



Published in final edited form as:

Biochemistry. 2011 November 1; 50(43): 9225–9236. doi:10.1021/bi201287e.

Exploring weak, transient protein-protein interactions in crowded *in vivo* environments by in-cell NMR spectroscopy†

Qinghua Wang^{§,†,⊥}, Anastasia Zhuravleva^{§,⊥}, and Lila M. Gierasch^{§,||,*}

[§]Department of Biochemistry and Molecular Biology, University of Massachusetts, Amherst, 01003

^{||}Department of Chemistry, University of Massachusetts, Amherst, 01003

[‡]Center for Bioinformatics and Computational Biology, Delaware Biotechnology Institute, University of Delaware, Newark, DE 19711

Abstract

Biology relies on functional interplay of proteins in the crowded and heterogeneous environment inside cells, and functional protein interactions are often weak and transient. Thus, methods are needed that preserve these interactions and provide information about them. In-cell NMR spectroscopy is an attractive method to study a protein's behavior in cells because it may provide residue-level structural and dynamic information. Yet several factors limit the feasibility of protein NMR spectroscopy in cells, and among them slow rotational diffusion has emerged as the most important. In this paper, we seek to elucidate the causes of the dramatically slow protein tumbling in cells and in so doing to gain insight into how the intracellular viscosity and weak, transient interactions modulate protein mobility. To address these questions, we characterized the rotational diffusion of three model globular proteins in *E. coli* cells using 2D heteronuclear NMR spectroscopy. These proteins have a similar molecular size and globular fold, but very different surface properties, and indeed, they show very different rotational diffusion in the *E. coli* intracellular environment. Our data are consistent with an intracellular viscosity approximately eight times that of water—too low to be a limiting factor to observing small globular proteins by in-cell NMR spectroscopy. Thus, we conclude that transient interactions with cytoplasmic components significantly and differentially affect the mobility of proteins and therefore their NMR detectability. Moreover, we suggest that an intricate interplay of total protein charge and hydrophobic interactions plays a key role in regulating these weak intermolecular interactions in cells.

Keywords

anomalous diffusion; intracellular mobility; in-cell NMR; macromolecular crowding; protein association; transient protein-protein interactions

†This work was supported by an NIH Director's Pioneer Award (OD000945).

*Corresponding author: Tel: (413) 545-6094; Fax: (413) 545-1289; gierasch@biochem.umass.edu.

⊥These authors contributed equally to this work.

Supporting Information Details about in-cell NMR experiments and leakage controls; amino acid sequences of the fusion constructs; Table S1 (experimental datasets), Table S2 (biophysical properties of the proteins used in this study); Figure S1 (determination of ³J(HN-HA) constants), Figure S2 (viscosity as a function of glycerol concentration), Figure S3 (viscosity corrections for lysate samples), Figure S4 (molecular tumbling and surface properties of dGB1 versus GB1), Figure S5 (protein leakage control for GB1 in-cell spectrum), Figure S6 (controls for ubiquitin in-cell spectrum), Figure S7 (lysate and in-cell NMR spectra), Figure S8 (surface properties of GB1, NmerA, and ubiquitin), and Figure S9 (surface changes for the Ubi3A mutant). Supplemental materials may be accessed free of charge online at <http://pubs.acs.org>.

Understanding biological systems requires knowledge of the behavior of their basic components and the interactions between them in cellular environments. A crucial difference between *in vivo* and *in vitro* conditions is the high concentration of macromolecules (1), which can range from 200 g/l in the eukaryotic cytoplasm to over 400 g/l in the cytoplasm of prokaryotes, where crowding seems most extreme (2). There have been a number of computational and experimental studies demonstrating that the complex cellular environment significantly modulates the behavior of macromolecules, affecting their structure, dynamics and stability (for reviews, see (3-5)). Many of these studies have relied on model crowding agents, either inert synthetic polymers or model proteins such as bovine serum albumen, and important principles have emerged from this work. For example, the crowded and heterogeneous cellular environment enhances the probability of promiscuous, nonspecific interactions compared with specific ones, and consequently, insuring specificity in biological processes requires more complicated regulation of protein behavior and interaction networks than has been envisioned in dilute solution studies (5-7). Other consequences of the high intracellular concentration of macromolecules are less well elucidated by studies with model crowding agents, such as the significant increase in intracellular viscosity, which will affect all intracellular processes that rely on diffusion-driven thermodynamics and kinetics, including macromolecular folding, recognition, binding, and catalysis (1, 4, 5). Recent work shows that the viscosities of highly concentrated protein solutions depend to a great extent on intermolecular interactions and factors affecting these interactions, such as charge, shape, and size (8). To some extent, bulk solution viscosity may be replaced by the local apparent viscosity, which includes crowding contributions to molecular diffusion (9). However, a single value for the apparent cytoplasmic viscosity cannot explain experimental observations of anomalous diffusion (*i.e.*, when the mean squared displacement does not increase linearly with time), nor the fact that proteins having the same size and shape, and consequently, the same hydrodynamic radius, display significantly different diffusion constants (1, 2, 10). It has been suggested that these deviations can be caused by the heterogeneity of intracellular environments (11) and/or macromolecular interactions such as protein-protein, protein-nucleic acid, or protein-lipid interactions (2).

In a provocative 1982 paper, McConkey pointed to the importance of weak but specific transient interactions in living systems. He coined the term 'quinary structure of proteins' to describe these interactions and emphasized that they were crucial to cellular organization and function, and that disruption of cellular integrity abolishes them, making it necessary to study them in intact cells (12). Weak interactions between proteins, *i.e.*, those with $K_d > 1 \mu\text{M}$ (13, 14), are in fact an unavoidable consequence of intracellular crowding, and thus evolutionary selection tunes their physiological roles in the living cell to enable their regulation and to suppress undesirable associations.

It is clearly important to better understand the nature of weak transient interactions between proteins in cells in order to elucidate how they shape protein functionality. A related question is how cells distinguish between specific (*i.e.*, physiologically productive) and nonspecific (of potential physiological harm) interactions. As the ability and propensity to participate in weak transient interactions is evolutionarily selected and encoded in protein sequences, it should be possible to elaborate principles to predict and account for a given protein's interaction profile. For example, it has been suggested that (de)solvation is a major physical factor in protein-protein interactions, and this suggestion has been supported by the discovery of significant correlation between the number of interactions made by a protein and the fraction of hydrophobic residues on its surface (15). Interestingly, no significant correlation was found between the percentages of charged amino acids on the surface and number of interactions (15). In a recent computational simulation of the diffusion behavior of proteins in the *E. coli* cytoplasm, McGuffee and Elcock (16) demonstrated that steric and

electrostatic interactions between the most abundant fifty *E. coli* proteins are not sufficient to predict realistic translational and rotational diffusion coefficients for a well-studied model protein, green fluorescent protein (GFP) in cells. As expected from the work of Deeds *et al.* (15), addition of short-range attractions between exposed hydrophobic atoms yielded results consistent with reported experimental measurements of GFP diffusion. Moreover, McGuffee and Elcock showed that electrostatic interactions resulted in only a small effect on calculated diffusion coefficients. Note that for these calculations, a single ‘effective charge’ was used instead of heterogeneous charge distributions on molecular surface for real proteins, while short-range attractions were computed more realistically, *i.e.*, between individual atoms. Consequently, the role of electrostatic forces could be underestimated or overshadowed by other contributions in those calculations. Indeed, Pielak and colleagues experimentally demonstrated that electrostatic forces contribute significantly to nonspecific interactions between barley chymotrypsin inhibitor 2 and a protein crowding agent, bovine serum albumen (17).

An experimental study of transient interactions in the extremely complex and heterogeneous intracellular environment is a very challenging task. NMR spectroscopy is a particularly attractive tool for studying a protein’s behavior in cells, because it provides information at the residue level (18-21). However, surprisingly few proteins have yielded interpretable in-cell NMR spectra, primarily because of extreme signal broadening and consequent reduced sensitivity and resolution (22, 23). There are several factors limiting the feasibility of obtaining an NMR spectrum in cells (22), and among them slow rotational diffusion, as demonstrated by Pielak and colleagues, has emerged as the most important one (24). However, what causes the dramatically slow protein tumbling in cells remains elusive, and it remains unclear the extent to which the intracellular viscosity and/or transient interactions modulate protein intracellular mobility.

To address these questions and to understand how cells control weak transient interactions between macromolecules, we characterized the rotational diffusion in the *E. coli* cytoplasm of three model proteins (or protein domains) — protein G B1 domain (GB1), the N-terminal metal-binding domain of mercuric ion reductase (NmerA), and ubiquitin — using 2D heteronuclear NMR spectroscopy. These proteins have a similar molecular size and globular fold, but very different surface properties, and indeed, they show very different rotational diffusion in the *E. coli* intracellular environment. Our data are consistent with an intracellular viscosity approximately eight times that of water, which would not be a limiting factor to observe small globular proteins by in-cell NMR spectroscopy. Thus, we conclude that transient interactions with cytoplasmic components significantly and differentially affect the in-cell mobility of proteins, and constitute a major contributor to the drastic line broadening and drop in spectral sensitivity in in-cell NMR. Moreover, as a result of our analysis of three proteins of similar size and shape that yield distinct in-cell NMR signatures, we suggest that an intricate interplay of total protein charge and hydrophobic interactions plays a key role in regulating weak transient interactions between proteins in cells.

