci.* **2014**, *39*, 130–140.

(280) Peleh, V.; Riemer, J.; Dancis, A.; Herrmann, J. M. Protein Oxidation in the Intermembrane Space of Mitochondria is Substrate-Specific Rather than General. *Microb. Cell* **2014**, *1*, 81–93.

(281) Linden, M.; Andersson, G.; Gellerfors, P.; Nelson, B. D. Subcellular Distribution of Rat Liver Porin. *Biochim. Biophys. Acta* **1984**, *770*, 93–96.

(282) Cortese, J. D.; Voglino, A. L.; Hackenbrock, C. R. Ionic Strength of the Intermembrane Space of Intact Mitochondria as Estimated with Fluorescein-BSA Delivered by Low pH Fusion. *J. Cell Biol.* **1991**, *113*, 1331–1340.

(283) Mochan, E. The Nature of Complex Formation Between Cytochrome *c* and Cytochrome *c* Peroxidase. *Biochim. Biophys. Acta - Bioenerg.* **1970**, *216*, 80–95.

(284) Smith, H. T.; Ahmed, A. J.; Millett, F. Electrostatic Interaction of Cytochrome *c* with Cytochrome *c*1 and Cytochrome Oxidase. *J. Biol. Chem.* **1981**, *256*, 4984–4990.

(285) Stonehuerner, J.; Williams, J. B.; Millett, F. Interaction between Cytochrome *c* and Cytochrome *b*5. *Biochemistry* **1979**, *18*, 5422–5427.

(286) Hazzard, J. T.; Cusanovich, M. A.; Tainer, J. A.; Getzoff, E. D.; Tollin, G. Kinetic Studies of Reduction of a 1:1 Cytochrome *c*-Flavodoxin Complex by Free Flavin Semiquinones and Rubredoxin. *Biochemistry* **1986**, *25*, 3318–3328.

(287) Lange, C.; Hunte, C. Crystal Structure of the Yeast Cytochrome *bc*1 Complex with its Bound Substrate Cytochrome *c*. *Proc. Natl. Acad. Sci.* **2002**, *99*, 2800–2805.

(288) Crowley, P. B.; Kyne, C.; Monteith, W. B. Simple and Inexpensive Incorporation of $^{19}$F-Tryptophan for Protein NMR Spectroscopy. *Chem. Commun*. **2012**, 10681–10683.

(289) Van Gelder, B.; Slater, E. C. The Extinction Coefficient of Cytochrome *c*. *Biochim. Biophys. Acta* **1962**, *58*, 593–595.

(290) Schuck, P. Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys. J.* **2000**, *78*, 1606–1619.

(291) Sezonov, G.; Joseleau-Petit, D.; D'Ari, R. *Escherichia coli* Physiology in Luria-Bertani Broth. *J. Bacteriol.* **2007**, *189* (23), 8746–8749.

(292) Weigelt, J. Single Scan, Sensitivity- and Gradient-Enhanced TROSY for Multidimensional NMR Experiments. *J. Am. Chem. Soc.* **1998**, *120*, 10778–10779.

(293) Thurlkill, R. L.; Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. pK Values of the Ionizable Groups of Proteins. *Protein Sci.* **2006**, *15*, 1214–1218.

(294) Fitch, C. A.; Platzer, G.; Okon, M.; Garcia-Moreno, B. E.; McIntosh, L. P. Arginine: Its pKa Value Revisited. *Protein Sci.* **2015**, *24*, 752–761.

(295) Winzor, D. J. Determination of the Net Charge (Valence) of a Protein: A Fundamental but Elusive Parameter. *Anal. Biochem.* **2004**, *325*, 1–20.

(296) Fülöp, V.; Sam, K. A.; Ferguson, S. J.; Ginger, M. L.; Allen, J. W. A. Structure of a Trypanosomatid Mitochondrial Cytochrome *c* with Heme Attached via only One Thioether Bond and Implications for the Substrate Recognition Requirements of Heme Lyase. *FEBS J.* **2009**, *276*, 2822–2832.

(297) Durant, J. A.; Chen, C.; Laue, T. M.; Moody, T. P.; Allison, S. A. Use of T4 Lysozyme Charge Mutants to Examine Electrophoretic Models. *Biophys. Chem.* **2002**, *101-102*, 593–609.

(298) Hirota, S.; Hattori, Y.; Nagao, S.; Taketa, M.; Komori, H.; Kamikubo, H.; Wang, Z.; Takahashi, I.; Negi, S.; Sugiura, Y.; Kataoka, M.; Higuchi, Y. Cytochrome *c* Polymerization by Successive Domain Swapping at the C-Terminal Helix. *Proc. Natl. Acad. Sci.* **2010**, *107*, 12854–12859.

(299) Marcus, Y. Ionic Radii in Aqueous Solutions. *Chem. Rev.* **1988**, *88*, 1475–1498.

(300) Hirsch, A. K. H.; Fischer, F. R.; Diederich, F. Phosphate Recognition in Structural Biology. *Angew. Chem. Int. Ed.* **2007**, *46*, 338–352.

(301) Copley, R. R.; Barton, G. J. A Structural Analysis of Phosphate and Sulphate Binding Sites in Proteins. Estimation of Propensities for Binding and Conservation of Phosphate Binding Sites. *J. Mol. Biol.* **1994**, *242*, 321–329.

(302) Denessiouk, K. A.; Johnson, M. S.; Denesyuk, A. I. Novel $C^{\alpha}NN$ Structural Motif for Protein Recognition of Phosphate Ions. *J. Mol. Biol.* **2005**, *345*, 611–629.

(303) Subramanian, M.; Jutila, A.; Kinnunen, P. K. J. Binding and Dissociation of Cytochrome *c* to and from Membranes Containing Acidic Phospholipids. *Biochemistry* **1998**, *37*, 1394–1402.

