Published in final edited form as:

J Proteomics. 2019 January 16; 191: 202–211. doi:10.1016/j.jprot.2018.02.006.

Interaction Proteomics by using In-cell NMR Spectroscopy

Leonard Breindel¹, David S. Burz¹, and Alexander Shekhtman¹

¹Department of Chemistry, University at Albany, State University of New York, Albany, NY 12222, USA

Abstract

A synopsis of in-cell NMR spectroscopic approaches to study interaction proteomics in prokaryotic and eukaryotic cells is presented. We describe the use of in-cell NMR spectroscopy to resolve high resolution protein structures, discuss methodologies for determining and analyzing high and low affinity protein-target structural interactions, including intrinsically disordered proteins, and detail important functional interactions that result from these interactions.

1. Introduction

The holistic study of proteins that are produced and utilized by a cell, tissue or organism, *i.e.* the proteome, is called proteomics. The composition of the proteome changes temporally and spatially with cell growth and development, and in response to internal and external stimuli. The resulting cellular physiology arises from networks of interactions mediated by macromolecular assemblies of proteins and small regulatory molecules and ions. The formation of macromolecular assemblies, in turn, depends on the relative abundance of the constituents, which drives the states of association and propensity to interact. The study of how networks of protein interactions regulate biological activity is called interaction proteomics.

The combinatorial nature of protein interactions results in an overwhelming number of possibilities that one must consider to describe interaction proteomics in live cells. A more tractable approach is to focus on a particular target protein and the interactions it engages in. From the perspective of a singular protein, interaction proteomics can be divided into two broad categories, specific high affinity interactions with dissociation constants $<1 \mu\text{M}$, and specific low affinity interactions with dissociation constants $>1 \mu\text{M}$, with specificity implying that the same target protein binding surface is involved in the interaction. The specific low affinity interactions that are omnipresent in live cells are also called quinary interactions to distinguish them from high affinity, tertiary and quaternary structure forming interactions.

Correspondence to: Alexander Shekhtman.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The analysis of interaction proteomics requires in-cell methodologies because isolation techniques can easily disrupt protein complexes with limited thermodynamic stability[1, 2]. The interior of a living cell is an extraordinarily dense environment containing up to 400 mg/mL of macromolecules[3, 4]. Hydrophobic and hydrophilic interactions are perturbed, there is limited bulk water, cytosolic water forms solvent shells on protein surfaces[5] and the distances between macromolecules are less than the characteristic Debye screening length of ion charges[6]. Excluded volume effects and molecular crowding increase the effective concentrations of cytosolic species, virtually assuring the likelihood of transient low-affinity interactions. In-cell NMR spectroscopy[7] has proven to be a reliable technique to identify protein interaction surfaces under these conditions[8-11].

In-cell NMR spectroscopy provides atomic level resolution of molecular structures under physiological conditions by observing isotopically-labelled target proteins within the environment of a living cell, unlike *in vitro* NMR, which examines purified targets under idealized solution conditions. In-cell NMR also differs from *in vivo* NMR in two respects: first, targets are commonly over-expressed in or introduced to the cell at greater than physiological concentrations, although advances in the field now permit data collection at or near physiological levels; second, heterologous targets can be studied in different cellular environments.

In-cell NMR spectra are inherently noisier than spectra acquired *in vitro*. The low signal to noise ratio arises from the myriad of interactions between the target protein and components of the cytosol. These interactions can be specific, involving well-defined surfaces, or non-specific, resulting from the general aggregating properties of biological macromolecules, particularly when present in a heterologous environment[12, 13]. ^{15}N correlated heteronuclear single quantum coherence, $^1\text{H} \{^{15}\text{N}\}$ -HSQC, experiments typically used to study proteins *in vitro* may not be as informative when applied in-cell (Figure 1). Exceptions to this include non-interacting proteins and intrinsically disordered proteins, IDPs, which do not suffer significant loss of signal to noise and can provide structural details comparable to those obtained *in vitro*[14, 15].

The low signal to noise ratio results from transient, specific low-affinity quinary interactions, mediated primarily by cellular RNA and ribosomes, giving rise to high molecular weight complexes that exhibit slower tumbling times, a dramatic increase in the apparent molecular weight of the target and extensive broadening of many or all of the in-cell NMR peaks [16-20]. *In vitro* NMR studies of reconstituted systems have confirmed the role of RNA and ribosomes in protein quinary interactions[9, 19, 21]. Recent reviews summarize the salient features of in-cell NMR spectroscopy including labeling techniques, sample preparation, cell types, dynamics and specialized applications[16, 22-27].

The ability to study protein quinary interactions by using conventional in-cell NMR is seriously limited by the molecular size of cellular complexes[15, 28, 29]. To circumvent the limitations of molecular size, in-cell NMR can be performed on extensively deuterated, REDPRO[30], targets by using cross-correlated relaxation enhanced polarization transfer-based, CRINEPT[31], relaxation optimized ^{15}N -edited CRINEPT-heteronuclear multiple quantum coherence, HMQC, and transverse relaxation optimized spectroscopy, TROSY,

pulse sequences. These ^1H - ^{15}N CRINEPT-HMQC-TROSY experiments are relatively insensitive to magnetic field inhomogeneity[32] and provide superior sensitivity to NMR signals associated with slow-tumbling, large molecular weight complexes and are thus well-suited to elucidating protein quinary interactions[19] (Figure 1C).

Because the cellular content of RNA depends on growth conditions[33-36], protein quinary interactions have the potential to globally fine-tune diverse biological processes in a way that cannot be deduced from *in vitro* observations[19, 37, 38]. For example, in bacteria the ribosome concentration is linearly proportional to the growth rate and can reach tens of micromolar in rapidly dividing cells[39]. At this high concentration even a weak, *i.e.* micromolar, protein affinity for ribosomes will create an intracellular distribution of free and bound target proteins, each likely to have different activities. The net activity observed under these conditions will help refine models of signal transduction, regulation of cellular metabolism, and actions of drug-like compounds.

Specific high affinity interactions between a target and interactor protein are studied by using STINT-NMR [8, 40], which elucidates STructural INteractions between proteins within their native environment by using in-cell NMR [7, 40-43]. STINT-NMR identifies the interacting surface of a target protein with a single interactor protein (ligand) binds to it. This is accomplished in prokaryotes by over-expressing the target protein on [U - ^{15}N] medium, then sequentially over-expressing interactor protein on unlabeled medium. Samples are collected at various times and the in-cell NMR spectrum of the target protein is acquired. As the concentration of the interactor increases, the spectrum of the target changes to reflect the different chemical environment of the residues that have been affected by the binding interaction; the data provide a complete titration of the interaction. Analysis of spectral changes using Singular Value Decomposition, SVD, identifies the residues that comprise the interacting surfaces.

2. In-cell atomic resolution protein structures

To accurately identify the interaction surface of a target protein with an interactor molecule through changes in the NMR spectrum of a target molecule, it is imperative that the resonance assignments of the target protein be known beforehand, and that the target is stable in the absence of the interactor. Methods such as *in vitro* NMR, X-ray crystallography, and cryo-electron microscopy have classically provided atomic resolution structures of proteins outside of their native environment. The long sampling times required to collect in-cell NMR spectra makes *de novo* protein structural determination difficult. Target interactions with the cytosol can reduce NMR signals and spectral resolution, and temporal changes in the cellular environment can give rise to changes in protein conformation and dynamics thereby confounding spectral peak assignments. Advances in data acquisition and sampling, combined with high levels of target protein over-expression can overcome these problems to allow the structure of a protein to be determined at atomic level in the interior of a living cell.