MATERIALS AND METHODS

Design of Protein Constructs

The GB1 construct (pET21a-GB1) was created by including a site-directed mutation on the GEV2 vector (based on pET21a originally) (25) to generate a stop codon following the GB1 domain, using PCR primers: 5′ - CCT TCA CGG TAA CCG AAT AGG TTC CGC GTG GAT CC - 3′; 5′ - GGA TCC ACG CGG AAC CTA TTC GGT TAC CGT GAA GG - 3′. This and all other constructs were confirmed by DNA sequencing (Genewiz).

The NmerA construct (in pET-11a vector) was a gift from the Dötsch lab (22). The ubiquitin D77 construct (in pET3a vector) was a gift from the Walters lab (26, 27). The ubiquitin triple-point mutant I8A/I44A/V70A (Ubi3A) (28) was created by sub-cloning the synthesized gene with triple Ala mutations into a pET16b vector, but neither this mutant nor the above ubiquitin D77 construct encodes extra tags.

dGB1 was constructed as follows: The GB1 gene from pET21a-GB1 was cloned into pET16b. The resulting plasmid is named pET16b-GB1-C. Then, DNA coding for full-length GB1 (amino acids 1-56) was PCR amplified from pET21a-GB1, digested and ligated into pET16b-GB1-C.

Construction of GB1-L₁₅-NmerA fusion: DNA coding for full-length GB1 (amino acids 1-56) plus 15-aa (*i.e.*, -SerGlySer(Gly)₁₁His-) at the C-terminus was PCR amplified and ligated into pET16b-NmerA-C to obtain a GB1-NmerA fusion construct (*i.e.*, GB1 and NmerA linked through a 15-aa linker -SerGlySer(Gly)₁₁His-). The final construct was confirmed by DNA sequencing.

NMR Sample Preparation

Preparation of cell samples over-expressing uniformly ¹⁵N-labeled [U-¹⁵N] - protein of interest—If not specified, the BL21 (DE3) (Novagen) cell line was used for expression. An overnight culture was grown at 30 or 37 °C from 5 mL Luria-Bertani medium (LB) supplemented with 100 mg/L ampicillin by inoculating the culture with a colony from newly transformed cells. Then, 100 mL LB was inoculated with 3 mL overnight culture and grown to an OD₆₀₀ of 0.8 or so. The cells were harvested by gentle centrifugation (~1,400 g for 10 min) and re-suspended to an OD₆₀₀ of 0.5-0.6 in 100 mL of M9 containing [U-¹⁵N] ammonium chloride (1.0 g/L) and glucose (4 g/L) as the sole nitrogen and carbon sources, along with 100 mg/L ampicillin. The culture was incubated at 37 °C for 10-15 minutes before protein expression was induced by adding 1/2000th culture volume of 1.0 M IPTG. Induction was allowed to proceed for 3 hours (if not otherwise specified). Following the 3-hour induction, the 100 mL culture was centrifuged gently (~1,400 g for 10 to 15 min), after which the resulting cell pellets were left sitting in ~500 µL M9 minimal medium together with ~50 µL D₂O for 15 min or so, and then the cells were very gently re-suspended well with a pipette. Finally the cell sample was transferred to an NMR tube and used immediately for NMR experiments.

Preparation of cell lysate—The cell samples prepared as above were flash frozen with liquid nitrogen and stored in -80 °C. Then they were thawed on ice, sonicated in an ice bath, with cooling between sonication bursts. The sonication cycle was repeated until the viscosity of the sample decreased significantly or the suspension became partially clear. Then the sample was centrifuged in a table-top mini-centrifuge at 13K rpm for 10 min or so at 4 °C, until the supernatant and pellet were well separated. The supernatant was carefully transferred to an NMR tube as a cleared cell lysate sample.

Purification of GB1—GB1 without any tag was successfully purified by following the method in (29). Basically, the cell slurry (~40 mL, harvested from 2-L M9 minimal medium culture) (*N.B.*, the cell sample was neither sonicated nor microfluidized) was heated at 80 °C for five minutes, immediately chilled on ice for ten minutes and then centrifuged at 16,000 rpm (Beckman JS 5.3 rotor) for 30 minutes at 4 °C. The supernatant was passed through a 0.45 µm syringe filter, dialyzed against 4 L milliQ water at 4 °C twice (for at least 1 hour each time), adjusted to 50 mM sodium phosphate (pH 5.6), concentrated using Amicon Ultra-4 5K centrifugal filter devices (Millipore Corp.), and loaded onto a Superdex-75 column (Hiload 16/60 prep grade; Amersham Biosciences) pre-equilibrated in 50 mM

sodium phosphate buffer (pH 5.6) at a flow rate of 0.5 - 1 mL/min at 4 °C. The identity of purified protein was verified by mass spectrometry.

Purification of His-tagged ubiquitin—Cells for the expression of uniformly ^{15}N labeled His-tagged ubiquitin were grown at 37 °C in M9 minimal medium, with the sole nitrogen source being $^{15}\text{NH}_4\text{Cl}$. The cells were induced to over-express ubiquitin by adding 0.5 mM IPTG, and were grown for additional 4 hours, harvested by centrifugation, and stored at -80 °C. The cell extract was affinity-purified by a Ni-NTA column (Qiagen). The ubiquitin-containing eluant was concentrated, and the buffer was exchanged into 10 mM potassium phosphate, pH 7.0.

NMR Spectroscopy

All NMR experiments were recorded on Bruker AVANCE 600-MHz NMR spectrometers equipped with cryo-probes and at 298 K. All [^1H , ^{15}N] HSQC data sets collected for the purpose of measurements of $^1\text{H}^{\text{N}}$ linewidths were acquired with ^1H resolution of about 7 to 8 Hz. For each FID 8 to 32 transients were collected, corresponding to total experiment durations of 10 to 40 minutes. No apodization function, baseline correction, or linear prediction was applied in the proton dimension to avoid an artificial disturbance of $^1\text{H}^{\text{N}}$ lineshapes; zero-filling was applied to set a digital resolution in the ^1H dimension of about 1 Hz.

All ^{15}N TROSY and anti-TROSY data were measured using a 2D [^1H , ^{15}N] TROSY experiment (30) by changing the phase cycle for the selection of upfield and downfield components in ^{15}N dimensions. Data sets were acquired with ^{15}N resolution of about 9 Hz, and zero-filling was applied to set the digital resolution to about 2 Hz. For in-cell NMR samples, total experiment duration is 2 hours each for TROSY and anti-TROSY. All spectra were processed with NMRPipe (31). GB1, NmerA and ubiquitin backbone amide assignments were transferred from previous assignments (BMRB accession codes 7280, 16208 and 6466) using CARA (32).

$^1\text{H}^{\text{N}}$ Linewidth Data Analysis

For linewidth analysis a set of 1D proton slices from a [^1H , ^{15}N] HSQC spectrum was selected for non-overlapping, assigned peaks. Each slice was positioned at the peak maximum in the ^{15}N dimensions and centered about the peak maximum in the ^1H dimension; the width of each slice was about 3 times the peak linewidth. $^3\text{J}(\text{HN-HA})$ coupling was included in the simulation of the ^1H lineshapes, and experimental data points were fitted into a sum of two Lorentzians:

$$F(\nu) = \frac{A_1 (\Delta\nu/2)}{(\Delta\nu+2)^2 + (\nu_0 - ^3\text{J}/2 - \nu)^2} + \frac{A_2 (\Delta\nu/2)}{(\Delta\nu+2)^2 + (\nu_0 + ^3\text{J}/2 - \nu)^2} \quad (\text{Eq. 1})$$

where $(\nu_0 + ^3\text{J}/2)$ and $(\nu_0 - ^3\text{J}/2)$ are ^1H frequencies (in Hz) of the maxima of a peak doublet separated by the $^3\text{J}(\text{HN-HA})$ coupling constant, $\Delta\nu$ is the full width at half-height, and A_1 and A_2 are normalization constants. $^3\text{J}(\text{HN-HA})$ constants for individual residues were fixed at values obtained as described below. Uncertainties of the $^1\text{H}^{\text{N}}$ linewidths were set to be the larger of the fit uncertainty and $\Delta\nu/S$, where S is the signal to noise ratio in an NMR spectrum.