(304) Cohen, D. S.; Pielak, G. J. Stability of Yeast Iso-1-Ferricytochrome *c* as a Function of pH and Temperature. *Protein Sci.* **1994**, *3*, 1253–1260.

(305) Brych, S. R.; Gokarn, Y. R.; Hultgen, H.; Stevenson, R. J.; Rajan, R.; Matsumura, M. Characterization of Antibody Aggregation: Role of Buried, Unpaired Cysteines in Particle Formation. *J. Pharm. Sci.* **2010**, *99*, 764–781.

(306) Horst, R.; Horwich, A. L.; Wuthrich, K. Translational Diffusion of Macromolecular Assemblies Measured Using Transverse-Relaxation-Optimized Pulsed Field Gradient NMR. *J. Am. Chem. Soc.* **2011**, *133*, 16354–16357.

(307) Laue, T. Proximity Energies: A Framework for Understanding Concentrated Solutions. *J. Mol. Recognit.* **2012**, *25*, 165–173.

(308) Crowley, P. B.; Brett, K.; Muldoon, J. NMR Spectroscopy Reveals Cytochrome *c*–Poly(Ethylene Glycol) Interactions. *ChemBioChem* **2008**, *9*, 685–688.

(309) Kuntz, I. D. Hydration of Macromolecules. III. Hydration of Polypeptides. *J. Am. Chem. Soc.* **1971**, *93*, 514–516.

(310) Trevino, S. R.; Scholtz, J. M.; Pace, C. N. Amino Acid Contribution to Protein Solubility: Asp, Glu, and Ser Contribute More Favorably than the Other Hydrophilic Amino Acids in RNase Sa. *J. Mol. Biol.* **2007**, *366*, 449–460.

(311) Von Hippel, P. H.; Berg, O. G. Facilitated Target Location in Biological Systems. *J. Biol. Chem.* **1989**, *264*, 675–678.

(312) Kornberg, R. D. Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* **1974**, *184*, 868–871.

(313) Han, M.; Chang, M.; Kim, U.-J.; Grunstein, M. Histone H2B Repression Causes Cell-Cycle-Specific Arrest in Yeast: Effects on Chromosomal Segregation, Replication, and Transcription. *Cell* **1987**, *48*, 589–597.

(314) Locke, G.; Tolkunov, D.; Moqtaderi, Z.; Struhl, K.; Morozov, A. V. High-Throughput Sequencing Reveals a Simple Model of Nucleosome Energetics. *Proc. Natl. Acad. Sci.* **2010**, *107*, 20998–21003.

(315) Shivaswamy, S.; Bhinge, A.; Zhao, Y.; Jones, S.; Hirst, M.; Iyer, V. R. Dynamic Remodeling of Individual Nucleosomes Across a Eukaryotic Genome in Response to Transcriptional Perturbation. *PLoS Biol.* **2008**, *6*, e65.

(316) Libertini, L. J.; Small, E. W. Effects of pH on Low-Salt Transition of Chromatin Core Particles. *Biochemistry* **1982**, *21*, 3327–3334.

(317) Yager, T. D.; McMurray, C. T.; Van Holde, K. E. Salt-Induced Release of DNA from Nucleosome Core Particles. *Biochemistry* **1989**, *28*, 2271–2281.

(318) Oliva, R.; Bazett-Jones, D. P.; Locklear, L.; Dixon, G. H. Histone Hyperacetylation can Induce Unfolding of the Nucleosome Core Particle. *Nucleic Acids Res.* **1990**, *18*, 2739–2747.

(319) Xu, F.; Zhang, K.; Grunstein, M. Acetylation in Histone H3 Globular Domain Regulates Gene Expression in Yeast. *Cell* **2005**, *121*, 375–385.

(320) Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution. *Nature* **1997**, *389*, 251–260.

(321) Arents, G.; Moudrianakis, E. N. The Histone Fold: A Ubiquitous Architectural Motif Utilized in DNA Compaction and Protein Dimerization. *Proc. Natl. Acad. Sci.* **1995**, *92*, 11170–11174.

(322) Sullivan, S. A.; Landsman, D. Characterization of Sequence Variability in Nucleosome Core Histone Folds. *Proteins Struct. Funct. Bioinforma.* **2003**, *52*, 454–465.

(323) Hadnagy, A.; Beaulieu, R.; Balicki, D. Histone Tail Modifications and Noncanonical Functions of Histones: Perspectives in Cancer Epigenetics. *Mol. Cancer Ther.* **2008**, *7*, 740–748.

(324) Gurley, L. R.; London, J. E.; Valdez, J. G. High-Performance Capillary Electrophoresis of Histones. *J. Chromatogr.* **1991**, *559*, 431–443.

(325) Kim, H. W.; Perez, J. a; Ferguson, S. J.; Campbell, I. D. The Specific Incorporation of Labelled Aromatic Amino Acids into Proteins Through Growth of Bacteria in the Presence of Glyphosate. Application to Fluorotryptophan Labelling to the H(+)-ATPase of *Escherichia coli* and NMR Studies. *FEBS Lett.* **1990**, *272*, 34–36.

(326) Waugh, D. S. Genetic Tools for Selective Labeling of Proteins with α-$^{15}$N-Amino Acids. *J. Biomol. NMR* **1996**, *8*, 184–192.

(327) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the Genetic Code of *Escherichia coli*. *Science* **2001**, *292*, 498–500.

(328) Tugarinov, V.; Kanelis, V.; Kay, L. E. Isotope Labeling Strategies for the Study of High-Molecular-Weight Proteins by Solution NMR Spectroscopy. *Nat. Protoc.* **2006**, *1*, 749–754.