The low sensitivity of in-cell protein NMR requires very long sampling times. Without fresh supplies of nutrients and gases, and waste removal, *E. coli* cells cannot survive for more than

a few hours. Much of the effort for solving structures in-cell has been devoted to reducing data acquisition times. Sakakibara *et al.* (2009)[44] employed a novel non-linear sampling scheme[45] to solve the structure of a putative heavy-metal binding protein from *T. thermophiles*, TTHA1718, in *E. coli*. Acquisition times for 3D spectra were reduced to 2-3 hours, as compared to 1-2 days required to collect data using traditional *in vitro* triple resonance experiments[46].

TTHA1718 was over-expressed in *E. coli* to 3-4 mM. To improve the signal to noise ratio, each 3D experiment was performed several times and the data sets were combined to enhance the protein signal. The 3D experiments were bracketed by a 2D $^1\text{H} \{^{15}\text{N}\}$ -HSQC experiment, which acted as a control to ensure that only data collected from intact cells were included in the combined data set. Each control spectrum following a 3D experiment was compared to the original reference HSQC spectra collected at the beginning of the experiment. If the control spectrum was significantly different from the reference spectrum, the previous 3D spectrum that was collected was not added to the composite data. The final calculated structure was similar to the structure determined *in vitro* for purified TTHA1718, and had a root mean squared deviation, RMSD, of 0.96 Å and a backbone RMSD of 1.16 Å from the *in vitro* structure (Figure 2).

Ikeya *et al.* (2016)[47] describes advances in methodology for in-cell protein structure determination of *Streptococcus* protein GB1 domain at near physiological concentrations with limited data sets of conformational restraints. The Quantitative Maximum Entropy, QME, method was used to process the indirect dimensions of 3D nonlinearly sampled data[43]. The fully automated structure determination of proteins in solution[48], FLYA, algorithm allowed the authors to supplement manual chemical shift assignments with the automated assignments obtained by using nuclear Overhauser effect spectroscopy, NOESY. Structure determination was accomplished by using the program CYANA[49, 50], enhanced by the Bayesian inference module, CYBAY, which facilitates the use of broad line shapes and low intensity peaks, which are common in in-cell NMR protein spectra, for structure determination.

Müntener *et al.* (2016)[51] developed a method for determining the structure of a protein in eukaryotic cells using a single set of 2D NMR spectra. They designed a metal ion chelator, DOTA-M7Py, that can be covalently bound to cysteines to introduce a paramagnetic tag onto the protein. Paramagnetic tags induce pseudocontact shifts, PCS, and residual dipolar couplings, RDC, which provide information about the distance and orientation of nuclei from the tag. Purified mutants of the B1 domain of Streptococcal protein G, were tagged with DOTA-M7Py[Tm]-GB1, and microinjected into *X. laevis* oocytes to an intracellular concentration of ~50 µM. A single 2D $^1\text{H} \{^{15}\text{N}\}$ -HSQC spectra was collected for each sample.

PCS and RDC data were input into GPS-Rosetta software. Libraries of 3- and 9-residue fragments were created and used to generate 10,000 structures; the 100 lowest energy conformations were compared with known GB1 structures determined by X-ray crystallography and solution NMR spectroscopy. A PCS-based weighting function was applied and the data subjected to energy minimization. Finally, RDC data was included into

the calculations to determine the final structure. During the analysis the median backbone C α RMSD decreased from 1.85 Å to 0.98 Å and 0.64 Å for the closest model and X-ray structure. The ability to acquire PCS and RDC data from 2D NMR experiments on samples at low in-cell concentration makes this methodology attractive and extremely useful for generating protein structures in-cell.

3. Structural Interactions

In-cell NMR spectroscopy provides invaluable information regarding the structural interactions of proteins inside live cells. Both high affinity and quinary interactions can be monitored to reveal how biological processes are conducted at atomic resolution.

3.1 STINT-NMR

To detect high affinity interactions by using in-cell NMR spectroscopy the interactor proteins have to be present in a stoichiometric ratio with the target protein so that the bulk of the labeled population is bound. In prokaryotes target and interactor proteins are over-expressed 10- to 100-fold greater than physiological concentrations and over-expression in eukaryotic cells can attain concentrations up to 100 μ M [26, 41]. Several methods have been developed to introduce ^{15}N labeled target proteins into both cells including cell-penetrating peptides [52], toxins [53], micro injections [54] [55], and electroporation [19, 56] in concentrations up to 50 μ M. Each method has its advantages and disadvantages, but all have proven effective in delivering sufficient quantities of labeled target protein to acquire reproducible in-cell NMR spectra.

Burz *et al.* [8] introduced a methodology called STructural INteractions NMR (STINT-NMR) that can be used to map protein interaction surfaces within the cellular environment. STINT-NMR uses selectively inducible expression plasmids to sequentially over-express target and interactor proteins. The target is first over-expressed on [U - ^{15}N] medium; the cells are then washed and over-expression of the interactor protein(s) is induced under non-labelling conditions, which renders the interactor protein cryptic. The uniform [U - ^{15}N] labeling strategy results in diminished background resonance peaks because bacterial growth is significantly reduced during protein over-expression. In-cell ^1H { ^{15}N }-HSQC spectra are collected before and during over-expression of the interactor [57].

As the concentration of the interactor increases, the spectrum of the target changes to reflect the different chemical environment of the residues that are affected by the binding interaction. The predominant spectral changes are reduced crosspeak intensities of surface residues, with chemical shift changes typically being very small. The residues involved in the binding interaction are identified from spectral changes between the free target and the final target following full over-expression of the interactor. The resulting NMR data provide a complete titration of the interaction and define structural details of the interacting surfaces at atomic resolution. Changes in chemical shifts and differential broadening of non-surface residues reflect rearrangements of secondary, tertiary or quaternary structural elements.

There are two important factors that need to be controlled when performing in-cell NMR. One is cell lysis, which allows labeled target to leak from the cells [58, 59] and limits the

acquisition time for in-cell NMR experiments. Protein leakage results in sharp signals from the target in the less viscous extracellular medium, thereby masking the broader signals obtained from protein in the cytosol. To ensure that NMR signals arise from proteins located inside the cell, following data acquisition, the cells are centrifuged and a $^1\text{H}\{^{15}\text{N}\}$ -HSQC spectrum of the sample supernatant is collected. If there is no leakage of target from the cells, the spectrum will show no protein NMR signals above noise levels.

The other factor is cell viability. Cells undergo changes in metabolism as they die, and these changes can affect the target spectrum. Cell viability is monitored by using a standard colony plating test in which the number of colonies grown on antibiotic selection plates inoculated with the in-cell sample before NMR spectroscopy is compared with the number of colonies grown using cells plated after acquiring experimental data. The cells are considered to be viable if the number of colonies on the plates are within 10% of each other[60].

In their original work, Burz *et al.* (2006)[8] described the use of STINT-NMR to identify the interacting surface of Ubiquitin, Ubq, for two interactors that contain a Ubq interacting motif[61, 62], UIM, in *E. coli*. The 28-amino-acid peptide from ataxin 3 protein, AUIM, and the signal transducing adapter molecule 2, STAM2, which contains a UIM and another Ubq binding region in the Vps27/Hrs/STAM, VHS, domain. Chemical shift changes were measured for ^{15}N -labelled Ubq by using $^1\text{H}\{^{15}\text{N}\}$ -HSQC NMR, as the intracellular concentration of unlabeled AUIM increased with time of over-expression. The largest changes arose exclusively from surface-exposed residues and were mapped onto the three dimensional structure of Ubq. The resulting perturbation map consisted of residues known to interact with sequences containing the UIM motif. The results were in agreement with *in vitro* STINT-NMR performed using purified components. The in-cell titration of ^{15}N -labelled Ubq with STAM2 produced uniform broadening of selected crosspeaks. The resulting surface perturbation map revealed two distinct interaction surfaces on Ubq. The additional interaction surface was proposed to underlie the greater affinity of STAM2 versus that of AUIM for Ubq[63].