Determination of $^3\text{J}(\text{HN-HA})$ constants—For small proteins in buffer solution and at low glycerol concentrations, $^3\text{J}(\text{HN-HA})$ coupling contributes significantly to $^1\text{H}^{\text{N}}$ lineshapes, resulting in resolved or unresolved peak doublets (Fig. S1). We used GB1 and

NmerA lysate samples and a purified His-tagged ubiquitin sample to obtain $^3J(\text{HN-HA})$ constants for individual residues. Experimental data from two or three different samples were fitted using Equation (1), with 3J as an adjustable parameter, and the resulting values were averaged over all samples. Finally, to check the reliability of the obtained $^3J(\text{HN-HA})$ values, they were compared with the $^3J(\text{HN-HA})$ values predicted from corresponding 3D protein structures (Fig. S1).

Viscosity dependence of $^1\text{H}^{\text{N}}$ linewidths in GB1—To determine how the GB1 spectra change with viscosity, we obtained $^1\text{H}^{\text{N}}$ linewidths for individual peaks in a corresponding [$^1\text{H}, ^{15}\text{N}$] HSQC spectrum (see above). The interpretation of absolute values for $^1\text{H}^{\text{N}}$ linewidths, which are mostly determined by dipole-dipole interactions with surrounding protons (33), is not a simple task, requiring knowledge of many parameters, including local and global dynamics and structure (34). Therefore, we drew on glycerol titrations to determine the relationship between residue $^1\text{H}^{\text{N}}$ linewidths and solution viscosity. Three data sets were used for glycerol titrations: purified GB1 (35 - 60 wt% glycerol, dataset 1), on a GB1/NmerA mixed lysate (0-40% glycerol, dataset 2), and on dGB1 lysate (0-40% glycerol, dataset 3) (Table S1). For all titrations we used D8-glycerol (D8 99%, from Cambridge Isotope Laboratories) (Fig. S2) to avoid strong background signals in NMR spectra. The solution viscosity for each sample was calculated based on known glycerol concentrations (see Supporting Information for more details). Taking into account lysate viscosities and the difference in molecular tumbling between GB1 and dGB1 (35), we corrected the viscosities of GB1 and dGB1 lysate samples at each glycerol concentrations (Table S1, Fig. S3, Fig. S4). For each sample from dataset 1-3, we obtained $^1\text{H}^{\text{N}}$ linewidths for individual peaks, corresponded to unambiguously assigned GB1 residues. For final analysis, we chose 26 GB1 residues that showed reliable data in at least two datasets (*i.e.*, they have non-overlapped peaks in HSQC spectra and $^1\text{H}^{\text{N}}$ linewidth errors of less than 20%, and each dataset exhibited linear dependence of $^1\text{H}^{\text{N}}$ linewidths on viscosity). To obtain $^1\text{H}^{\text{N}}$ linewidth *vs.* viscosity calibration slopes for each residue, $^1\text{H}^{\text{N}}$ linewidths for the three data sets were fitted simultaneously as a function of the solution viscosity. Finally, these calibration slopes were used for the estimation of the unknown intracellular viscosity from $^1\text{H}^{\text{N}}$ linewidths in the in-cell HSQC spectrum (see below).

Determination of the intracellular viscosity using $^1\text{H}^{\text{N}}$ linewidth analysis for GB1— $^1\text{H}^{\text{N}}$ linewidths for individual residues were obtained from HSQC spectra for three different GB1 in-cell samples analyzed as described above. For each residue the resulting values were averaged over all three in-cell data sets, and uncertainties were set to be the larger of the maximum uncertainty for individual samples and the standard deviation between different samples. Glycerol viscosity calibration slopes for individual residues (see above) were used to calculate intracellular viscosity from the in-cell $^1\text{H}^{\text{N}}$ linewidths averaged over all three in-cell HSQC spectra. The resulting viscosity values were averaged over all the analyzed residues, and an uncertainty was calculated as the standard deviation between different residues.

Prediction of $^1\text{H}^{\text{N}}$ linewidths for NmerA in-cell—First, we fitted $^1\text{H}^{\text{N}}$ linewidths for 23 NmerA residues as a function of solution viscosity for the GB1/NmerA mixture lysate dataset (0-40% glycerol, dataset 2 from Table S1) to obtain viscosity calibration slopes, and based on the slopes we calculated $^1\text{H}^{\text{N}}$ linewidths expected for the intracellular viscosity obtained from the $^1\text{H}^{\text{N}}$ linewidth analysis of GB1 (see above).

BSA titration of the GB1/NmerA lysate sample—To mimic intracellular molecular crowding with a protein crowding agent, $^1\text{H}^{\text{N}}$ linewidths for individual residues in GB1 and NmerA were obtained from HSQC spectra for the GB1/NmerA mixture lysate in the

presence of 100 and 200 g/L of BSA (ACROS Organics). To obtain an average linewidth for each protein, the resulting values were averaged over all analyzed residues. For each GB1 residue a glycerol viscosity calibration slope was used to obtain the viscosity in the 100 and 200 g/L BSA samples (see above). The resulting values were averaged over all analyzed residues, and an uncertainty was calculated as a standard deviation between different residues. The apparent molecular weights for GB1 and NmerA were calculated as $MW^{app} = MW \cdot \eta^{BSA} / \eta_0$, where MW is the protein molecular weight, η^{BSA} and η_0 are the viscosity of the lysate solution in the presence of 100 or 200 g/L of BSA and the viscosity of water, respectively. η^{BSA} was obtained in the same way as described for determination of the intracellular viscosity using $^1H^N$ linewidth analysis for GB1 (see above).

Average $^1H^N$ linewidth analysis for lysate samples—To understand whether molecular weight is the primary determinant of protein $^1H^N$ linewidths or whether other factors (*e.g.*, transient interactions) significantly affect linewidths, and how these contributions vary from protein to protein, we plotted the average $^1H^N$ linewidth for GB1, NmerA, ubiquitin and Ubi3A as a function of the protein molecular weight. $^1H^N$ linewidths of individual residues for GB1, NmerA, ubiquitin and Ubi3A were obtained from HSQC spectra of their lysates. To obtain an average linewidth for each protein, the resulting values were averaged over all analyzed residues, and the uncertainty was calculated as the standard deviation between different residues. The apparent molecular weights were calculated as $MW^{app} = MW \cdot \eta^{lysate} / \eta_0$, where MW is the protein molecular weight, η^{lysate} and η_0 are the viscosities of the lysate and water, respectively.

TROSY/anti-TROSY Data Analysis

For data analysis, a set of 1D ^{15}N slices from each [$^1H, ^{15}N$] TROSY and anti-TROSY spectrum was selected for analysis of non-overlapped peaks with known assignments. Each slice was positioned at the peak maximum in the 1H dimension and centered about the peak maximum in the ^{15}N dimension; the width of each slice was about 3 times as great as the peak linewidth. Experimental data points were fitted to a Lorentzian function:

$$F(\nu) = \frac{A(\Delta\nu/2)}{(\Delta\nu/2)^2 + (\nu_0 - \nu)^2} \quad (\text{Eq. 2}),$$

where ν_0 is the ^{15}N frequency (in Hz) at the peak maximum, $\Delta\nu$ is the full width at half-height, and A is a normalization constant. Uncertainties of the obtained linewidths were set as the maximum of fit uncertainties and $\Delta\nu/S$, where S is the signal to noise ratio.

The difference in linewidths between TROSY and anti-TROSY lines, $\Delta\Delta\nu^{TAT}$, was calculated as $\Delta\Delta\nu^{TAT} = \Delta\nu^{anti-TROSY} - \Delta\nu^{TROSY}$; uncertainties were estimated as

$\sqrt{(\sigma^{TROSY})^2 + (\sigma^{anti-TROSY})^2}$, where σ was set to the maximum of the fit uncertainty and $\Delta\Delta\nu/S$, where S is the signal to noise ratio in the corresponding TROSY or anti-TROSY spectrum.

Glycerol titrations of purified GB1 and determination of the intracellular viscosity from TROSY/anti-TROSY analysis—To determine the relationship between $\Delta\Delta\nu^{TAT}$ and solution viscosity for individual residues in GB1, we performed glycerol titrations of purified GB1 (35-60% glycerol by weight, dataset 1 from Table S1). For the final analysis, we chose 10 GB1 residues, which showed non-overlapped peaks and reliable data at all glycerol concentrations studied. For each residue the obtained $\Delta\Delta\nu^{TAT}$ values were fitted as a linear function of viscosity and the resulting viscosity calibration slope was used to estimate the apparent intracellular viscosity. The resulting values of apparent

intracellular viscosities were averaged over all 10 residues, and an uncertainty was calculated as the standard deviation between different residues.

All linewidth analysis has been conducted using homemade scripts in Mathematica 7.0 (Wolfram).