(329) O'Grady, C.; Rempel, B. L.; Sokaribo, A.; Nokhrin, S.; Dmitriev, O. Y. One-Step Amino Acid Selective Isotope Labeling of Proteins in Prototrophic *Escherichia coli* Strains. *Anal. Biochem.* **2012**, *426*, 126–128.

(330) Kigawa, T.; Muto, Y.; Yokoyama, S. Cell-Free Synthesis and Amino Acid-Selective Stable Isotope Labeling of Proteins for NMR Analysis. *J. Biomol. NMR* **1995**, *6*, 129–134.

(331) Smith, R.; Brereton, I. M.; Chai, R. Y.; Kent, S. B. Ionization States of the Catalytic Residues in HIV-1 Protease. *Nat. Struct. Biol.* **1996**, *3*, 946–950.

(332) Hattori, Y.; Furuita, K.; Ohki, I.; Ikegami, T.; Fukada, H.; Shirakawa, M.; Fujiwara, T.; Kojima, C. Utilization of Lysine $^{13}$C-Methylation NMR for Protein-Protein Interaction Studies. *J. Biomol. NMR* **2013**, *55*, 19–31.

(333) Serber, Z.; Ledwidge, R.; Miller, S. M.; Dötsch, V. Evaluation of Parameters Critical to Observing Proteins Inside Living *Escherichia coli* by In-Cell NMR Spectroscopy. *J. Am. Chem. Soc.* **2001**, *123*, 8895–8901.

(334) Tanio, M.; Tanaka, R.; Tanaka, T.; Kohno, T. Amino Acid-Selective Isotope Labeling of Proteins for Nuclear Magnetic Resonance Study: Proteins Secreted by *Brevibacillus Choshinensis*. *Anal. Biochem.* **2009**, *386*, 156–160.

(335) White, C. E.; Kempi, N. M.; Komives, E. A. Expression of Highly Disulfide-bonded Proteins in *Pichia Pastoris*. *Structure* **1994**, *2*, 1003–1005.

(336) Strauss, A.; Bitsch, F.; Cutting, B.; Fendrich, G.; Graff, P.; Liebetanz, J.; Zurini, M.; Jahnke, W. Amino–acid-Type Selective Isotope Labeling of Proteins Expressed in *Baculovirus*-Infected Insect Cells Useful for NMR Studies. *J. Biomol. NMR* **2003**, *26*, 367–372.

(337) Archer, S. J.; Bax, A.; Roberts, A. B.; Sporn, M. B.; Ogawa, Y.; Piez, K. A.; Weatherbee, J. A.; Tsang, M. L. S.; Lucas, R. Transforming Growth Factor β1: NMR Signal Assignments of the Recombinant Protein Expressed and Isotopically Enriched Using Chinese Hamster Ovary Cells. *Biochemistry* **1993**, *32*, 1152–1163.

(338) Sprangers, R.; Kay, L. E. Quantitative Dynamics and Binding Studies of the 20S Proteasome by NMR. *Nature* **2007**, *445*, 618–622.

(339) Jones, D. H.; Cellitti, S. E.; Hao, X.; Zhang, Q.; Jahnz, M.; Summerer, D.; Schultz, P. G.; Uno, T.; Geierstanger, B. H. Site-Specific Labeling of Proteins with NMR-Active Unnatural Amino Acids. *J. Biomol. NMR* **2010**, *46*, 89–100.

(340) Cattani, G.; Vogeley, L.; Crowley, P. B. Structure of a PEGylated Protein Reveals a Highly Porous Double-Helical Assembly. *Nat. Chem.* **2015**, *7*, 823–828.

(341) Tong, K. I.; Yamamoto, M.; Tanaka, T. A Simple Method for Amino Acid Selective Isotope Labeling of Recombinant Proteins in *E. coli*. *J. Biomol. NMR* **2008**, *42*, 59–67.

(342) Matchett, W. H. Inhibition of Tryptophan Synthetase by Indoleacrylic Acid. *J. Bacteriol.* **1972**, *110*, 146–154.

(343) Leone, M.; Rodriguez-Mias, R. A.; Pellecchia, M. Selective Incorporation of $^{19}$F-Labeled Trp Side Chains for NMR-Spectroscopy-Based Ligand-Protein Interaction Studies. *ChemBioChem* **2003**, *4*, 649–650.

(344) El Khattabi, M.; van Roosmalen, M. L.; Jager, D.; Metselaar, H.; Permentier, H.; Leenhouts, K.; Broos, J. *Lactococcus Lactis* as Expression Host for the Biosynthetic Incorporation of Tryptophan Analogues into Recombinant Proteins. *Biochem. J.* **2008**, *409*, 193–198.

(345) Wals, K.; Ovaa, H. Unnatural Amino Acid Incorporation in *E. coli*: Current and Future Applications in the Design of Therapeutic Proteins. *Front. Chem.* **2014**, *2*, 15–26.

(346) Jackson, J. C.; Hammill, J. T.; Mehl, R. A. Site-Specific Incorporation of a $^{19}$F-Amino Acid into Proteins as an NMR Probe for Characterizing Protein Structure and Reactivity. *J. Am. Chem. Soc.* **2007**, *129*, 1160–1166.

(347) Chen, H.; Viel, S.; Ziarelli, F.; Peng, L. $^{19}$F NMR: A Valuable Tool for Studying Biological Events. *Chem. Soc. Rev.* **2013**, *42*, 7971–7982.

(348) Smith, A. E.; Zhang, Z.; Pielak, G. J.; Li, C. NMR Studies of Protein Folding and Binding in Cells and Cell-Like Environments. *Curr. Opin. Struct. Biol.* **2015**, *30*, 7–16.