The versatility of STINT-NMR was expanded to include in-cell biochemical modifications of interactor proteins in Burz *et al.*, (2008)[64]. They identified the binding surfaces on Ubq for non-phosphorylated and phosphorylated interactor proteins STAM2, and hepatocyte growth factor regulated substrate, Hrs, both of which contain a UIM, in *E. coli*. STAM2 and Hrs form homodimers and STAM2 and Hrs can form heterodimers. The technique involves sequential induction of a third plasmid coding for biochemical activity, in this case the constitutively active Src-family tyrosine kinase, Fyn, in addition to expression plasmids for target and interactor proteins. STINT-NMR was performed on ^{15}N -labelled Ubq and unlabeled STAM2 and/or Hrs interactor proteins, followed by over-expression of Fyn, which resulted in phosphorylation of sites on STAM2 and Hrs. The Ubq surface map revealed additional interaction surfaces upon phosphorylation (Figure 3).

Luchinat *et al.*[65] described a system for selectively labeling target proteins in eukaryotic cells. HEK293T cells were stably-transfected with the human copper binding protein HAH1, and transiently transfected with vectors expressing human superoxide dismutase 1, SOD1.

HAH1 was first over-expressed in unlabeled medium, followed by over-expression of [U - ^{15}N]-SOD1. The ^1H { ^{15}N }-SOFAST-HMQC[66] spectra of SOD1 in HEK293T cells revealed undesired labeling of HAH1 had occurred during over-expression of the second protein. To minimize labelling of HAH1 during over-expression of the second protein, silencing vectors containing short hairpin RNA, shRNA, were co-transfected into HEK293T cells. The silencing vectors repressed constitutive production of HAH1 through RNA interference and allowed for selective labelling of SOD1. This technique demonstrated that specific silencing of a stably expressed protein is a successful strategy to obtain selective isotope labeling of only one protein. This work enabled the study of high affinity interactions between proteins in eukaryotic cells, thus expanding the repertoire of in-cell NMR methodology.

3.2 Quinary interactions

In-cell spectra of target molecules exhibit broadened crosspeaks and megadalton apparent molecular weights due to transient, specific low-affinity quinary interactions with components of the cytosol (Figure 1). In cells, transverse proton relaxation, T_2 , is so short for high molecular weight complexes that only proteins that are unstructured, and thus have a longer T_2 due to internal flexibility or that interact extremely weakly with the cytosol, can be studied by using two dimensional solution NMR such as ^1H { ^{15}N }-HSQC[46]. To overcome the limitations imposed by high molecular weight complexes ^1H - ^{15}N CRINEPT-HMQC-TROSY experiments[31, 67] are employed to optimize the effective transverse relaxation of amide proton signals (Figure 1).

Majumder *et al.* (2015)[19] used ^1H - ^{15}N CRINEPT-HMQC-TROSY experiments to map the quinary interaction surfaces of REDPRO-labeled, [U - ^{15}N]-thioredoxin, FKBP, adenylate kinase and ubiquitin, and showed that the underlying cause for peak broadening was due to target binding to cellular RNA in both *E. coli* and HeLa cells. Target protein in-cell spectra were compared to *in vitro* and/or cell lysate spectra to identify the residues that comprise the quinary interacting surface. To achieve maximal spectroscopic sensitivity, the optimal CRIP transfer delay, T_{opt} , was measured *in vitro* using glycerol-solutions (Figure 4).

The in-cell spectrum of thioredoxin, Trx, exhibits broad crosspeaks at positions close to those observed in cell lysates and a resolved apparent molecular weight of ~ 1 MDa. Some crosspeaks exhibited two multiplets, corresponding to fast and slow transverse components of the CRINEPT-HMQC crosspeaks, indicating that free Trx is in exchange with a large cytosolic complex. Mapping the interaction surface residues onto the molecular structure of Trx revealed that residues in and about the Trx active site are involved in quinary interactions samples (Figure 4).

Adenylate kinase, ADK, exists in substrate-free, open, and substrate-bound, closed, conformations. The in-cell spectrum of adenylate kinase, ADK, has a subset of partially or completely broadened crosspeaks that resembles the open, substrate-free conformation seen *in vitro*. These crosspeaks correspond to residues in the CORE domain, proximal to the AMP binding site. Unlike Trx, the active site of ADK is unaffected by quinary interactions and is free to bind substrates ATP and AMP. Residue crosspeaks involved in domain closing

are unchanged, suggesting that quinary interactions do not perturb the tertiary structure of the protein.

The interaction of Trx, FKBP, ADK and ubiquitin were examined in the presence of preparations of total prokaryotic and eukaryotic cellular RNA *in vitro*. Adding 10 µg/mL of total RNA to 15 µM REDPRO-labelled FKBP led to the complete disappearance of the NMR spectrum, while 20 µg/mL of total RNA was sufficient to render the NMR spectrum of 10 µM REDPRO-labelled ADK comparable to its in-cell spectrum, suggesting strong interactions with RNA. In comparison, 30 mg/mL of total RNA was required for the spectra of 15 µM Trx and 15 µM Ubq to resemble the corresponding in-cell spectra, suggesting that the strength of the quinary interactions are considerably weaker for these two proteins. Controls in which the Trx and total cellular RNA mixture was treated with RNase confirmed that short RNA oligonucleotides interact specifically with Trx, and that the quinary interaction surface is specific. This work produced an *in vitro* method for interrogating in-cell-type quinary interactions.

3.3 SVD analysis of protein-protein interactions

Analyses of in-cell NMR spectra of target proteins interacting with physiological partners are often complicated by low signal to noise ratios, and long over-expression times may lead to changes in the in-cell spectra over the course of the experiment. The process of distinguishing which spectral changes are due to specific binding generally considers the differences between only the spectrum of free target and the final target spectrum following full over-expression of the interactor to assess the change in intensity of a given crosspeak resonance[57]. This difference is used to infer whether or not the corresponding amino acid contributes to the principal binding mode of the target. Because individual signals change at different rates as the concentration of the interactor protein increases (Figure 5A), this analysis ignores changes in the in-cell spectra that arise over the time course of the experiment and from non-interactor and concentration-independent binding processes. This problem can be overcome by analyzing in-cell NMR data using Singular Value Decomposition, SVD.

SVD is a mathematical technique used to identify the principal components of an arbitrary matrix that contribute maximally to the variance of its elements[68]. Over the course of a STINT-NMR titration, a series of in-cell NMR spectra are collected, and a matrix, *M*, is created that contains the changes in target protein crosspeak intensities or chemical shifts versus the expression time of the unlabeled binding partner (Figure 5B). SVD analysis of matrix *M* discriminates between changes in the in-cell NMR spectra of the target protein due to specific and non-specific binding interactions and changes due to the presence of the complex cellular environment over the time course of interactor over-expression. The analysis objectively identifies the amino acid residues involved in the principal binding mode of a target protein with its interactor.

Majumder *et al.* (2014)[69] used SVD to analyze STINT-NMR data for the interaction between prokaryotic ubiquitin-like protein, Pup, and mycobacterial proteasome ATPase, Mpa. To minimize the signal to noise ratio, crosspeak intensities in the in-cell NMR spectra of free [*L*-¹⁵N] Pup and [*L*-¹⁵N] Pup-Mpa complexes were scaled to define matrix *M*.

Results of the analysis are presented as a Scree plot (Figure 5C), which shows the distribution of singular values that defines the relative contribution of each binding mode to the change in intensities or chemical shifts (Figure 5D). An abrupt drop in the singular values following the first binding mode indicates the presence of a single principal binding mode; a gradual decrease in singular values corresponds to random binding of target protein to components of the cytosol. The analysis indicated the existence of one principal binding mode and implicated a set of amino acids involved in the specific interaction that differed from previous NMR analyses but were in good agreement with crystallographic data.