RESULTS

Size Is Not Necessarily a Limiting Factor for Observing Protein NMR Signals in Cells

The promise of NMR spectroscopy as a method for studying protein behavior *in vivo* is confounded by several technical problems, including low stability of in-cell samples and protein leakage into the extracellular medium (36) (see Supporting Information for more details, Fig. S5). Most importantly, to date, surprisingly few proteins have been visible by in-cell NMR spectroscopy. There are several parameters that may be critical for observing a protein in the living cell, including protein concentration (22), the rotational correlation time in the cytoplasm (24), protein stability (17, 37, 38), conformational and internal dynamics (24), as well as oligomerization and interaction with other components of the cytoplasm (23). To understand the role of these factors we performed NMR spectroscopy studies in *E. coli* cells expressing isotopically labeled GB1, NmerA, and ubiquitin (Table S2). Despite the fact that all these proteins demonstrated high quality NMR spectra *in vitro*, we found that only two of them, GB1 and NmerA, showed a reasonably good in-cell NMR spectrum (Fig. 1, Fig. S7), while ubiquitin is invisible in cells (Fig. S6 and S7). [*N.B.*, Our results for ubiquitin agree with Li *et al.*, who reported that they saw no signals for ubiquitin in *E. coli* cells (39), but not with Burz *et al.* (18), who reported relatively high resolution spectra for ubiquitin in *E. coli* cells using a protocol that differed from ours by freezing cells before taking spectra and expressing significantly lower concentrations of ubiquitin.]

For both GB1 and NmerA we observed significant line broadening for in-cell spectra; however, GB1 demonstrated markedly better spectral quality than NmerA (Fig. 1). To understand why these two small globular proteins exhibited such different behavior in cells, we examined several factors, which could explain the greater line broadening and lower sensitivity of the NmerA spectrum, including the difference in GB1 and NmerA concentrations, the difference in their sizes, and differential propensities for transient interactions with intracellular components. Given the strong similarity in peak positions in-cell and *in vitro*, we suggest that neither GB1 nor NmerA forms specific high affinity complexes with other cytoplasmic macromolecules inside *E. coli* cells, and the in-cell structures of both GB1 and NmerA are similar to those *in vitro*. Even though GB1 and NmerA appeared to be expressed at similar levels in *E. coli* cells (inferred from the comparable NMR signal intensities of their lysate samples), to eliminate confusion arising from small differences in their expression level, we designed a chimeric fusion protein with a flexible tether between the linked GB1 and NmerA domains (see MATERIALS AND METHODS). Chemical shift analysis of the GB1-L₁₅-NmerA fusion protein (*i.e.*, the two domains connected by a linker of 15-amino acids) in buffer revealed that the GB1 and NmerA domains within the fusion protein adopt the same fold and rotate relatively independently of each other. For the GB1-L₁₅-NmerA fusion, cross peaks corresponding to both GB1 and NmerA were detected in the in-cell HSQC spectrum. However, the intensities of NmerA peaks were significantly lower than those of GB1 (data not shown), indicating that the difference in peak intensities is not a result of different protein intracellular concentrations.

It was previously suggested that macromolecular crowding slows down molecular tumbling and results in faster relaxation and broader in-cell NMR peaks (17, 23, 24). Indeed, the backbone amide peaks of NmerA became so broad that they were barely visible in the

NmerA in-cell spectrum, while NH₂ side-chains in NmerA, for which relaxation is mainly determined by fast internal dynamics and therefore is less dependent on molecular tumbling, displayed sharp, intense peaks (Fig. 1). Moreover, previously it was shown that deuteration, bringing about a reduction of ¹H^N relaxation, significantly improved the quality of in-cell NmerA spectra (22). These observations suggest that the reduced rotational correlation time of NmerA indeed plays an important role in the line broadening observed for in-cell NmerA spectra. The molecular tumbling is roughly proportional to a protein's size; however, it is not clear if a small difference in the molecular size between GB1 and NmerA (*i.e.*, 6.18 and 6.91 kDa, respectively) can account for the dramatic difference in the sensitivity of their in-cell spectra. To answer this question, we need to distinguish between the contribution to broadening from protein rotational diffusion and other possible contributing factors that affect protein NMR linewidths. To do this, we created a GB1 fusion construct with a one-residue linker between two linked GB1 domains, dGB1 (see MATERIALS AND METHODS), which has similar surface properties with GB1, but a rotational correlation time that is about twice that of GB1 (Fig. S4). As expected, with the increased rotational correlation time the in-cell NMR spectrum of dGB1 presented a lower signal to noise ratio (S/N) and broader peaks than the GB1 spectrum (Fig. 1). However, in cells this 12.5 kDa dGB1 protein construct exhibited much better resolved spectra than did the 6.91 kDa NmerA (Fig. 1). Moreover, ubiquitin, which is significantly smaller than dGB1 (Table S2), showed no backbone amide signals in cells (Fig. S6 and S7). All these findings indicate that, for many proteins, the reduced mobility from molecular size increase in the intracellular high viscosity environment is an important factor but not necessarily the limiting factor determining their NMR visibility in cells.

Determination of Apparent Intracellular Viscosity Using ¹H^N Linewidths and TROSY/anti-TROSY Analysis

Consistent with our observations, a recent in-cell NMR study by Crowley *et al.* concluded that GB1 had minimal interaction with other cytoplasmic macromolecules (40). Consequently we chose it to explore the contribution of global viscosity and molecular crowding on protein diffusion in the *E. coli* cell. There are several NMR techniques such as pulsed field gradient measurements and relaxation experiments that enable accurate and detailed characterization of protein diffusion. Unfortunately, most of them are not suitable for in-cell samples due to poor sample stability and the low signal-to-noise ratio of in-cell NMR spectra. By contrast, ¹H^N linewidths, which for many rigid protein systems are strongly correlated with overall rotational correlation time (33), can be accurately estimated from a [¹H,¹⁵N] HSQC spectrum, and consequently, can be used to probe sample viscosity. The interpretation of absolute values for ¹H^N linewidths, which are mostly determined by dipole-dipole interactions with surrounding protons (33), is not a simple task, requiring knowledge of many parameters, including local and global dynamics and structure (34). Instead, we drew on glycerol titrations to determine the relationship between residue ¹H^N linewidths and solution viscosity, using three data sets: purified GB1, a cell lysate containing a mixture of GB1 and NmerA, and cell lysate containing the dGB1 protein construct (see MATERIALS AND METHODS, and Table S1). The viscosity dependencies of ¹H^N linewidths obtained were used to estimate the apparent intracellular viscosity (Fig. 2A and B). We found that the ¹H^N linewidths obtained for GB1 in the *E. coli* cell corresponded to a viscosity about 11±2 times that of water (Fig. 3). This estimate of in-cell viscosity represents an upper limit, because other factors might also affect in-cell ¹H^N linewidths, including contributions linked to rotational diffusion, such as change of molecular size (*e.g.*, upon binding to other macromolecules) and protein shape (*e.g.*, upon unfolding), and factors unrelated to diffusion, *e.g.*, conformational exchange in the μs-ms time scale, as well as sample and magnetic field inhomogeneity. In order to exclude contributions not linked to molecular tumbling, we performed a linewidth analysis of GB1

using ^{15}N TROSY and anti-TROSY spectra. The difference between ^{15}N TROSY and anti-TROSY relaxation and consequently, their linewidths is determined primarily by interference between the ^1H - ^{15}N dipolar and ^{15}N CSA interactions (41). As a result, the difference between ^{15}N TROSY and anti-TROSY linewidths, $\Delta\Delta\nu^{\text{TAT}}$, is unaffected by contributions from chemical exchange and sample or magnetic field inhomogeneity. To avoid interpreting absolute $\Delta\Delta\nu^{\text{TAT}}$ values, we examined the viscosity dependence of $\Delta\Delta\nu^{\text{TAT}}$. Specifically we performed glycerol titrations of purified GB1 (dataset 1, Table S1), and found that $\Delta\Delta\nu^{\text{TAT}}$ values for individual GB1 residues can be fitted as a linear function of sample viscosity (Fig. 2C and D). These relationships were used to calculate the apparent viscosities in the *E. coli* cell based on individual residue data. The average apparent viscosity obtained from the TROSY/anti-TROSY analysis was about 30% lower than the one estimated from the $^1\text{H}^{\text{N}}$ linewidth analysis (Fig. 3), indicating that about 30% of the line broadening for in-cell GB1 spectra is not linked to molecular tumbling and comes from exchange and/or inhomogeneity contributions. Consequently, our analysis showed that GB1 rotationally diffuses about 8 ± 2 times more slowly in *E. coli* cells than in water. This decrease in the GB1 molecular tumbling rate includes contributions from the intracellular viscosity as well as other contributions linked to rotational diffusion, such as change of protein molecular size and protein shape. However, because GB1 shows minimal interaction with other cytoplasmic macromolecules (40), it is likely that intracellular viscosity plays a dominant role in the case of GB1 in reducing intracellular rotational diffusion. Nonetheless, one should keep in mind that even a small fraction of GB1 bound to large intracellular macromolecules would significantly affect the observed molecular tumbling, and consequently, our results provide only an upper limit of the apparent intracellular viscosity.