(349) Sykes, B. D.; Weingarten, H. I.; Schlesinger, M. J. Fluorotyrosine Alkaline Phosphatase from *Escherichia coli*: Preparation, Properties, and Fluorine-19 Nuclear Magnetic Resonance Spectrum. *Proc. Natl. Acad. Sci.* **1974**, *71*, 469–473.

(350) Frieden, C.; Hoeltzli, S. D.; Bann, J. G. The Preparation of $^{19}$F-Labeled Proteins for NMR Studies. *Methods Enzymol.* **2004**, *380*, 400–415.

(351) O'Hagan, D.; B. Harper, D. Fluorine-Containing Natural Products. *J. Fluor. Chem.* **1999**, *100*, 127–133.

(352) Lian, C.; Le, H.; Montez, B.; Patterson, J.; Harrell, S.; Laws, D.; Matsumura, I.; Pearson, J.; Oldfield, E. Fluorine-19 Nuclear Magnetic Resonance Spectroscopic Study of Fluorophenylalanine- and Fluorotryptophan-Labeled Avian Egg White Lysozymes. *Biochemistry* **1994**, *33*, 5238–5245.

(353) Liu, J. J.; Horst, R.; Katritch, V.; Stevens, R. C.; Wuthrich, K. Biased Signaling Pathways in β2-Adrenergic Receptor Characterized by $^{19}$F-NMR. *Science* **2012**, *335*, 1106–1110.

(354) Bann, J. G.; Pinkner, J.; Hultgren, S. J.; Frieden, C. Real-Time and Equilibrium $^{19}$F-NMR Studies Reveal the Role of Domain-Domain Interactions in the Folding of the Chaperone PapD. *Proc. Natl. Acad. Sci.* **2002**, *99*, 709–714.

(355) Evanics, F.; Bezsonova, I.; Marsh, J.; Kitevski, J. L.; Forman-Kay, J. D.; Prosser, R. S. Tryptophan Solvent Exposure in Folded and Unfolded States of an SH3 Domain by $^{19}$F and $^{1}$H NMR. *Biochemistry* **2006**, *45*, 14120–14128.

(356) Khan, F.; Kuprov, I.; Craggs, T. D.; Hore, P. J.; Jackson, S. E. $^{19}$F NMR Studies of the Native and Denatured States of Green Fluorescent Protein. *J. Am. Chem. Soc.* **2006**, *128*, 10729–10737.

(357) Li, C.; Lutz, E. A.; Slade, K. M.; Ruf, R. A. S.; Wang, G.-F.; Pielak, G. J. $^{19}$F NMR Studies of α-Synuclein Conformation and Fibrillation. *Biochemistry* **2009**, *48*, 8578–8584.

(358) Liu, L.; Byeon, I.-J. L.; Bahar, I.; Gronenborn, A. M. Domain Swapping Proceeds via Complete Unfolding: A $^{19}$F- and $^{1}$H-NMR Study of the Cyanovirin-N Protein. *J. Am. Chem. Soc.* **2012**, *134*, 4229–4235.

(359) Aramini, J. M.; Hamilton, K.; Ma, L.-C.; Swapna, G. V. T.; Leonard, P. G.; Ladbury, J. E.; Krug, R. M.; Montelione, G. T. $^{19}$F NMR Reveals Multiple Conformations at the Dimer Interface of the Nonstructural Protein 1 Effector Domain from Influenza A Virus. *Structure* **2014**, *22*, 515–525.

(360) Gee, C. T.; Koleski, E. J.; Pomerantz, W. C. K. Fragment Screening and Druggability Assessment for the CBP/p300 KIX Domain through Protein-Observed $^{19}$F NMR Spectroscopy. *Angew. Chem. Int. Ed.* **2015**, *54*, 3735–3739.

(361) Mishra, N. K.; Urick, A. K.; Ember, S. W. J.; Schönbrunn, E.; Pomerantz, W. C. Fluorinated Aromatic Amino Acids are Sensitive $^{19}$F NMR Probes for Bromodomain-Ligand Interactions. *ACS Chem. Biol.* **2014**, *9*, 2755–2760.

(362) Ye, Y.; Liu, X.; Chen, Y.; Xu, G.; Wu, Q.; Zhang, Z.; Yao, C.; Liu, M.; Li, C. Labeling Strategy and Signal Broadening Mechanism of Protein NMR Spectroscopy in *Xenopus Laevis* Oocytes. *Chem. – A Eur. J.* **2015**, *21*, 8686–8690.

(363) Browne, D. R.; Kenyon, G. L.; Hegeman, G. D. Incorporation of Monoflurotryptophans into Protein During the Growth of *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **1970**, *39*, 13–19.

(364) Salonen, L. M.; Ellermann, M.; Diederich, F. Aromatic Rings in Chemical and Biological Recognition: Energetics and Structures. *Angew. Chemie Int. Ed.* **2011**, *50*, 4808–4842.

(365) Bahadur, R. P.; Chakrabarti, P.; Rodier, F.; Janin, J. A Dissection of Specific and Non-Specific Protein-Protein Interfaces. *J. Mol. Biol.* **2004**, *336*, 943–955.

(366) Schneider, T. R.; Gerhardt, E.; Lee, M.; Liang, P. H.; Anderson, K. S.; Schlichting, I. Loop Closure and Intersubunit Communication in Tryptophan Synthase. *Biochemistry* **1998**, *37*, 5394–5406.

(367) Ngo, H.; Kimmich, N.; Harris, R.; Niks, D.; Blumenstein, L.; Kulik, V.; Barends, T. R.; Schlichting, I.; Dunn, M. F. Allosteric Regulation of Substrate Channeling in Tryptophan Synthase: Modulation of the L-Serine Reaction in Stage I of the β-Reaction by α-Site Ligands. *Biochemistry* **2007**, *46*, 7740–7753.