3.4 Intrinsically disordered proteins

The NMR spectra of intrinsically disordered proteins, IDPs, do not undergo extensive broadening in-cell, as a result each resonance peak is generally observable. Measuring spin relaxation parameters for individual residues presents a rigorous methodology for identifying quinary interactions between IDPs and components of the cytosol for individual residues.

Hough *et al.* (2015)[70] measured spin relaxation parameters to identify quinary interactions for segments of the phenylalanyl-glycyl-repeat-rich nucleoporin, FG Nup[71], derived from Nsp1 in *E. coli*. The N-terminal segment, FG-N[72], is Asn-rich and contains irregularly spaced FG repeats, and the central segment, FSFG-K[73], is Lys-rich and has regularly spaced FG repeats. FG Nups are IDPs that are implicated in the transport of transcription factors across the nuclear membrane. *In vitro* FG-N forms a high molecular weight hydrogel-like material that is not observed in-cell or in cell lysates or mimics. By measuring spin relaxation parameters, R_1 , R_2 , nOe , for [U- ^{15}N]-FG-N and [U- ^{15}N]-FSFG-K, they observed large increases in the R_2/R_1 ratio of phenylalanine and adjacent residues of each repeat indicating interactions with components of the cytosol while the inter FG regions remained largely unaffected (Figure 6). These quinary interactions stabilize the IDPs in a disordered state and inhibit the intramolecular interactions that promote aggregation. The results allowed the authors to propose a model for transport of transcription factors through the nuclear envelope.

Theillet *et al.* (2016)[56] showed that the protein dynamics of the IDP α -synuclein, α Syn, are independently modulated by weak hydrophobic and electrostatic quinary interactions affecting the terminal regions of the protein. They electroporated ^{15}N -labeled non-acetylated human α Syn into non-neuronal A2780 and neuronal SK-N-SH cells and used 1H { ^{15}N }-SOFAS-HMQC to identify signal broadening for N-terminal residues that are known to be acetylated *in vivo*, residues around Tyr39 and the C-terminus of the protein. The broadening was not observed in cell lysates or *in vitro*. Analysis of backbone amide relaxation parameters, R_1 , R_2 and nOe , in-cell and in artificially crowded solutions containing BSA or lysozyme explained the selective signal broadening of N- or C- terminal signals. Longer rotational correlation times indicated a uniform decrease in mobility in A2780 and SK-N-SH cells consistent with viscosity-driven forces. *In vitro* lysozyme-induced signal broadening was reduced with increasing salt, implicating electrostatic interactions in mediating the C-terminal interactions. Mutating α Syn residues Phe4 and Tyr39 to alanines and re-measuring the exchange dynamics in A2780 and SK-N-SH cells and *in vitro* in the presence of BSA

implicated hydrophobic interactions in mediating N-terminal residue signal broadening independent of C-terminal interactions. The results showed that intrinsic structural disorder can be maintained in mammalian cells and that different intracellular environments can mediate distinct interactions.

4. Functional Interactions

Functional interactions are inexorably linked to structural interactions. Live cells used in in-cell NMR spectroscopy can be manipulated by applying various stimuli and changes in growth media to evoke physiological changes. The influence of cellular changes on protein-protein interactions can be subsequently observed by in-cell NMR spectroscopy.

Majumder *et al.* [9] identified quinary interactions for ubiquitin, Ubq in *P. pastoris* (yeast). Ubq overexpressed from the methanol-induced *AOX1* promoter is localized in the cytosol and exhibits a well-resolved NMR spectrum[42]. Over-expressed Ubq grown on a mixed carbon source containing methanol and dextrose is partitioned into cytosolic vesicles and has a uniformly broadened in-cell $^1\text{H} \{^{15}\text{N}\}$ -HSQC spectrum[42]. RNA isolated from *P. pastoris* grown in buffered minimal methanol, RNA_{BMM}, contains a large amount of pre-processed mRNA and pre-processed large ribosomal subunit whereas these species are largely absent in RNA from cells grown using buffered minimal dextrose methanol, RNA_{BMDM}[9].

In vitro experiments confirmed the role of RNA in mediating the interaction with Ubq. Adding 30 mg/mL of RNA isolated from *P. pastoris* grown in BMM to purified [U - ^{15}N]-Ubq broadened a subset of crosspeaks, indicating a specific quinary interaction between Ubq and RNA. Adding 30 mg/mL of RNA isolated from cells grown in BMDM to purified [U - ^{15}N]-Ubq resulted in the complete disappearance of crosspeaks in the $^1\text{H} \{^{15}\text{N}\}$ -HSQC spectrum suggesting an increase in quinary interactions between Ubq and RNA. The results showed that the total cellular content of the cells was responsible for the cellular redistribution of Ubq indicated by the observed spectral changes, and, more broadly, that total cellular RNA may regulate biological activity.

DeMott *et al.* (2017)[10] found that the in-cell ^1H - ^{15}N CRINEPT-HMQC-TROSY spectrum of REDPRO labeled [U - ^2D , ^{15}N]-ADK identified specific quinary interaction surfaces that do not affect the substrate binding sites, and investigated the possibility that quinary interactions may affect protein activity. *In vitro*, 1 μM ribosomes resulted in selective broadening and chemical shift changes in the [U - ^2D , ^{15}N]-ADK spectrum that resemble those observed in-cell; the spectrum was completely broadened in the presence of 1 mg/mL of total RNA. The spectra are characteristic of intermediate exchange and imply a dissociation constant of $\sim 10 \mu\text{M}$ for the quinary interaction between ADK and ribosomes. ADK binding to the ribosome was confirmed by fluorescent titrations *in vitro* and yielded an apparent K_d of 3.7 μM .

The in-cell NMR spectrum of ADK is consistent with a nucleotide-free or open configuration[19]. Adding chloramphenicol to the growth medium increases the cellular production of ATP[74] and shifts the in-cell NMR spectrum of ADK from the open to closed, ATP-bound, configuration[19]. To test if ATP binding to ribosomes could reduce the

concentration of free ATP in *E. coli* to maintain ADK in the open configuration, they used *in vitro* ^1H $\{^3\text{P}\}$ -correlation NMR experiments[75] to resolve an apparent K_d of 6 μM for the binding of a non-cleavable ATP analog, AMP-PCP, to ribosomes.

The effect of ribosomes on ADK kinetics was evaluated *in vitro* by titrating ATP into purified 10 μM [U - ^2D , ^{15}N] ADK. The ^1H - ^{15}N CRINEPT-HMQC-TROSY spectra showed systematic changes in chemical shifts and concomitant broadening as the concentration of ATP was increased. At 80 μM ATP in the presence of 1 μM ribosome the spectrum of the ATP-bound ribosome-ADK complex resembled that of 40 μM ATP-bound ADK suggesting that ATP binding to the ribosome reduced the concentration of free ATP available for binding. *In vitro* kinetic assays performed in the absence and presence of ribosomes and found that the ribosome acted as an uncompetitive inhibitor of ADK activity (Figure 7A). Similar experiments performed using [U - ^2D , ^{15}N] dihydrofolate reductase, DHFR, showed that DHFR engages in micromolar binding affinity quinary interactions with ribosomes, and that its coenzyme, NADPH, binds to ribosomes with a K_d of ~ 4.5 μM [10]. The results showed that the ribosome acts as a competitive inhibitor of DHFR activity (Figure 7B).