Effect of Protein Crowding on NmerA Diffusion

The apparent intracellular viscosity obtained from GB1 data analysis predicts that proteins with molecular size about 13 kDa or smaller will be detectable by in-cell NMR spectroscopy. By contrast, even though the molecular weight of NmerA is only 6.91 kDa and it is visible in cells, its in-cell NMR signals were much broader than expected if only intracellular viscosity contributed to line broadening. Consequently, macromolecular crowding alone fails to explain the dramatic broadening observed for NmerA and loss of signals for ubiquitin (8.7 kDa). Therefore other factors affecting NMR linewidths in the cell must be considered. Recently, it was shown that transient nonspecific interactions with other proteins significantly broadens resonances, suggesting that this may be an important factor for protein NMR visibility in-cell (23). Indeed, these interactions slow down protein diffusion by increasing protein molecular size and/or changing protein shape due to binding. In addition, line broadening can be caused by μs -ms conformational dynamics and/or sample inhomogeneity, when several protein species with slightly different chemical shifts are present. Moreover, weak, transient interactions might vary from protein to protein and depend on rather unique protein surface features more than on molecular size. To test the contributions of weak, transient interactions for GB1 and NmerA, we compared their $^1\text{H}^{\text{N}}$ linewidths in cell lysate, which had about the same soluble protein components as the cell cytoplasm, but was about 8-10 times more dilute. As a result, the apparent viscosity of the cell lysate, obtained by $^1\text{H}^{\text{N}}$ linewidth analysis, was much lower than in cells and only slightly higher than in water (Table S1). Based on previous results of Pielak and others showing significant differences between the effect of synthetic polymer and protein crowding agents (23), which suggested that protein crowders (and particularly, BSA) are more suitable mimics of the intracellular environments(37), we recorded NMR spectra in the presence of 100 and 200 g/L BSA. To keep the same sample conditions for both proteins, we used a mixture of GB1 and NmerA lysate in the same sample in the presence of 100 or 200 g/L BSA.

We used average $^1\text{H}^{\text{N}}$ linewidths as signposts for weak, transient interactions in the lysate and in the presence of BSA. To this end, utilizing GB1 and dGB1 glycerol titrations we obtained a linear dependence of average linewidths for apparent molecular weight, which are linked to molecular tumbling and linearly proportional to molecular size and viscosity (Fig. 4A, black). Transient interactions in the BSA-spiked lysates should significantly affect $^1\text{H}^{\text{N}}$ linewidths, resulting in positive deviations from this linear dependence, *i.e.*, when the higher bulk viscosity or/and molecular interactions increase(s) the average $^1\text{H}^{\text{N}}$ linewidth more than predicted.

Fig. 4A shows the average $^1\text{H}^{\text{N}}$ linewidths for GB1 and NmerA as a function of the apparent molecular weight and clearly demonstrates that the average NmerA and GB1 linewidth agrees very well with their predicted values. This result indicates the absence of significant broadening in the NmerA lysate spectrum comparing with GB1, and consequently, the absence of significant weak, transient interactions for both proteins in their lysate solution.

In the presence of the protein crowder, BSA, the $^1\text{H}^{\text{N}}$ linewidths of GB1 indicated that the apparent viscosities of the 100 and 200 g/L BSA samples were about 1.90 and 4.25 cP, respectively, which are higher than the bulk viscosities previously estimated in pure protein solution (<http://www.rheosense.com>), namely, about 1.4 and 2.2 cP for 100 and 200 g/L of BSA, respectively. Because the viscosity of our lysate samples was only slightly higher than that of water (0.98 cP vs. 0.92 cP, see Table S1 and Fig. S2), we expected our lysate-BSA samples to have about the same viscosity as the BSA in-water bulk viscosity. Interestingly, Pielak and co-authors demonstrated that transient interactions in the presence of BSA significantly reduce the rotational diffusion of the chymotrypsin inhibitor 2 (CI2) (23). In line with this previous result, the positive deviation of the apparent viscosity obtained from the $^1\text{H}^{\text{N}}$ linewidth analysis of GB1 from its bulk value likely indicates that the crowding of BSA resulted in some transient interactions between GB1 and BSA. Interestingly, another protein, 12.8 kDa cytochrome C (cyt c) is able to interact with a synthetic crowding agent, poly(ethylene glycol) (PEG) (8 and 20 kDa) without experiencing significant NMR linewidth changes (an increase of only 25-35% on average in NMR spectral line width at 200 and 300 g/L PEG 8000) (42). However, one should keep in mind that it is the size difference between the protein observed and an interacting crowding agent that leads to changes in the protein's NMR linewidth (assuming fast exchange on the NMR time scale between the free and bound states). In case of GB1/BSA interactions, the participation of even a very small fraction of GB1 (6.18 kDa) in complex formation with BSA (69.3 kDa) should result in significant increase of GB1 linewidths: For example, even if only 10% of the GB1 formed a transient complex with BSA in solution, the GB1 linewidths and the resultant apparent viscosity would increase by a factor of ~ 2 . Consequently, our results indicate that no more than 10% of GB1 forms a complex with BSA in the presence of 200 g/L BSA (and no more than 5% does so for 100 g/L BSA). The question arises whether weak, transient interactions between NmerA and BSA are more favored, and consequently, result in more significant line broadening?

To answer this question, we compared how GB1, dGB1, and NmerA linewidths change in the presence of BSA. We suggested that the bulk viscosity, which is roughly equal to the apparent viscosity obtained from the GB1 $^1\text{H}^{\text{N}}$ linewidth analysis, depends only on the BSA concentration. Consequently, for each protein we calculated an apparent molecular weight, $MW^{app} = MW \cdot \eta^{BSA} / \eta_0$, where MW is the protein molecular weight, η^{BSA} is the viscosity of the lysate solution in the present of 100 or 200 g/L of BSA and η_0 is the viscosity of water. Fig. 4B demonstrates that, like GB1, the dGB1 average linewidths increased as predicted from the bulk viscosity of BSA samples. The average linewidth for NmerA is slightly larger than predicted from its molecular weight. However, this difference is smaller than experimental errors and contributes no more than 30% to NmerA linewidths, which

corresponds to the presence of no more than 5% of NmerA forming a NmerA-BSA complex.

Finally, we studied whether the intracellular environment affects NmerA linewidths differently from lysate or BSA solutions. To estimate how broad the in-cell NmerA spectrum should be based on molecular crowding and viscosity considerations only, we used dataset 2 (Table S1) to obtain per residue dependencies of NmerA $^1\text{H}^{\text{N}}$ linewidth for viscosity, which were used to calculate the expected linewidths in cells (Fig. 4C). To allow for changes in molecular tumbling and possible sample inhomogeneity, we used the apparent intracellular viscosity, obtained from the $^1\text{H}^{\text{N}}$ linewidth analysis of GB1, *i.e.*, 11 times the viscosity of water (Fig. 3). In contrast with experimental observations (Fig. 1), such $^1\text{H}^{\text{N}}$ linewidth analysis predicted that NmerA in-cell should have a less broad and more intense spectrum than the in-cell dGB1 sample, and only slightly broader than the GB1 sample. This disagreement between experimental and predicted linewidths for in-cell NmerA indicates that weak interactions between the protein molecules and components of the cell cytoplasm play a more significant role for NmerA than for GB1, which result in the increased broadening in the NmerA spectrum. Consequently, inside the cell weak, transient interactions have significantly larger influence on NmerA than GB1, while in the cell lysate contributions of these interactions are about the same for both proteins. Bearing in mind that concentrations of NmerA and other cytoplasmic proteins in the lysate and inside the cell were in range of several hundred μM to a few mM, respectively, we estimate that NmerA interacts with cytoplasmic macromolecules with a K_{d} in the tens of mM range, which agrees with K_{d} of 35 mM for nonspecific interactions between CI2 and BSA found previously (43). In fact, despite its small size (7.3 kDa) CI2 is invisible by in-cell NMR spectroscopy in our hands, which is consistent with previous reports (44).

Effect of Surface Properties on Ubiquitin Diffusion

To understand how transient interactions depend on protein surface properties, we analyzed the HSQC spectra for a third model small globular protein, ubiquitin, and a triple mutant of ubiquitin, Ubi3A, which has three substitutions of surface hydrophobic amino acids to alanine, *viz.* L8A, I44A, and V70A (Fig. S9, the three residues are on a molecular surface shown to be highly involved in interactions with ubiquitin's binding partners both *in vitro* and *in vivo* (45)), which gave sharper spectra than wild-type ubiquitin in mammalian cells (28). Following our protocol, native ubiquitin inside *E. coli* cells shows no detectible resonances in an HSQC spectrum (Fig. S6). Note that our expression system yields approximately millimolar ubiquitin inside the cell (consistent with a previous report (36)). In *E. coli* cell lysate, which is about 10-fold diluted relative to the in-cell sample, ubiquitin showed a reasonably good HSQC spectrum. However, $^1\text{H}^{\text{N}}$ peak linewidths for the lysate sample were about two times larger than those predicted from its molecular size (Fig. 4A, green). Because the chemical shift values for the lysate sample closely matched those of the purified protein, we suggested that the observed line broadening is caused by weak, transient interactions between ubiquitin and other cytoplasmic proteins. Interestingly, while the average proton linewidth of ubiquitin was increased by a factor of ≈ 2 , the average peak intensity was reduced by more than 5 fold, which indicates a significant signal loss attributable to ubiquitin bound to large macromolecules that is invisible by NMR.