(368) Rodriguez-Mias, R. A.; Pellecchia, M. Use of Selective Trp Side Chain Labeling to Characterize Protein-Protein and Protein-Ligand Interactions by NMR Spectroscopy. *J. Am. Chem. Soc.* **2003**, *125*, 2892–2893.

(369) Broos, J.; Gabellieri, E.; Biemans-Oldehinkel, E.; Strambini, G. B. Efficient Biosynthetic Incorporation of Tryptophan and Indole Analogs in an Integral Membrane Protein. *Protein Sci.* **2003**, *12*, 1991–2000.

(370) Goss, R. J. M.; Newill, P. L. A. A Convenient Enzymatic Synthesis of L-Halotryptophans. *Chem. Commun.* **2006**, 47, 4924–4925.

(371) Pollock, W. B.; Rosell, F. I.; Twitchett, M. B.; Dumont, M. E.; Mauk, A. G. Bacterial Expression of a Mitochondrial Cytochrome *c*. Trimethylation of

Lys72 in Yeast Iso-1-Cytochrome *c* and the Alkaline Conformational Transition. *Biochemistry* **1998**, *37*, 6124–6131.

(372) Cobas, J. C.; Sardina, F. J. Nuclear Magnetic Resonance Data Processing. MestRe-C: A Software Package for Desktop Computers. *Concepts Magn. Reson. Part A* **2003**, *19A*, 80–96.

(373) Acchione, M.; Lee, Y.-C.; DeSantis, M. E.; Lipschultz, C. A.; Wlodawer, A.; Li, M.; Shanmuganathan, A.; Walter, R. L.; Smith-Gill, S.; Barchi, J. J. J. Specific Fluorine Labeling of the HyHEL10 Antibody Affects Antigen Binding and Dynamics. *Biochemistry* **2012**, *51*, 6017–6027.

(374) Campos-Olivas, R.; Aziz, R.; Helms, G. L.; Evans, J. N. S.; Gronenborn, A. M. Placement of $^{19}$F into the Center of GB1: Effects on Structure and Stability. *FEBS Lett.* **2002**, *517*, 55–60.

(375) Santiveri, C. M.; Jimenez, M. A. Tryptophan Residues: Scarce in Proteins but Strong Stabilizers of β-Hairpin Peptides. *Biopolymers* **2010**, *94*, 779–790.

(376) Piñero-Fernandez, S.; Chimerel, C.; Keyser, U. F.; Summers, D. K. Indole Transport across *Escherichia coli* Membranes *J. Bacteriol.* **2011**, *193*, 1793–1798.

(377) Bergmann, E. D.; Hoffmann, E. 549 6-Fluoro-, 6-Methoxy-, and 7-Methoxy-Tryptophan. *J. Chem. Soc.* **1962**, 2827–2829.

(378) Wang, G.; Zhang, Z.-T.; Jiang, B.; Zhang, X.; Li, C.; Liu, M. Recent Advances in Protein NMR Spectroscopy and their Implications in Protein Therapeutics Research. *Anal. Bioanal. Chem.* **2014**, *406*, 2279–2288.

(379) Paulini, R.; Muller, K.; Diederich, F. Orthogonal Multipolar Interactions in Structural Chemistry and Biology. *Angew. Chem. Int. Ed.* **2005**, *44*, 1788–1805.

(380) Witter, R.; Nozirov, F.; Sternberg, U.; Cross, T. A.; Ulrich, A. S.; Fu, R. Solid-State $^{19}$F NMR Spectroscopy Reveals that Trp41 Participates in the Gating Mechanism of the M2 Proton Channel of Influenza A Virus. *J. Am. Chem. Soc.* **2008**, *130*, 918–924.

(381) 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.

(382) Rowe, A. J. Ultra-Weak Reversible Protein-Protein Interactions. *Methods* **2011**, *54*, 157–166.

(383) Vaynberg, J.; Qin, J. Weak Protein–Protein Interactions as Probed by NMR Spectroscopy. *Trends Biotechnol.* **2006**, *24*, 22–27.

(384) Barbieri, L.; Luchinat, E.; Banci, L. Protein Interaction Patterns in Different Cellular Environments are Revealed by In-Cell NMR. *Sci. Rep.* **2015**, *5*, 14456–14465.

(385) 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*, 163–179.

(386) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR* **1995**, *6*, 277–293.

(387) Latham, M. P.; Kay, L. E. Is Buffer a Good Proxy for a Crowded Cell-Like Environment? A Comparative NMR Study of Calmodulin Side-Chain Dynamics in Buffer and *E. coli* Lysate. *PLoS One* **2012**, *7*, e48226.

(388) Latham, M. P.; Kay, L. E. Probing Non-Specific Interactions of $Ca^{2+}$-Calmodulin in *E. coli* Lysate. *J. Biomol. NMR* **2013**, *55*, 239–247.

(389) Sanfelice, D.; Politou, A.; Martin, S. R.; De Los Rios, P.; Temussi, P.; Pastore, A. The Effect of Crowding and Confinement: A Comparison of Yfh1 Stability in Different Environments. *Phys. Biol.* **2013**, *10*, 45002–45009.

(390) Miklos, A. C.; Li, C.; Sharaf, N. G.; Pielak, G. J. Volume Exclusion and Soft Interaction Effects on Protein Stability under Crowded Conditions. *Biochemistry* **2010**, *49*, 6984–6991.

(391) Fiaux, J.; Bertelsen, E. B.; Horwich, A. L.; Wuthrich, K. NMR Analysis of a 900K GroEL GroES Complex. *Nature* **2002**, *418*, 207–211.

(392) Kranz, J. K.; Lee, E. K.; Nairn, A. C.; Wand, A. J. A Direct Test of the Reductionist Approach to Structural Studies of Calmodulin Activity: Relevance of Peptide Models of Target Proteins. *J. Biol. Chem.* **2002**, *277*, 16351–16354.