DHFR and thymidylate synthase, TS, are functionally linked in the thymidylate synthetic pathway, and TS interacts with total cellular *E. coli* RNA, mRNA[76] and ribosomes[77]. DeMott *et al.* (2017), examined the possibility that ribosome-mediated quinary interactions also affect TS activity. *In vitro* TS kinetic activity increased with increasing concentration of ribosomes, the kinetic profile was characteristic of uncompetitive substrate inhibition (Figure 7C). Control experiments confirmed the specificity of these ribosome-binding interactions. This study revealed that ribosome-mediated quinary interactions are likely to play an as yet unrealized role in regulating metabolic enzyme activities.

Banci *et al.*[78] showed that co-expression of thioredoxin 1, Trx1, or glutaredoxin 1, Grx1, along with mitochondrial intermembrane space import and assembly protein 40, Mia40, changes the redox state of Mia40. Mia40 maturation requires crossing the outer mitochondrial membrane in a reduced, unfolded conformation. In HEK293T cells 85% of Mia40 is found to be in an oxidized, folded state as determined by using ^1H -NMR. Increasing Grx1 co-expression decreased oxidized Mia40 to 25%; co-expressing Trx1 to comparable levels decreased oxidized Mia40 levels to 50%. Transient expression was accomplished by transfecting HEK293T cells with varying amounts of DNA for each gene and grown under [U - ^{15}N]-labelling conditions. The redox state was monitored by using ^1H – ^{15}N SOFAST-HMQC NMR spectroscopy. A comparable effect of Grx1 or Trx1 co-expression on cytochrome c oxidase copper chaperone, Cox17, was observed by Mercatelli *et al.* (2016)[79]. The effects occur at sub-stoichiometric amounts of Grx1 and Trx1, implying a catalytic role for each. Grx1 and Trx1 crosspeaks resulting from co-expression with Mia40 were broadened, due to presumed interactions with cytosolic components, to such an extent as to not be detectable by NMR in-cell and to not interfere with the signals from ^{15}N -Mia40.

Banci *et al.*[41] monitored the maturation of human superoxide dismutase 1, SOD1, in the absence and presence of copper chaperone for SOD1, CCS in HEK293T cells. Maturation of SOD1 entails a sequence of steps including zinc binding, homodimerization, copper binding,

and formation of an intrasubunit disulfide bond. HEK293T cells were transiently infected with varying ratios of DNA/PEI to maximize expression of each protein. ^1H NMR and ^1H - ^{15}N SOFAST-HMQC NMR spectroscopy were used to determine the structural state of the protein. In cells without metal supplements SOD1 exists in two forms, a monomeric metal-free species, apo-SOD1, and a dimeric species with one Zn^{2+} ion bound to each monomer, E,Zn-SOD1, neither of which contained the intrasubunit disulfide bridge. The ion-bound complex was likely due to residual zinc in the culture medium. Adding Zn^{2+} to the growth medium resulted in a spectrum consistent with only E,Zn-SOD1 present in the cells. *In vitro* or in lysates a mixture of apo-SOD1, E,Zn-SOD1 and Zn,Zn-SOD1 was observed implying that selective zinc binding occurs only in the cellular environment. The addition of Cu^{2+} to the growth medium resulted in ~25% Cu(I),Zn-SOD1, 75% E,Zn-SOD1 and ~20% intrasubunit disulfide bond formation.

Co-expression and labeling of SOD1 and CCS in HEK293T cells in growth medium containing zinc yielded dimeric E,Zn-SOD1. CCS remained cryptic in these experiments due to crosspeak broadening. The intrasubunit disulfide bond was present at ~50%, consistent with CCS-mediated disulfide oxidation in the absence of copper binding. The addition of Cu^{2+} to the growth medium resulted in the formation of Cu(I), Zn-SOD1 and complete oxidation of the disulfide bond. The concentration dependence of SOD1 maturation was examined by altering the expression levels of SOD1 and CCS and found to limit the stoichiometry of each species but not the maturation sequence.

Mercatelli *et al.* (2016)[79] compared the distributions of unfolded and folded Cox17, Mia40 and SOD1, containing reduced oxidized disulfide bonds respectively, in HEK293T and *E. coli* cells with the distribution predicted from the intracellular glutathione, GSH, redox potential and the Nernst equation. HEK293T cells were transiently infected with varying ratios of DNA/PEI to maximize expression of [^{15}N]-cysteine Cox17, [^{15}N]-Cys Mia40, [U - ^{15}N]-E,Zn-SOD1 and [^{15}N]-Cys Cu,Zn-SOD1. The ratios of folded to total Cox17 and Mia40 were determined from [^{15}N]-Cys intensities using ^1H - ^{15}N SOFAST NMR spectroscopy. The results showed 70% of Cox17 and 75% of Mia40 were in the folded, oxidized state. The results were higher than predicted favoring the oxidized state. Co-expression of Grx1 or Trx1 altered the distribution to that predicted by the Nernst equation indicating that these proteins establish a thermodynamic equilibrium between Cox17, Mia40 and the GSH pool.

The oxidized to reduced ratio of [U - ^{15}N]-E,Zn-SOD1 and [^{15}N]-cysteine Cu,Zn-SOD1 in HEK293T cells was obtained by integrating crosspeaks in the ^1H - ^{15}N SOFAST NMR spectra corresponding to the oxidized and the reduced species. In the absence of co-expressed CCS, 85% of E,Zn-SOD1 was found to be reduced. CCS binding to SOD1 forms a disulfide bridge between Cys244 of CCS and Cys 57 of SOD1 that is transferred to form the intrasubunit Cys57-Cys146 in SOD1. Co-expressing CCS in the presence of excess Cu^{2+} resulted in 80% oxidized Cu,Zn-SOD1. In bacterial cells zinc binding and homo-dimerization of [U - ^{15}N]-SOD1 occurred spontaneously, and copper binding occurred when Cu^{2+} was added to the growth medium. In BL21 (DE3) cells [U - ^{15}N]-E,Zn-SOD1 was found to be mostly reduced and [U - ^{15}N]-Cu,Zn-SOD1 was 60% oxidized. In Origami B cells E,Zn-SOD1 was 85% oxidized and Cu,Zn-SOD1 was 100% oxidized. SOD1 never

attained thermodynamic equilibrium with the GSH pool. The CCS-SOD1 interaction skews the distribution of SOD1 redox states further from thermodynamic equilibrium implying that disulfide exchange between these species is uncoupled from the GSH pool, and that the redox potential of CCS is not driven by the GSH redox pathway.

In *E. coli* BL21 cells, ^1H - ^{15}N SOFAST NMR spectroscopy revealed that individually over-expressed [U - ^{15}N]-Cox17, Mia40 and SOD1 were almost completely in the unfolded, reduced state, and SOD1 is partially oxidized in copper treated cells. In Origami B, which lack glutathione and thioredoxin reductases, and is genetically designed to facilitate folding of disulfide-containing proteins ~45% of Cox17 and virtually all of Mia40 and SOD1 were found to exist in the folded, oxidized state. These measurements were performed in lysates due to the extreme broadening of the folded domain signal in bacterial cells. Thus, in bacteria the distribution of the redox states for these proteins is not in equilibrium with the GSH pool, and is probably controlled by the thioredoxin system, which does not function in Origami B. An alternative explanation is that the midpoint potential of GSH in *E. coli* is lower than reported.

In summary, in-cell NMR provides a rich opportunity to study proteins engaged in functionally important interactions at atomic resolution under close to physiological conditions.

5. Conclusions

In the past ten years, in-cell NMR spectroscopy has been developed to reliably obtain atomic resolution information about protein interactions in live cells[8, 41, 44, 52, 56]. In-cell NMR allows us to solve the atomic structure of an individual protein inside the complex cellular milieu and analyze the interactions of a protein of interest with the rest of the proteome. The technology can be applied to any protein that is observable by in-cell NMR spectroscopy in either prokaryotic or eukaryotic cells. The use of NMR spectroscopy to study systems of interest by using both low sensitivity/high information-content in-cell techniques and high sensitivity/low information-content *in vitro* techniques provides a means to achieve a detailed understanding of functional interactions. This knowledge is critical to comprehend in-cell basic biochemistry and to develop effective modulators of biochemical interactions for therapeutic applications.