By contrast, for the Ubi3A mutant no line broadening or signal loss was observed in lysate, relative to expectations for the viscosity (Fig. 4A). The residues substituted with alanine in this mutant are located on a solvent-exposed hydrophobic surface of ubiquitin (the I44-containing surface, Fig. S9) that is responsible for substrate recognition and binding (45). Consequently, our observation that transient interactions become significantly weaker in Ubi3A relative to wild-type ubiquitin implicates the I44-containing surface in these binding interactions for diluted lysate samples. Interestingly, inside the cell no backbone amide

signals for either ubiquitin or its Ubi3A mutant were detected (Fig. S5B). Consequently, in contrast to the more dilute lysate samples, at higher in-cell concentrations transient interactions (which may or may not be related to the I44-containing surface) play a significant role for both ubiquitin and its Ubi3A mutant.

DISCUSSION

In *E. coli* more than 70-90% of the most abundant proteins are acidic or neutral with isoelectric points between 4 and 7 (11, 46), which strongly suggests that their surfaces are anionic at ambient pH in the cell. For GB1 molecules, which have a largely negatively-charged interface (Fig. S6), a high net charge (-4) would lead to significant self-repulsion, and more hard interactions, thereby outweighing short-range attractive forces. Consequently, GB1 diffusion is mostly determined by steric interactions. Indeed, GB1 shows almost no interactions with *E. coli* macromolecules even at mM intracellular concentrations (40), and as a result, its molecular tumbling in *E. coli* cells reports on molecular crowding and intracellular viscosity (which we estimated to be up to 8 times as high as in water) (Fig. 3). This result predicts that the slow molecular tumbling for small proteins with a molecular size of up to 13 kDa is not a limiting factor to observe them inside the cell. Indeed, 12.4-kDa dGB1, which tumbles about two times more slowly as GB1, demonstrates a high quality in-cell NMR spectrum (Fig. 1). By contrast, NMR signals for many other smaller proteins are too broad to be observed in our and others' experience, indicating the presence of attractive interactions between a protein of interest and other intracellular macromolecules. Indeed, intermolecular interactions – (including transient binding/association) lead to a change in the effective molecular size and consequently slower molecular tumbling and broader NMR signals. Intriguingly, a similar level of reduction in protein translational diffusion in the *E. coli* cytoplasm was observed by fluorescence experiments: The translational diffusion of relatively inert concatemers of green fluorescent protein (GFP) (made of 2 to 6 GFP molecules that are covalently and linearly linked; net charge of GFP at neutral pH is -5.3) in the *E. coli* cytoplasm was an order of magnitude slower than that in water (47), which qualitatively agrees with the reduced rotational diffusion of GB1 obtained in this study. Moreover, the diffusion coefficient varied with protein size roughly as would be predicted from Einstein-Stokes equation. However, protein interactions have been suggested to dramatically restrict protein mobility and lead to anomalous diffusion (1, 10, 47). For example, fusion proteins containing native *E. coli* cytoplasmic proteins attached to YFP led to a much steeper reduction in the fusion protein's mobility in *E. coli* cells than a linear prediction based on molecular mass (48), presumably because of specific interactions that are lacking in the GFP constructs.

Interestingly, Crowley et al. (40) showed that a large positive charge significantly reduces cytochrome C diffusion in lysate: Strong electrostatic interactions of the highly basic cytochrome C (with a net charge of +7) with the negatively charged *E. coli* cytosolic proteins result in formation of a complex with molecular weight greater than 150 kDa, which was undetectable by NMR; however, charge-inverted mutants in cytochrome C or elevated salt concentrations led to disruption of cytochrome C-cytosolic protein interactions. Indeed, a highly basic protein would likely interact with the negatively charged *E. coli* environment, containing acidic proteins, negatively charged nucleic acids and component cellular membranes, and these transient binding/association events would lead to a change in an effective molecular size and consequently, protein rotational diffusion.

In contrast to GB1 and cytochrome C, NmerA and ubiquitin have nearly no net charge (Table S2, Fig. S8); and consequently, short-range attractive forces between hydrophobic side-chains play a decisive role in their propensity for self-association and interactions with other cytoplasmic molecules. It has been suggested that 'stickiness' of a protein, *i.e.*, its

tendency to form weak, transient nonspecific interactions and the strength of these interactions, largely depends on the total hydrophobic surface that is screened from water upon formation of a nonspecific complex (7, 15, 49). Even though about the same number of hydrophobic residues are displayed on the surfaces of NmerA and ubiquitin, they are distributed differently: While for ubiquitin almost all surface hydrophobic residues are located on one hydrophobic patch near the C-terminus, for NmerA surface hydrophobic residues are more dispersed and less clustered (Fig. S8). We conclude that these differences in surface distribution of hydrophobic groups for the two proteins lead to different degrees of 'stickiness'. Our results demonstrate that they indeed show a significantly different propensity to form weak, transient complexes. NmerA is involved in relatively weak transient interactions with a K_d of tens of mM, such that complexes are observed only under over-expression conditions inside the cells, but not in the diluted lysate. For ubiquitin, transient interactions have a significant impact even in the cell lysate, where cytoplasmic proteins as well as ubiquitin itself were diluted about ten times compared to in cells, and therefore the weak interactions for ubiquitin are estimated to be about a hundred times stronger than those for NmerA (*i.e.*, they have K_d of hundreds of μ M). We found that three central residues of ubiquitin, *i.e.* Leu8, Ile44 and Val70, which comprise a hydrophobic patch (named the I44-containing surface (45), Fig. S9), play a key role in its interactions. Indeed, mutating these residues to Ala significantly weakened transient interactions. Strikingly, in eukaryotic cells the hydrophobic side chains of the I44-containing surface interact with hydrophobic surfaces of ubiquitin's physiological partners with an affinity corresponding to $K_d = 0.1 - 1$ mM (50), which is very close to the affinity observed for nonspecific interactions in the *E. coli* cytoplasm. Interestingly, the ability of ubiquitin to form relatively stable and promiscuous hydrophobic complexes was shown to be general (51). Consequently, it is not surprising that ubiquitin formed weak, transient interactions in the *E. coli* cytoplasm even in the absence of any known ubiquitin substrate. This salient feature of ubiquitin seems to be important for its functionality under physiological conditions: While most proteins take part in associations with a small number of partners, ubiquitin interacts with more than 150 eukaryotic cellular binding partners from more than twenty different families (45). The question arises as to what mechanisms facilitate promiscuous hydrophobic interactions for ubiquitin? It has been shown that conformational dynamics on the hydrophobic patch on the I44-containing surface play an essential role for recognition of different structures by ubiquitin (52). Indeed, ubiquitin initially binds by a conformational selection mechanism, which provides a wide range of binding partners and results in promiscuous interactions (53). However, for physiological interactions the subsequent induced fit results in strong complexes and high specificity.

Many questions are raised by the observations in our study and related work in the literature: What factors define protein 'stickiness'? How are the affinities and dynamics of weak interactions determined? Which interactions are specific (*i.e.*, physiologically productive) and which are nonspecific (of potential physiological harm)? Cells must avoid transient nonspecific interactions. Consequently, proteins with high propensity for promiscuous nonfunctional interactions are less abundant in cells, whereas more 'inert' proteins can be present at higher concentrations. Consistent with this, it has recently been shown that the intracellular abundance of a protein anti-correlates with the number of its promiscuous nonfunctional partners (7). Also, the most abundant *E. coli* proteins are mostly acidic so that they repel each other and other intracellular components. In general, we suggest that, for highly charged proteins, short-range hydrophobic forces are significantly overshadowed by electrostatic repulsions. Indeed, the other most abundant intracellular macromolecules, such as nucleic acids and anionic components of cellular membranes are negatively charged. Consequently, acidic proteins together with anionic components of cellular membranes result in the negatively charged *E. coli* environment, which is, in turn, inert to acid proteins. As a result, nonspecific interactions are minimized, and in the absence of specific

interactions, diffusion of negatively charged proteins in the cytoplasm is expected to be relatively fast and affected mostly by molecular crowding. Interestingly, to avoid nonspecific interactions under high physiological concentrations acidic proteins likely sacrifice their stability. Indeed, a net charge on a protein leads to lower stability, and consequently, under physiological conditions (but in the absence of transient interactions) many acidic proteins including GB1 are destabilized (54). However, inside the cell nonspecific interactions may result in significant protein destabilization and even prevent folding of global proteins (38). Consequently, loss in protein stability for acidic protein at physiological pH can be compensated by the absence of nonspecific interactions.