(393) McGuffee, S. R.; Elcock, A. H. Atomically Detailed Simulations of Concentrated Protein Solutions: The Effects of Salt, pH, Point Mutations, and Protein Concentration in Simulations of 1000-Molecule Systems. *J. Am. Chem. Soc.* **2006**, *128*, 12098–12110.

(394) Pernodet, N.; Maaloum, M.; Tinland, B. Pore Size of Agarose Gels by Atomic Force Microscopy. *Electrophoresis* **1997**, *18*, 55–58.

(395) Zhu, Z. Y.; Karlin, S. Clusters of Charged Residues in Protein Three-Dimensional Structures. *Proc. Natl. Acad. Sci.* **1996**, *93*, 8350–8355.

(396) Xu, Y.; Wang, H.; Nussinov, R.; Ma, B. Protein Charge and Mass Contribute to the Spatio-Temporal Dynamics of Protein-Protein Interactions in a Minimal Proteome. *Proteomics* **2013**, *13*, 1339–1351.

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

(398) Popovic, M.; Sanfelice, D.; Pastore, C.; Prischi, F.; Temussi, P. A.; Pastore, A. Selective Observation of the Disordered Import Signal of a Globular Protein by In-Cell NMR: The Example of Frataxins. *Protein Sci.* **2015**, *24*, 996–1003.

(399) Schreiber, S. L. The Small-Molecule Approach to Biology. *Chem. Eng. News* **2003**, *81*, 51–61.

(400) Pastore, A. New Challenges in Structural Biology: Catching the Complexity of Dynamic Nanomachines. *Front. Mol. Biosci.* **2014**, *1*, 3–4.

(401) Higurashi, M.; Ishida, T.; Kinoshita, K. Identification of Transient Hub Proteins and the Possible Structural Basis for their Multiple Interactions. *Protein Sci.* **2008**, *17*, 72–78.

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

(403) Bolis, D.; Politou, A. S.; Kelly, G.; Pastore, A.; Andrea Temussi, P. Protein Stability in Nanocages: A Novel Approach for Influencing Protein Stability by Molecular Confinement. *J. Mol. Biol.* **2004**, *336*, 203–212.

(404) Phillip, Y.; Sherman, E.; Haran, G.; Schreiber, G. Common Crowding Agents Have Only a Small Effect on Protein-Protein Interactions. *Biophys. J.* **2009**, *97*, 875–885.

(405) Ådén, J.; Wittung-Stafshede, P. Folding of an Unfolded Protein by Macromolecular Crowding *in Vitro*. *Biochemistry* **2014**, *53*, 2271–2277.

(406) Chakraborty, K.; Chatila, M.; Sinha, J.; Shi, Q.; Poschner, B. C.; Sikor, M.; Jiang, G.; Lamb, D. C.; Hartl, F. U.; Hayer-Hartl, M. Chaperonin-Catalyzed Rescue of Kinetically Trapped States in Protein Folding. *Cell* **2010**, *142*, 112–122.

(407) Yu, Y.; Wang, J.; Liu, J.; Ling, D.; Xia, J. Functional Assembly of Protein Fragments Induced by Spatial Confinement. *PLoS One* **2015**, *10*, e0122101.

(408) 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.

(409) Schlarb-Ridley, B. G.; Mi, H.; Teale, W. D.; Meyer, V. S.; Howe, C. J.; Bendall, D. S. Implications of the Effects of Viscosity, Macromolecular Crowding, and Temperature for the Transient Interaction Between Cytochrome F and Plastocyanin from the Cyanobacterium *Phormidium Laminosum*. *Biochemistry* **2005**, *44*, 6232–6238.

(410) Minton, A. P. The Influence of Macromolecular Crowding and Macromolecular Confinement on Biochemical Reactions in Physiological Media. *J. Biol. Chem.* **2001**, *276*, 10577–10580.

(411) Matthew, J. B.; Weber, P. C.; Salemme, F. R.; Richards, F. M. Electrostatic Orientation During Electron Transfer Between Flavodoxin and Cytochrome *c*. *Nature* **1983**, *301*, 169–171.

(412) Kozer, N.; Kuttner, Y. Y.; Haran, G.; Schreiber, G. Protein-Protein Association in Polymer Solutions: From Dilute to Semidilute to Concentrated. *Biophys. J.* **2007**, *92*, 2139–2149.

(413) Kozer, N.; Schreiber, G. Effect of Crowding on Protein-Protein Association Rates: Fundamental Differences Between Low and High Mass Crowding Agents. *J Mol Biol* **2004**, *336*, 763–774.

(414) Li, C.; Wang, Y.; Pielak, G. J. Translational and Rotational Diffusion of a Small Globular Protein under Crowded Conditions. *J. Phys. Chem. B.* **2009**, *113*, 13390–13392.

(415) Volkov, A. N.; Nicholls, P.; Worrall, J. A. R. The Complex of Cytochrome *c* and Cytochrome *c* Peroxidase: The End of the Road? *Biochim. Biophys. Acta - Bioenerg.* **2011**, *1807*, 1482–1503.

(416) Pastore, A.; Salvadori, S.; Temussi, P. A. Peptides and Proteins in a Confined Environment: NMR Spectra at Natural Isotopic Abundance. *J. Pept. Sci.* **2007**, *13*, 342–347.

(417) Wang, Y. *Macromolecular Crowding and Protein Chemistry: Views from Inside and Outside Cells.* University of North Carolina at Chapel Hill, North Carolina, PhD Thesis, **2012**.

(418) Li, C.; Pielak, G. J. Using NMR to Distinguish Viscosity Effects from Nonspecific Protein Binding under Crowded Conditions. *J. Am. Chem. Soc.* **2009**, *131*, 1368–1369.