References

1. Wirth AJ, Gruebele M. Quinary protein structure and the consequences of crowding in living cells: leaving the test-tube behind. *Bioessays*. 2013; 35(11):984–93. [PubMed: 23943406]
2. Sarkar M, Smith AE, Pielak GJ. Impact of reconstituted cytosol on protein stability. *Proc Natl Acad Sci U S A*. 2013; 110(48):19342–7. [PubMed: 24218610]
3. Goodsell D. The machinery of life. Springer; New York: 2009.
4. Milo R, Phillips R. Cell Biology by the numbers. 1st. Garland Science; New York, NY, USA: 2016.
5. Persson E, Halle B. Cell water dynamics on multiple time scales. *Proc Natl Acad Sci U S A*. 2008; 105(17):6266–71. [PubMed: 18436650]
6. Benedek GB, Villars FMH. Physics: With Illustrative Examples from Medicine and Biology: Electricity and Magnetism. Springer-Verlag; New York: 2000.

7. Serber Z, Dotsch V. In-cell NMR spectroscopy. *Biochemistry*. 2001; 40(48):14317–23. [PubMed: 11724542]
8. Burz DS, Dutta K, Cowburn D, Shekhtman A. Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). *Nat Methods*. 2006; 3(2):91–3. [PubMed: 16432517]
9. Majumder S, DeMott CM, Reverdatto S, Burz DS, Shekhtman A. Total Cellular RNA Modulates Protein Activity. *Biochemistry*. 2016; 55(32):4568–73. [PubMed: 27456029]
10. DeMott CM, Majumder S, Burz DS, Reverdatto S, Shekhtman A. Ribosome mediated quinary interactions modulate in-cell protein activities. *Biochemistry*. 2017; 56(32):4117–4126. [PubMed: 28715177]
11. Wu H, Finger LD, Feigon J. Structure determination of protein/RNA complexes by NMR. *Methods Enzymol*. 2005; 394:525–45. [PubMed: 15808236]
12. Smith AE, Zhang Z, Pielak GJ, Li C. NMR studies of protein folding and binding in cells and cell-like environments. *Curr Opin Struct Biol*. 2015; 30:7–16. [PubMed: 25479354]
13. Li C, Pielak GJ. Using NMR to distinguish viscosity effects from nonspecific protein binding under crowded conditions. *J Am Chem Soc*. 2009; 131(4):1368–9. [PubMed: 19140727]
14. 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 (IDPs). *Chem Rev*. 2014; 114(13):6661–714. [PubMed: 24901537]
15. Luh LM, Hansel R, Lohr F, Kirchner DK, Krauskopf K, Pitzius S, Schafer B, Tufar P, Corbeski I, Guntert P, Dotsch V. Molecular crowding drives active pin1 into nonspecific complexes with endogenous proteins prior to substrate recognition. *J Am Chem Soc*. 2013; 135(37):13796–803. [PubMed: 23968199]
16. Maldonado AY, Burz DS, Shekhtman A. In-cell NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc*. 2011; 59(3):197–212. [PubMed: 21920217]
17. Li C, Wang GF, Wang Y, Creager-Allen R, Lutz EA, Scronce H, Slade KM, Ruf RA, Mehl RA, Pielak GJ. Protein (19)F NMR in *Escherichia coli*. *J Am Chem Soc*. 2010; 132(1):321–7. [PubMed: 20050707]
18. Crowley PB, Chow E, Papkovskaia T. Protein interactions in the *Escherichia coli* cytosol: an impediment to in-cell NMR spectroscopy. *Chembiochem*. 2011; 12(7):1043–8. [PubMed: 21448871]
19. Majumder S, Xue J, DeMott CM, Reverdatto S, Burz DS, Shekhtman A. Probing protein quinary interactions by in-cell nuclear magnetic resonance spectroscopy. *Biochemistry*. 2015; 54(17):2727–38. [PubMed: 25894651]
20. Breindel L, DeMott C, Burz DS, Shekhtman A. Real-Time In-Cell Nuclear Magnetic Resonance: Ribosome-Targeted Antibiotics Modulate Quinary Protein Interactions. *Biochemistry*. 2018; doi: 10.1021/acs.biochem.7b00938
21. Kyne C, Ruhle B, Gautier VW, Crowley PB. Specific ion effects on macromolecular interactions in *Escherichia coli* extracts. *Protein Sci*. 2015; 24:310–318. [PubMed: 25492389]
22. Freedberg DI, Selenko P. Live cell NMR. *Annu Rev Biophys*. 2014; 43:171–92. [PubMed: 24895852]
23. Luchinat E, Banci L. A Unique Tool for Cellular Structural Biology: In-cell NMR. *J Biol Chem*. 2016; 291(8):3776–84. [PubMed: 26677229]
24. Burz DS, DeMott CM, Aldousary A, Dansereau S, Shekhtman A. Quantitative Determination of Interacting Protein Surfaces in Prokaryotes and Eukaryotes by Using In-Cell NMR Spectroscopy. *Methods Mol Biol*. 2018; 1688:423–444. [PubMed: 29151221]
25. Serber Z, Selenko P, Hansel R, Reckel S, Lohr F, Ferrell JE Jr, Wagner G, Dotsch V. Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. *Nat Protoc*. 2006; 1(6):2701–9. [PubMed: 17406526]
26. Barbieri L, Luchinat E, Banci L. Characterization of proteins by in-cell NMR spectroscopy in cultured mammalian cells. *Nat Protoc*. 2016; 11(6):1101–11. [PubMed: 27196722]
27. Felli IC, Gonnelli L, Pierattelli R. In-cell (1)(3)C NMR spectroscopy for the study of intrinsically disordered proteins. *Nat Protoc*. 2014; 9(9):2005–16. [PubMed: 25079425]