As net charge becomes smaller, hydrophobic forces become more significant and transient attractive interactions are enhanced, resulting in K_d 's of hundreds of μM to tens of mM. The resulting interactions could be tuned to favor productive interactions, for example in metabolic or signaling pathways. However, the balance between such specific interactions and nonspecific interactions is crucial, as a substantial fraction of nonspecific complexes may form between proteins of low net charge under intracellular conditions, and the concentration of a protein in cells becomes an important factor to insure that it form specific and not nonspecific interactions (7). Taken together, our results suggest that the likelihood of nonspecific interactions (and apparently protein abundance) is determined not by one factor (*e.g.*, the amount of hydrophobic surface exposed on a protein), but by several factors, such as an overall charge, distribution of hydrophobic residues, and conformational flexibility.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Cathy H. Wu for her generous support of this work and Tatyana Polenova for her assistance with experiments performed at the University of Delaware. We acknowledge Weiguo Hu and Steve Bai for their help with the NMR instrumentation. We are grateful to Kylie Walters, Volker Dötsch and many others for their kind sharing of many constructs we used in this study.

Abbreviations

GB1	protein G B1 domain
NmerA	N-terminal domain of mercuric ion reductase (MerA)
HSQC	heteronuclear single quantum coherence
S/N	signal-to-noise ratio
TROSY	transverse relaxation optimized spectroscopy

References

1. Verkman AS. Solute and macromolecule diffusion in cellular aqueous compartments. *Trends Biochem. Sci.* 2002; 27:27–33. [PubMed: 11796221]
2. Mika JT, Poolman B. Macromolecule diffusion and confinement in prokaryotic cells. *Curr. Opin. Biotechnol.* 2011; 22:117–126. [PubMed: 20952181]
3. Elcock AH. Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Curr. Opin. Struct. Biol.* 2010; 20:196–206. [PubMed: 20167475]
4. Gershenson A, Gierasch LM. Protein folding in the cell: challenges and progress. *Curr. Opin. Struct. Biol.* 2011; 21:32–41. [PubMed: 21112769]

5. Zhou H-X, Rivas G, Minton AP. Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* 2008; 37:375–397. [PubMed: 18573087]
6. Ellis RJ. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr. Opin. Struct. Biol.* 2001; 11:114–119. [PubMed: 11179900]
7. Heo MY, Maslov S, Shakhnovich E. Topology of protein interaction network shapes protein abundances and strengths of their functional and nonspecific interactions. *Proc. Natl. Acad. Sci. U.S.A.* 2011; 108:4258–4263. [PubMed: 21368118]
8. Yadav S, Shire SJ, Kalonia DS. Factors affecting the viscosity in high concentration solutions of different monoclonal antibodies. *J. Pharm. Sci.* 2010; 99:4812–4829. [PubMed: 20821382]
9. Zorrilla S, Hink MA, Visser AJ, Lillo MP. Translational and rotational motions of proteins in a protein crowded environment. *Biophys. Chem.* 2007; 125:298–305. [PubMed: 17007994]
10. Elowitz MB, Surette MG, Wolf PE, Stock JB, Leibler S. Protein mobility in the cytoplasm of *Escherichia coli*. *J. Bacteriol.* 1999; 181:197–203. [PubMed: 9864330]
11. 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. [PubMed: 19487732]
12. McConkey EH. Molecular evolution, intracellular organization, and the quinary structure of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 1982; 79:3236–3240. [PubMed: 6954476]
13. Rowe AJ. Ultra-weak reversible protein-protein interactions. *Methods.* 2011; 54:157–166. [PubMed: 21338686]
14. Vaynberg J, Qin J. Weak protein-protein interactions as probed by NMR spectroscopy. *Trends in Biotechnology.* 2006; 24:22–27. [PubMed: 16216358]
15. Deeds EJ, Ashenberg O, Shakhnovich EI. A simple physical model for scaling in protein-protein interaction networks. *Proc. Natl. Acad. Sci. U.S.A.* 2006; 103:311–316. [PubMed: 16384916]
16. McGuffee SR, Elcock AH. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput. Biol.* 2010; 6:e1000694. [PubMed: 20221255]
17. Miklos AC, Sarkar M, Wang YQ, Pielak GJ. Protein crowding tunes protein stability. *J. Am. Chem. Soc.* 2011; 133:7116–7120. [PubMed: 21506571]
18. Burz DS, Dutta K, Cowburn D, Shekhtman A. Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). *Nat. Methods.* 2006; 3:91–93. [PubMed: 16432517]
19. 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; 458:102–105. [PubMed: 19262674]
20. Selenko P, Wagner G. Looking into live cells with in-cell NMR spectroscopy. *J. Struct. Biol.* 2007; 158:244–253. [PubMed: 17502240]
21. Serber Z, Selenko P, Hansel R, Reckel S, Lohr F, Ferrell JE, Wagner G, Dotsch V. Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. *Nat. Protoc.* 2006; 1:2701–2709. [PubMed: 17406526]
22. Serber Z, Ledwidge R, Miller SM, Dotsch 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. [PubMed: 11552796]
23. Wang Y, Li C, Pielak GJ. Effects of proteins on protein diffusion. *J. Am. Chem. Soc.* 2010; 132:9392–9397. [PubMed: 20560582]
24. Barnes CO, Monteith WB, Pielak GJ. Internal and global protein motion assessed with a fusion construct and in-cell NMR spectroscopy. *ChemBioChem.* 2011; 12:390–391. [PubMed: 21290539]
25. Huth JR, Bewley CA, Jackson BM, Hinnebusch AG, Clore GM, Gronenborn AM. Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. *Protein Sci.* 1997; 6:2359–2364. [PubMed: 9385638]
26. Zhang N, Wang Q, Ehlinger A, Randles L, Lary JW, Kang Y, Haririnia A, Storaska AJ, Cole JL, Fushman D, Walters KJ. Structure of the s5a:k48-linked diubiquitin complex and its interactions with rpn13. *Mol. Cell.* 2009; 35:280–290. [PubMed: 19683493]

27. Piotrowski J, Beal R, Hoffman L, Wilkinson KD, Cohen RE, Pickart CM. Inhibition of the 26 S proteasome by polyubiquitin chains synthesized to have defined lengths. *J. Biol. Chem.* 1997; 272:23712–23721. [PubMed: 9295315]
28. Inomata K, Ohno A, Tochio H, Isogai S, Tenno T, Nakase I, Takeuchi T, Futaki S, Ito Y, Hiroaki H, Shirakawa M. High-resolution multidimensional NMR spectroscopy of proteins in human cells. *Nature.* 2009; 458:106–109. [PubMed: 19262675]
29. Byeon I-JL, Louis JM, Gronenborn AM. A captured folding intermediate involved in dimerization and domain-swapping of GB1. *J. Mol. Biol.* 2004; 340:615–625. [PubMed: 15210358]
30. Pervushin K, Riek R, Wider G, Wuthrich K. Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. U.S.A.* 1997; 94:12366–12371. [PubMed: 9356455]
31. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* 1995; 6:277–293. [PubMed: 8520220]
32. Keller, RLJ. The computer aided resonance assignment tutorial. 1st ed. CANTINA Verlag; 2004.
33. Cavanagh, J.; Fairbrother, WJ.; Palmer, AG., III; Skelton, NJ.; M., R. Protein NMR spectroscopy: Principles and practice. 2nd ed. Elsevier Academic Press; 2007.
34. Markus MA, Dayie KT, Matsudaira P, Wagner G. Effect of deuteration on the amide proton relaxation rates in proteins - Heteronuclear NMR experiments on villin 14T. *J. Magn. Reson., ser. B.* 1994; 105:192–195. [PubMed: 7952934]
35. Walsh JD, Meier K, Ishima R, Gronenborn AM. NMR studies on domain diffusion and alignment in modular GB1 repeats. *Biophys. J.* 2010; 99:2636–2646. [PubMed: 20959105]
36. Barnes CO, Pielak GJ. In-cell protein NMR and protein leakage. *Proteins.* 2011; 79:347–351. [PubMed: 21120863]
37. Miklos AC, Li CG, Sharaf NG, Pielak GJ. Volume exclusion and soft interaction effects on protein stability under crowded conditions. *Biochemistry.* 2010; 49:6984–6991. [PubMed: 20672856]
38. Schlesinger AP, Wang Y, Tadeo X, Millet O, Pielak GJ. Macromolecular crowding fails to fold a globular protein in cells. *J. Am. Chem. Soc.* 2011; 133:8082–8085. [PubMed: 21534566]
39. 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; 132:321–327. [PubMed: 20050707]
40. Crowley PB, Chow E, Papkovskaia T. Protein interactions in the Escherichia coli cytosol: An impediment to in-cell NMR spectroscopy. *ChemBioChem.* 2011; 12:1043–1048. [PubMed: 21448871]
41. Palmer AG, Kroenke CD, Loria JP. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Nuclear Magnetic Resonance of Biological Macromolecules, Pt. B.* 2001; 339:204–238.
42. Crowley PB, Brett K, Muldoon J. NMR spectroscopy reveals cytochrome c-poly(ethylene glycol) interactions. *ChemBioChem.* 2008; 9:685–688. [PubMed: 18260072]
43. Li C, Pielak GJ. Using NMR to distinguish viscosity effects from nonspecific protein binding under crowded conditions. *J. Am. Chem. Soc.* 2009; 131:1368–1369. [PubMed: 19140727]
44. Li C, Charlton LM, Lakkavaram A, Seagle C, Wang G, Young GB, Macdonald JM, Pielak GJ. 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. [PubMed: 18419123]
45. Dikic I, Wakatsuki S, Walters KJ. Ubiquitin-binding domains - from structures to functions. *Nat. Rev. Mol. Cell Biol.* 2009; 10:659–671. [PubMed: 19773779]
46. Link AJ, Robison K, Church GM. Comparing the predicted and observed properties of proteins encoded in the genome of Escherichia coli K-12. *Electrophoresis.* 1997; 18:1259–1313. [PubMed: 9298646]
47. Nenninger A, Mastroianni G, Mullineaux CW. Size Dependence of Protein Diffusion in the Cytoplasm of Escherichia coli. *Journal of Bacteriology.* 2010; 192:4535–4540. [PubMed: 20581203]