(419) Zuiderweg, E. R. P. Mapping Protein-Protein Interactions in Solution by NMR Spectroscopy. *Biochemistry* **2002**, *41*, 1–7.

(420) Barik, A.; C, N.; Pilla, S. P.; Bahadur, R. P. Molecular Architecture of Protein-RNA Recognition Sites. *J. Biomol. Struct. Dyn.* **2015**, *33*, 2738–2751.

(421) Fiorucci, S.; Zacharias, M. Prediction of Protein-Protein Interaction Sites Using Electrostatic Desolvation Profiles. *Biophys. J.* **2010**, *98*, 1921–1930.

(422) Niwa, T.; Ying, B.-W.; Saito, K.; Jin, W.; Takada, S.; Ueda, T.; Taguchi, H. Bimodal Protein Solubility Distribution Revealed by an Aggregation Analysis of the Entire Ensemble of *Escherichia coli* Proteins. *Proc. Natl. Acad. Sci.* **2009**, *106*, 4201–4206.

(423) Chan, P.; Curtis, R. A.; Warwicker, J. Soluble Expression of Proteins Correlates with a Lack of Positively-Charged Surface. *Sci. Rep.* **2013**, *3*, 3333–3336.

(424) Dyson, H. J.; Wright, P. E. Intrinsically Unstructured Proteins and their Functions. *Nat Rev Mol Cell Biol* **2005**, *6*, 197–208.

(425) Tan, R.; Chen, L.; Buettner, J. A.; Hudson, D.; Frankel, A. D. RNA Recognition by an Isolated α-Helix. *Cell* **1993**, *73*, 1031–1040.

(426) Birtalan, S.; Zhang, Y.; Fellouse, F. A.; Shao, L.; Schaefer, G.; Sidhu, S. S. The Intrinsic Contributions of Tyrosine, Serine, Glycine and Arginine to the Affinity and Specificity of Antibodies. *J. Mol. Biol.* **2008**, *377*, 1518–1528.

(427) Harms, M. J.; Schlessman, J. L.; Sue, G. R.; García-Moreno E., B. Arginine Residues at Internal Positions in a Protein are Always Charged. *Proc. Natl. Acad. Sci.* **2011**, *108*, 18954–18959.

(428) Jones, S.; Thornton, J. M. Principles of Protein-Protein Interactions. *Proc. Natl. Acad. Sci.* **1996**, *93*, 13–20.

(429) Rodier, F.; Bahadur, R. P.; Chakrabarti, P.; Janin, J. Hydration of Protein-Protein Interfaces. *Proteins* **2005**, *60*, 36–45.

(430) Schreiber, G.; Fleishman, S. J. Computational Design of Protein–Protein Interactions. *Curr. Opin. Struct. Biol.* **2013**, *23*, 903–910.

(431) Kosloff, M.; Travis, A. M.; Bosch, D. E.; Siderovski, D. P.; Arshavsky, V. Y. Integrating Energy Calculations with Functional Assays to Decipher the Specificity of G-Protein Inactivation by RGS Proteins. *Nat. Struct. Mol. Biol.* **2011**, *18*, 846–853.

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

(433) Eisenberg, D.; Schwarz, E.; Komaromy, M.; Wall, R. Analysis of Membrane and Surface Protein Sequences with the Hydrophobic Moment Plot. *J. Mol. Biol.* **1984**, *179*, 125–142.

(434) Kramer, R. M.; Shende, V. R.; Motl, N.; Pace, C. N.; Scholtz, J. M. Toward a Molecular Understanding of Protein Solubility: Increased Negative Surface Charge Correlates with Increased Solubility. *Biophys. J.* **2012**, *102*, 1907–1915.

(435) Iwahara, J.; Esadze, A.; Zandarashvili, L. Physicochemical Properties of Ion Pairs of Biological Macromolecules. *Biomolecules* **2015**, *5*, 2435–2463.

(436) Bar-Even, A.; Noor, E.; Flamholz, A.; Buescher, J. M.; Milo, R. Hydrophobicity and Charge Shape Cellular Metabolite Concentrations. *PLoS Comput. Biol.* **2011**, *7*, e1002166.

(437) Grabarczyk, D. B.; Chappell, P. E.; Johnson, S.; Stelzl, L. S.; Lea, S. M.; Berks, B. C. Structural Basis for Specificity and Promiscuity in a Carrier Protein/Enzyme System from the Sulfur Cycle. *Proc. Natl. Acad. Sci.* **2015**, *112*, E7166–E7175.

(438) Winkel, B. S. J. Metabolic Channeling in Plants. *Annu. Rev. Plant Biol.* **2004**, *55*, 85–107.

(439) Williams, K. Interactions of Polyamines with Ion Channels. *Biochem. J.* **1997**, *325*, 289–297.

(440) Miyamoto, S.; Kashiwagi, K.; Ito, K.; Watanabe, S.; Igarashi, K. Estimation of Polyamine Distribution and Polyamine Stimulation of Protein Synthesis in *Escherichia coli*. *Arch. Biochem. Biophys.* **1993**, *300*, 63–68.

(441) Miller-Fleming, L.; Olin-Sandoval, V.; Campbell, K.; Ralser, M. Remaining Mysteries of Molecular Biology: The Role of Polyamines in the Cell. *J. Mol. Biol.* **2015**, *427*, 3389–3406.

# Acknowledgements

My PhD experience has been transformative and illuminating. This journey was enriched by many people:

Firstly, I thank my supervisor **Dr. Peter Crowley** for the zeal, creativity and dedication that he has toward all aspects of his work. His enthusiasm and scientific expertise have been majorly influential to my development as a researcher. I appreciate all of the training and professional and moral support that he has provided. I am lucky to be his student.