28. Harada R, Tochio N, Kigawa T, Sugita Y, Feig M. Reduced native state stability in crowded cellular environment due to protein-protein interactions. *J Am Chem Soc.* 2013; 135(9):3696–701. [PubMed: 23402619]
29. Ye Y, Liu X, Zhang Z, Wu Q, Jiang B, Jiang L, Zhang X, Liu M, Pielak GJ, Li C. (19) F NMR spectroscopy as a probe of cytoplasmic viscosity and weak protein interactions in living cells. *Chemistry.* 2013; 19(38):12705–10. [PubMed: 23922149]
30. Shekhtman A, Ghose R, Goger M, Cowburn D. NMR structure determination and investigation using a reduced proton (REDPRO) labeling strategy for proteins. *FEBS Lett.* 2002; 524(1-3):177–82. [PubMed: 12135763]
31. Riek R, Wider G, Pervushin K, Wuthrich K. Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proc Natl Acad Sci U S A.* 1999; 96(9):4918–23. [PubMed: 10220394]
32. Wider G. NMR techniques used with very large biological macromolecules in solution. *Methods Enzymol.* 2005; 394:382–98. [PubMed: 15808229]
33. Darnell J. *RNA: Life's Indispensable Molecule*. 1st. Laboratory Press; Cold Spring Harbor, New York, NY, USA: 2011.
34. Roeder RG. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci.* 1996; 21(9):327–35. [PubMed: 8870495]
35. Pardee AB. Multiple molecular levels of cell cycle regulation. *J Cell Biochem.* 1994; 54(4):375–8. [PubMed: 8014185]
36. Waldron C, Lacroute F. Effect of growth rate on the amounts of ribosomal and transfer ribonucleic acids in yeast. *J Bacteriol.* 1975; 122(3):855–65. [PubMed: 1097403]
37. Barbieri L, Luchinat E, Banci L. Protein interaction patterns in different cellular environments are revealed by in-cell NMR. *Sci Rep.* 2015; 5:14456. [PubMed: 26399546]
38. Kyne C, Crowley PB. *Grasping the Nature of the Cell Interior: From Physiological Chemistry to Chemical Biology*. FEBS J. 2016
39. Milo R, Philips R. *Cell Biology by the Numbers*. 1st. Garland Science; New York, USA: 2015.
40. Burz DS, Shekhtman A. The STINT-NMR method for studying in-cell protein-protein interactions. *Curr Protoc Protein Sci.* 2010; Chapter 17 Unit 17 11.
41. Banci L, Barbieri L, Bertini I, Luchinat E, Secci E, Zhao Y, Aricescu AR. Atomic-resolution monitoring of protein maturation in live human cells by NMR. *Nat Chem Biol.* 2013; 9(5):297–9. [PubMed: 23455544]
42. Bertrand K, Reverdatto S, Burz DS, Zitomer R, Shekhtman A. Structure of proteins in eukaryotic compartments. *J Am Chem Soc.* 2012; 134(30):12798–806. [PubMed: 22758659]
43. Hamatsu J, O'Donovan D, Tanaka T, Shirai T, Hourai Y, Mikawa T, Ikeya T, Mishima M, Boucher W, Smith BO, Laue ED, Shirakawa M, Ito Y. High-resolution heteronuclear multidimensional NMR of proteins in living insect cells using a baculovirus protein expression system. *J Am Chem Soc.* 2013; 135(5):1688–91. [PubMed: 23327446]
44. Sakakibara D, Sasaki A, Ikeya T, Hamatsu J, Hanashima T, Mishima M, Yoshimasu M, Hayashi N, Mikawa T, Walchli M, Smith BO, Shirakawa M, Guntert P, Ito Y. Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature.* 2009; 458(7234):102–5. [PubMed: 19262674]
45. Ikeya T, Sasaki A, Sakakibara D, Shigemitsu Y, Hamatsu J, Hanashima T, Mishima M, Yoshimasu M, Hayashi N, Mikawa T, Nietlispach D, Walchli M, Smith BO, Shirakawa M, Guntert P, Ito Y. NMR protein structure determination in living *E. coli* cells using nonlinear sampling. *Nat Protoc.* 2010; 5(6):1051–60. [PubMed: 20539281]
46. Cavanagh JF, J W, Palmer AG, Rance M, Skelton NJ. *Protein NMR spectroscopy*. Academic Press: New York: 2007.
47. Ikeya T, Hanashima T, Hosoya S, Shimazaki M, Ikeda S, Mishima M, Guntert P, Ito Y. Improved in-cell structure determination of proteins at near-physiological concentration. *Sci Rep.* 2016; 6:38312. [PubMed: 27910948]
48. Schmidt E, Guntert P. A new algorithm for reliable and general NMR resonance assignment. *J Am Chem Soc.* 2012; 134(30):12817–29. [PubMed: 22794163]

49. Guntert P. Automated NMR structure calculation with CYANA. *Methods Mol Biol.* 2004; 278:353–78. [PubMed: 15318003]
50. Guntert P. Automated structure determination from NMR spectra. *Eur Biophys J.* 2009; 38(2):129–43. [PubMed: 18807026]
51. Muntener T, Haussinger D, Selenko P, Theillet FX. In-Cell Protein Structures from 2D NMR Experiments. *J Phys Chem Lett.* 2016; 7(14):2821–5. [PubMed: 27379949]
52. 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(7234):106–9. [PubMed: 19262675]
53. Ogino S, Kubo S, Umemoto R, Huang S, Nishida N, Shimada I. Observation of NMR Signals from Proteins Introduced into Living Mammalian Cells by Reversible Membrane Permeabilization Using a Pore-Forming Toxin, Streptolysin O. *J Am Chem Soc.* 2009; 131(31):10834–5. [PubMed: 19603816]
54. Selenko P, Serber Z, Gadea B, Ruderman J, Wagner G. Quantitative NMR analysis of the protein G B1 domain in *Xenopus laevis* egg extracts and intact oocytes. *Proc Natl Acad Sci U S A.* 2006; 103(32):11904–9. [PubMed: 16873549]
55. Sakai T, Tochio H, Tenno T, Ito Y, Kokubo T, Hiroaki H, Shirakawa M. In-cell NMR spectroscopy of proteins inside *Xenopus laevis* oocytes. *J Biomol NMR.* 2006; 36(3):179–88. [PubMed: 17031531]
56. Theillet FX, Binolfi A, Bekei B, Martorana A, Rose HM, Stuiver M, Verzini S, Lorenz D, van Rossum M, Goldfarb D, Selenko P. Structural disorder of monomeric alpha-synuclein persists in mammalian cells. *Nature.* 2016; 530(7588):45–50. [PubMed: 26808899]
57. Burz DS, Dutta K, Cowburn D, Shekhtman A. In-cell NMR for protein-protein interactions (STINT-NMR). *Nat Protoc.* 2006; 1(1):146–52. [PubMed: 17406226]
58. Cruzeiro-Silva C, Albernaz FP, Valente AP, Almeida FC. In-Cell NMR spectroscopy: inhibition of autologous protein expression reduces *Escherichia coli* lysis. *Cell Biochem Biophys.* 2006; 44(3):497–502. [PubMed: 16679537]
59. Barnes CO, Pielak GJ. In-cell protein NMR and protein leakage. *Proteins.* 2011; 79(2):347–51. [PubMed: 21120863]
60. Maniatis T, Fritsch EF, Sambrook J. *Molecular cloning : a laboratory manual.* Cold Spring Harbor Laboratory; Cold Spring Harbor, N.Y.: 1982.
61. Shekhtman A, Cowburn D. A ubiquitin-interacting motif from Hrs binds to and occludes the ubiquitin surface necessary for polyubiquitination in monoubiquitinated proteins. *Biochemical and Biophysical Research Communications.* 2002; 296(5):1222–1227. [PubMed: 12207904]
62. Di Fiore PP, Polo S, Hofmann K. When ubiquitin meets ubiquitin receptors: a signalling connection. *Nat Rev Mol Cell Biol.* 2003; 4(6):491–7. [PubMed: 12778128]
63. Mizuno E, Kawahata K, Kato M, Kitamura N, Komada M. STAM proteins bind ubiquitinated proteins on the early endosome via the VHS domain and ubiquitin-interacting motif. *Mol Biol Cell.* 2003; 14(9):3675–89. [PubMed: 12972556]
64. Burz DS, Shekhtman A. In-cell Biochemistry using NMR spectroscopy. *PloS One.* 2008; 3(7):e2571.doi: 10.1371/journal.pone.0002571 [PubMed: 18626516]
65. Luchinat E, Secci E, Cencetti F, Bruni P. Sequential protein expression and selective labeling for in-cell NMR in human cells. *Biochim Biophys Acta.* 2016; 1860(3):527–33. [PubMed: 26723170]
66. Schanda P, Kupce E, Brutscher B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR.* 2005; 33(4):199–211. [PubMed: 16341750]
67. Riek R, Fiaux J, Bertelsen EB, Horwich AL, Wuthrich K. Solution NMR techniques for large molecular and supramolecular structures. *J Am Chem Soc.* 2002; 124(41):12144–53. [PubMed: 12371854]
68. Golub GH, Van Loan CF. *Matrix Computations.* 4. The Johns Hopkins University Press; Baltimore, USA: 2012.
69. Majumder S, DeMott CM, Burz DS, Shekhtman A. Using singular value decomposition to characterize protein-protein interactions by in-cell NMR spectroscopy. *Chembiochem.* 2014; 15(7):929–33. [PubMed: 24692227]