48. Kumar M, Mommer MS, Sourjik V. Mobility of Cytoplasmic, Membrane, and DNA-Binding Proteins in *Escherichia coli*. *Biophys. J.* 2010; 98:552–559. [PubMed: 20159151]
49. Hoskins J, Lovell S, Blundell TL. An algorithm for predicting protein-protein interaction sites: Abnormally exposed amino acid residues and secondary structure elements. *Protein Sci.* 2006; 15:1017–1029. [PubMed: 16641487]
50. Fisher RD, Wang B, Alam SL, Higginson DS, Robinson H, Sundquist WI, Hill CP. Structure and ubiquitin binding of the ubiquitin-interacting motif. *J. Biol. Chem.* 2003; 278:28976–28984. [PubMed: 12750381]
51. Van Horn WD, Ogilvie ME, Flynn PF. Reverse micelle encapsulation as a model for intracellular crowding. *J. Am. Chem. Soc.* 2009; 131:8030–8039. [PubMed: 19469539]
52. Boehr DD, Nussinov R, Wright PE. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* 2009; 5:789–796. [PubMed: 19841628]
53. Wlodarski T, Zagrovic B. Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proc. Natl. Acad. Sci. U.S.A.* 2009; 106:19346–19351. [PubMed: 19887638]
54. Lindman S, Xue WF, Szczepankiewicz O, Bauer MC, Nilsson H, Linse S. Salting the charged surface: pH and salt dependence of protein G B1 stability. *Biophys. J.* 2006; 90:2911–2921. [PubMed: 16443658]

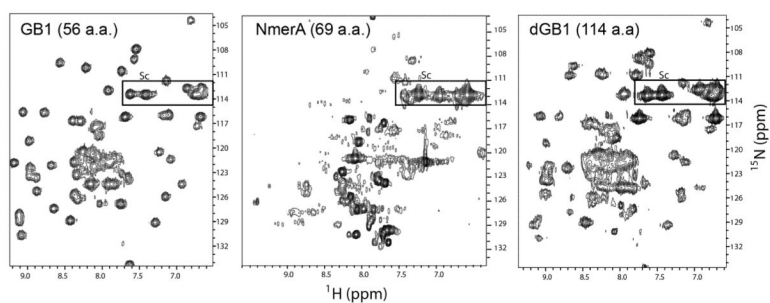


Figure 1. Effect of protein size on the quality of in-cell NMR spectra. In-cell $[^1\text{H}, ^{15}\text{N}]$ HSQC spectra of GB1, NmerA, and GB1-L1-GB1 (dGB1). Spectra were measured with 8, 32, and 16 scans, corresponding to total acquisition times of 10, 40 and 20 minutes for GB1, NmerA, and dGB1, respectively. The boxed regions show sharp peaks arising from flexible NH_2 side-chains.

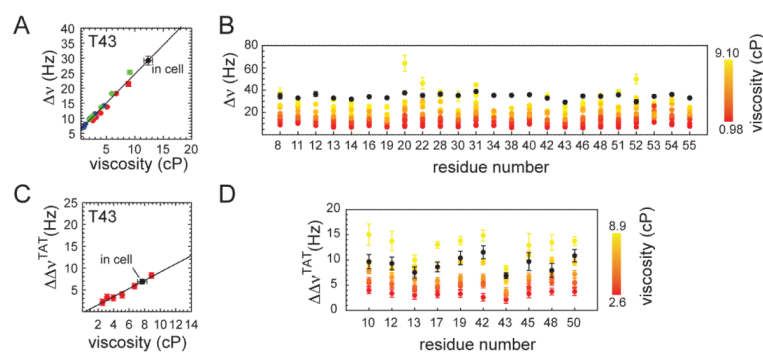


Figure 2.

Determination of the apparent intracellular viscosity. (A, C) A representative example of a measured (A) $^1\text{H}^{\text{N}}$ linewidth, $\Delta\nu$, and (C) difference between ^{15}N TROSY and anti-TROSY linewidths, $\Delta\Delta\nu^{\text{TAT}}$, as a function of the solution viscosity for residue T43 in GB1. Red, blue and green colors correspond to datasets 1, 2 and 3 (Table S1), respectively. Solid lines show the best linear fit obtained using all titration data (A) or only dataset 1 (C). (B, D) Histograms showing $\Delta\nu$ (B) and $\Delta\Delta\nu^{\text{TAT}}$ (D) as a function of GB1 residue number. Colors indicate the bulk viscosity in solution from lowest (red) to highest (yellow). In-cell data are shown by black and labeled.

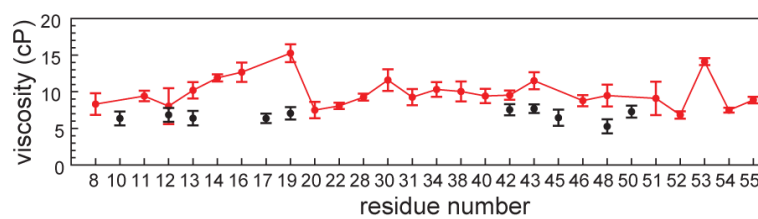


Figure 3. Apparent intracellular viscosity as a function of residue number, obtained from $^1\text{H}^{\text{N}}$ linewidth (red) and TROSY/anti-TROSY (black) analysis on GB1.

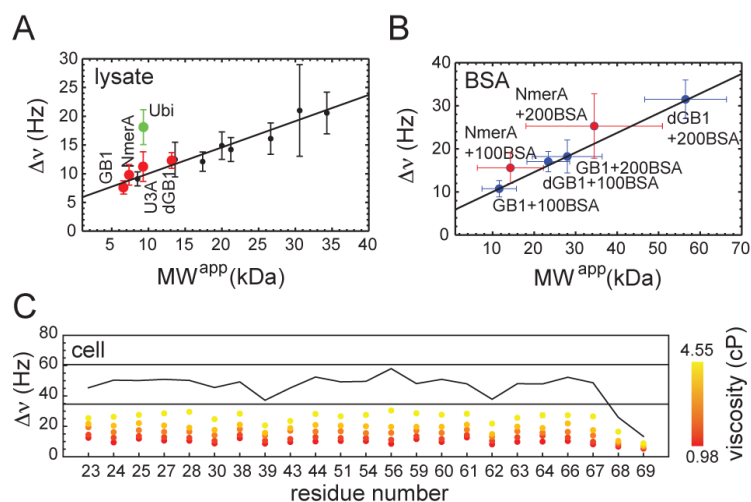


Figure 4.

Effect of intracellular environments and molecular crowding on protein mobility. (A) Average $^1\text{H}^{\text{N}}$ linewidths in lysate samples of GB1, NmerA, Ubi3A (U3A in the figure), and dGB1 (red), ubiquitin (green), and average $^1\text{H}^{\text{N}}$ linewidths (black) obtained from glycerol titrations of GB1 and dGB1 lysates from datasets 2 and 3 (Table S1) as a function of an apparent molecular weight, $MW^{\text{app}} = MW \cdot \eta / \eta_0$, where MW , η , and η_0 are the protein molecular weight, the apparent sample viscosity, and the viscosity of water, respectively. A solid black line is the best linear fit of GB1/dGB1 glycerol titrations. (B) Average $^1\text{H}^{\text{N}}$ linewidths for GB1 and dGB1 (blue), and NmerA (red) lysate samples in the presence of 100 and 200 g/L BSA as a function of an apparent molecular weight, $MW^{\text{app}} = MW \cdot \eta / \eta_0$, where MW , η , and η_0 are a protein molecular weight, the bulk viscosity and the viscosity of water, respectively. A solid black line is the best linear fit of GB1/dGB1 glycerol titrations (same as in Fig. 4A). (C) Histograms showing experimental $^1\text{H}^{\text{N}}$ linewidths, $\Delta\nu$, as a function of NmerA residue number. Colors indicate viscosity in solution from lowest (red) to highest (yellow). Horizontal lines show the average $^1\text{H}^{\text{N}}$ linewidths for in-cell GB1 and dGB1 samples. A black solid line shows theoretically predicted NmerA linewidths expected for intracellular environment 11 times as viscous as water.