I am grateful to the **National University of Ireland, Galway Hardiman Scholarship** for funding my PhD and establishing the Hardiman Scholar community, through which, I have met some fantastic friends.

I am also grateful to the **National University of Ireland Travelling Studentship** for funding my research in the Laue laboratory and my participation at six high-profile conferences. This funding has exposed me to the variety of work throughout the field of protein science and has enabled me to meet leaders across the field.

I am grateful to all of the past and current members of the **Crowley Research group** for their help in the lab and useful discussions.

I thank **Dr. Róisín Doohan** for NMR spectrometer maintenance and helpful comments on NMR experiment design. I also thank **Mr Gerard Fahy** for the acquisition of mass spectrometry data.

I wish to thank **Profs Thomas Laue and Gary Pielak** for allowing me to work in their laboratories and their continued discussions and support. I thank the **Laue and Pielak group members** for their training and assistance.

Finally, I am extremely grateful to my **family and friends** for their support and encouragement. In particular, I thank my parents, **Tim and Noreena**, for teaching me the value of hard work and perseverance. I thank my siblings, **Tadhg, Aoife and Proinnsias**, for their unwavering support, guidance and patience. Special thanks go to **Ethan and Timothy** for the colour that they bring to everything. Most importantly, I dedicate this thesis to my granny, **Mary Kyne**. I am still inspired by the many nature walks, projects and stories that you shared with me.

# *Curriculum Vitae*

Ciara Kyne was born on September the 11th 1989 in Galway, Ireland. During her BSc studies at NUI Galway, Ciara worked as a protein chemist under the supervision of Dr Peter Crowley. The work involved using NMR spectroscopy to determine the p$K_a$ values of ionisable groups in a protein and in-cell NMR studies. Ciara continued this project for her undergraduate thesis.

In 2011, before graduating with a 1st class honours degree in chemistry, Ciara was awarded a Korner Travelling Fellowship. This funding enabled her to perform a 6 week placement in Prof. Gary J. Pielak's laboratory at the University of North Carolina, Chapel Hill, USA. The work explored how cell growth conditions could be optimized to achieve high quality NMR spectra of a test protein in live *E. coli*. Later that year, Ciara joined the Crowley laboratory at NUI Galway to pursue her PhD. The aim of her research was to understand protein interactions under physiological conditions and to elaborate principles for describing cytoplasmic structuring. As part of her PhD, Ciara performed a 10 week placement in the laboratory of Prof. Thomas M. Laue at the University of New Hampshire, U.S.A. The work involved characterizing the charge and association state of cytochrome *c* and its charge-inverted mutants. The results of these projects are presented in this thesis.

# *Publications*

**Kyne C**, Crowley PB

Effects of crowding and confinement on quinary-like protein interactions.

*To be submitted*

**Kyne C**, Jordan K, Filoti DI, Laue TM, Crowley PB

Protein charge determination and implications for interactions in cell extracts

*To be submitted*

**Kyne C**, Crowley PB

Grasping the nature of the cell interior.

*FEBS J.* **2016***, In press*

**Kyne C**, Ruhle B, Gautier VW, Crowley PB.

Specific ion effects on macromolecular interactions in *Escherichia coli* extracts.

*Protein Sci.* **2015**, *24*, 310-318. *Highlighted Article*

Wettstein C, **Kyne C**, Doolan AM, Möhwald H, Crowley PB*, Lisdat F.

Study of cytochrome c-DNA interaction: evaluation of binding sites on the redox protein.

*Nanoscale* **2014**, *6*, 13779-13786.

Theillet FX, Binolfi A, Frembgen-Kesner T, Hingorani K, Sarkar M, **Kyne C**, Li C, Crowley PB, Gierasch L, Pielak GJ, Elcock AH, Gershenson A, Selenko P.

Physicochemical properties of cells and their effects on intrinsically disordered proteins.

*Chem. Rev.* **2014**, *114*, 6661-6714.

PB Crowley*, **C Kyne** and WB Monteith.

Simple and inexpensive incorporation of $^{19}$F-tryptophan into proteins.

*Chem. Commun.* **2012**, *48*, 10681-10683.

## *Oral Presentations*

1. Kyne, C. (2015, April) Assessing Protein Stickiness in the *E. coli* cytoplasm, NUI, Galway, School of Chemistry **BOC Postgraduate Award**

2. Kyne, C. (2013, April) Dissecting Biomolecular Machines inside *E. coli* Cells, NUI Galway – UL Alliance Research Day. **Prize for best poster**

3. Kyne, C (2013, July) Charge and Interaction Behaviour of Cytochrome *c* and its Charge Mutants, University of New Hampshire, Laue Group Seminar

## *Conferences & Workshops*

07/2015- Protein Society Symposium, Barcelona, Spain[*†]

10/2014- Disordered Motifs and Domains in Cell Control, Dublin, Ireland[†]

09/2014- Soft-Inter 2014: Soft Interactions in Biological and Biomimetic Self-Assemblies,Saint Malo, France[†]

12/2013- In-cell NMR Analysis of Biomolecular Structure and Function (Instruct Workshop), Florence, Italy[†]

06/2013- Gordon Research Conference: Proteins, New Hampshire, USA[*†]

09/2012- FASEB Protein Folding in the Cell, Vermont, USA[*†]

05/2012- EuroMAR, Dublin, Ireland[*]

[*]Selected to Present a Poster

[†]Funded by NUI Travelling Studentship

## *Extracurricular Contributions & Outreach Work*

06/2012-06/2013: Auditor of NUI Galway Chemistry Society

Coordinated a 14-part seminar series involving national/international researchers from Industry and Academia

06/2013-06/2014: Director of "Kitchen Chemistry", a postgraduate-led outreach initiative to promote Science in primary and secondary schools.