70. Hough LE, Dutta K, Sparks S, Temel DB, Kamal A, Tetenbaum-Novatt J, Rout MP, Cowburn D. The molecular mechanism of nuclear transport revealed by atomic-scale measurements. *Elife*. 2015; 4
71. Yamada J, Phillips JL, Patel S, Goldfien G, Calestagne-Morelli A, Huang H, Reza R, Acheson J, Krishnan VV, Newsam S, Gopinathan A, Lau EY, Colvin ME, Uversky VN, Rexach MF. A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. *Mol Cell Proteomics*. 2010; 9(10):2205–24. [PubMed: 20368288]
72. Ader C, Frey S, Maas W, Schmidt HB, Gorlich D, Baldus M. Amyloid-like interactions within nucleoporin FG hydrogels. *Proc Natl Acad Sci U S A*. 2010; 107(14):6281–5. [PubMed: 20304795]
73. Patel SS, Belmont BJ, Sante JM, Rexach MF. Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. *Cell*. 2007; 129(1):83–96. [PubMed: 17418788]
74. Schneider DA, Gourse RL. Relationship between growth rate and ATP concentration in *Escherichia coli*: a bioassay for available cellular ATP. *J Biol Chem*. 2004; 279(9):8262–8. [PubMed: 14670952]
75. Luy B, Marino JP. (1)H-(31)P CPMG-correlated experiments for the assignment of nucleic acids. *J Am Chem Soc*. 2001; 123(45):11306–7. [PubMed: 11697980]
76. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsveld J, Hentze MW. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*. 2012; 149(6):1393–406. [PubMed: 22658674]
77. Simsek D, Tiu GC, Flynn RA, Byeon GW, Leppek K, Xu AF, Chang HY, Barna M. The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. *Cell*. 2017; 169(6):1051–1065 e18. [PubMed: 28575669]
78. Banci L, Barbieri L, Luchinat E, Secci E. Visualization of redox-controlled protein fold in living cells. *Chem Biol*. 2013; 20(6):747–52. [PubMed: 23790485]
79. Mercatelli E, Barbieri L, Luchinat E, Banci L. Direct structural evidence of protein redox regulation obtained by in-cell NMR. *Biochim Biophys Acta*. 2016; 1863(2):198–204. [PubMed: 26589182]

Significance

The ultimate goal of structural and biochemical research is to understand how macromolecular interactions give rise to and regulate biological activity in living cells. The challenge is formidable due to the complexity that arises not only from the number of proteins (genes) expressed by the organism, but also from the combinatorial interactions between them. Despite ongoing efforts to decipher the complex nature of protein interactions, new methods for structurally characterizing protein complexes are needed to fully understand molecular networks. With the onset of in-cell NMR spectroscopy, molecular structures and interactions can be studied under physiological conditions shedding light on the structural underpinning of biological activity.

Figure 1.

NMR spectra of Thioredoxin. A. In-cell ^1H $\{^{15}\text{N}\}$ -HSQC. Only small metabolites can be observed due to an increase in the apparent molecular weight of Thioredoxin. Quinary interactions with cytosolic components are responsible for this increase. B. *In vitro* ^1H $\{^{15}\text{N}\}$ -HSQC. C. Overlay of in-cell ^1H - ^{15}N CRINEPT-HMQC-TROSY spectra acquired in lysates (red) and in-cell (blue). Insets show individual crosspeaks for selected residues, black contours are crosspeaks acquired *in vitro*. Adapted from Majumder *et al.*[19].

Figure 2. NMR solution structure of TTHA1718 in living *E. coli* cells. **A.** Superposition of the 20 final structures of TTHA1718 determined in *E. coli* showing the backbone (N, C) atoms. **B.** Superposition of the 20 final structures of purified TTHA1718 determined *in vitro*. **C.** A Comparison of TTHA1718 structures from living *E. coli* cells and *in vitro*. The best fit superposition of backbone atoms are shown with the same color code as in **A** and **B**. Adapted from Sakakibara *et al.* [44].

Figure 3.

Interaction surface maps of Ubiquitin-ligand complexes. The interaction surface of Ubiquitin is mapped onto the three-dimensional structure of Ubiquitin (PDB code 1D3Z). Individual residues exhibiting either a chemical shift change >0.05 ppm or significant differential broadening are indicated in red. All perturbed residues lie on the Ubiquitin surface and therefore reflect changes in the interaction surface of the molecule rather than changes in tertiary or quaternary structure. **a)** STAM2-Ubq interaction; **b)** Hrs-Ubq interaction; **c)** STAM2-Hrs-Ubq interaction; **d)** phosphorylated STAM2-Ubq interaction (YP-STAM2); **e)** phosphorylated Hrs-Ubq interaction (YP-Hrs); **f)** phosphorylated STAM2-Hrs-Ubq interaction (YP-STAM2-Hrs). Ubiquitin ligands are indicated in each panel. Adapted from Burz *et al.*[64].

Figure 4. Quinary interactions of Trx in *E. coli* occlude the active site

(a) Residues involved in the quinary interactions (red) are mapped onto the molecular surface of Trx (PDB code 1X0B); active site residues, C33 and G34, are in bold. **(b)** The relative volumes of the G52, G66 and G85 crosspeaks in the in-cell ^1H - ^{15}N CRINEPT-HMQC-TROSY spectra of Trx are plotted against the CRINEPT transfer delay times. An endogenous tryptophan indole amide peak in the in-cell spectra is used as a reference. The optimum CRINEPT transfer delay for Trx is ~ 1.3 ms, which corresponds to an apparent molecular weight of 1.1 MDa. Adapted from Majumder *et al*[19].

Figure 5.

SVD analysis of in-cell Pup-Mpa interactions. **A.** The intensity of individual crosspeaks of Pup change at different rates during interactor protein, Mpa over-expression. SVD analysis evaluates the magnitude of the contribution of an intensity change to the NMR data over the experimental time course to identify Mpa concentration-dependent interactions. **B.** SVD of the experimental data matrix M of size $m \times n$ yields the matrices U , Σ , and V^T with sizes $m \times m$, $m \times n$, and $n \times n$, respectively, where m is the number of target protein amino acid residues used in the NMR analysis, n is the number of time course NMR datasets, and V^T is the transpose of matrix V . **C.** The Scree plot shows the distribution of singular values for each dataset index (binding mode) from 1 to 6. The root mean square deviation, RMSD, values between respective components and the complete dataset are indicated by solid circles. **D.** The weighted contribution of each Pup amino acid residue to the Mpa principal binding modes, calculated as a product of a corresponding singular value and left singular, is shown for the 1st (black) and 2nd (hatched) binding modes. A threshold of 0.15 was chosen to highlight the amino acids that exhibit the largest changes. Negative values are due to spectral overlap between the target protein and cellular metabolites. Adapted from Majumder *et al* [69].

Figure 6.

FG-N and FSFG-K constructs engage in interactions with the cellular milieu and exhibit very rapid motions. (Upper panels) R_2/R_1 ratios indicate overall motion and the effects of multiple environments on chemical exchange in *E. coli*. (Lower panels) Heteronuclear nuclear Overhauser effect (nOe) reveals backbone motions. Gray columns indicate the locations of the FG motifs. Adapted from Hough *et al.*[70].

Figure 7.

Ribosome interactions modulate enzymatic activity *in vitro*. **A.** ADK initial velocity versus [ATP] without (black) and with (red) 1 μM ribosome. **B.** DHFR initial velocity versus [DHF] without (black) and with (red) 0.5 μM ribosome. **C.** TS initial velocity versus [dUMP] without (black) and with (red) 0.5 μM ribosome. DHF is dihydrofolate a substrate for DHFR and dUMP is deoxyuridylate, a substrate for TS. Adapted from DeMott *et al.*[10].