BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Exploring Weak, Transient Protein−Protein Interactions in Crowded In Vivo Environments by In-Cell Nuclear Magnetic Resonance Spectroscopy

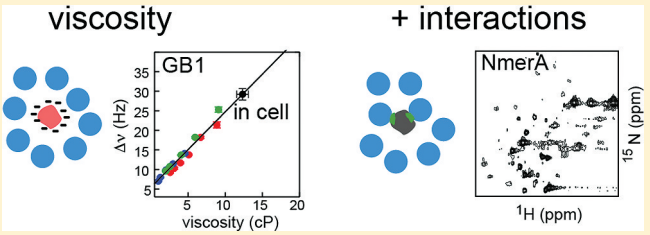
Qinghua Wang,[†,§] Anastasia Zhuravleva,[†] and Lila M. Gierasch*[,†,‡]

[†]Department of Biochemistry and Molecular Biology and [‡]Department of Chemistry, University of Massachusetts, Amherst, Massachusetts 01003, United States

[§]Center for Bioinformatics and Computational Biology, Delaware Biotechnology Institute, University of Delaware, Newark, Delaware 19711, United States

Supporting Information

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 that preserve these interactions and provide information about them are needed. In-cell nuclear magnetic resonance (NMR) spectroscopy is an attractive method for studying 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; 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 *Escherichia coli* cells using two-dimensional 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 8 times that of water, too low to be a limiting factor for observation of 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.



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 >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 refs 3−5). Many of these studies have relied on model crowding agents, either inert synthetic polymers or model proteins such as bovine serum albumin, 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, ensuring 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) or 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

Received: August 12, 2011
Revised: September 22, 2011
Published: September 26, 2011

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 values of >1 μM,[13,14] are in fact an unavoidable consequence of intracellular crowding, and thus, evolutionary selection tunes their physiological roles in the living cell to allow their regulation and to suppress undesirable associations.

It is clearly important to improve our understanding of the nature of weak transient interactions between proteins in cells 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 are 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 the number of interactions.[15] In a recent computational simulation of the diffusion behavior of proteins in the *Escherichia coli* cytoplasm, McGuffee and Elcock[16] demonstrated that steric and electrostatic interactions between the 50 most abundant *E. coli* proteins are not sufficient for predicting 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 the 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 albumin.[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 the extent to which the intracellular viscosity and/or transient interactions modulate protein intracellular mobility remains unclear.

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 two-dimensional (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 8 times that of water, which would not be a limiting factor for observing 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 polymerase chain reaction (PCR) primers: 5′-CCT TCA CGG TAA CCG AAT AGG TTC CGC GTG GAT CC-3′ and 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 subcloning the synthesized gene with triple Ala mutations into a pET16b vector, but neither this mutant nor the 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 encoding full-length GB1 (amino acids 1−56) was amplified via PCR from pET21a-GB1, digested, and ligated into pET16b-GB1-C.

For construction of the GB1-L_{15}-NmerA fusion, DNA encoding full-length GB1 (amino acids 1−56) and 15 amino acids [i.e., -SerGlySer$(Gly)_{11}$His-] at the C-terminus was amplified via PCR and ligated into pET16b-NmerA-C to obtain a GB1-NmerA fusion construct [i.e., GB1 and NmerA linked through a 15-amino acid linker, -SerGlySer$(Gly)_{11}$His-]. The final construct was confirmed by DNA sequencing.

NMR Sample Preparation. *Preparation of Cell Samples Overexpressing the Uniformly ^{15}N-Labeled 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 of Luria-Bertani medium (LB) supplemented with 100 mg/L ampicillin by inoculating the culture with a colony from newly transformed cells. Then, 100 mL of LB was inoculated with a 3 mL overnight culture and grown to an OD_{600} of ∼0.8. The cells were harvested by gentle centrifugation (∼1400*g* for 10 min) and resuspended to an OD_{600} of 0.5−0.6 in 100 mL of M9 containing [U-^{15}N]-ammonium chloride (1.0 g/L) and glucose (4 g/L) as the sole nitrogen and carbon sources, respectively, along with 100 mg/L ampicillin. The culture was incubated at 37 °C for 10−15 min before protein expression was induced by adding 1/2000 of a culture volume of 1.0 M IPTG. Induction was allowed to proceed for 3 h (if not otherwise specified). Following the 3 h induction, the 100 mL culture was centrifuged gently (∼1400*g* for 10−15 min), after which the resulting cell pellets were left sitting in ∼500 μL of M9 minimal medium together with ∼50 μL of D_2O for ∼15 min, and then the cells were very gently resuspended well with a pipet. 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 described above were flash-frozen with liquid nitrogen and stored in −80 °C. Then they were thawed on ice and 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 minicentrifuge at 13K rpm for ∼10 min 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 described in ref 29. Basically, the cell slurry (∼40 mL, harvested from a 2 L M9 minimal medium culture) (the cell sample was neither sonicated nor microfluidized) was heated at 80 °C for 5 min, immediately chilled on ice for 10 min, and then centrifuged at 16000 rpm (Beckman JS 5.3 rotor) for 30 min at 4 °C. The supernatant was passed through a 0.45 μm syringe filter, dialyzed against 4 L of Milli-Q water at 4 °C twice (for at least 1 h 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 the 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 $^{15}NH_4Cl$. The cells were induced to overexpress ubiquitin via addition of 0.5 mM IPTG and were grown for an additional 4 h, harvested by centrifugation, and stored at −80 °C. The cell extract was affinity-purified with 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 conducted on Bruker AVANCE 600 MHz NMR spectrometers equipped with cryoprobes at 298 K. All ^{1}H−^{15}N HSQC data sets collected for the purpose of measurements of $^1H^N$ line widths were acquired with a ^{1}H resolution of ∼7−8 Hz. For each FID, 8−32 transients were collected, corresponding to total experiment durations of 10−40 min. No apodization function, baseline correction, or linear prediction was applied in the proton dimension to avoid an artificial disturbance of $^1H^N$ line shapes; zero filling was applied to set a digital resolution in the ^{1}H dimension of ∼1 Hz.

All ^{15}N TROSY and anti-TROSY data were measured using a 2D ^{1}H−^{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 a ^{15}N resolution of ∼9 Hz, and zero filling was applied to set the digital resolution to ∼2 Hz. For in-cell NMR samples, the total experiment duration is 2 h 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 entries 7280, 16208, and 6466, respectively) using CARA.[32]

$^1H^N$ Line Width Data Analysis. For line width analysis, a set of one-dimensional (1D) proton slices from a ^{1}H−^{15}N HSQC spectrum was selected for nonoverlapping, assigned peaks. Each slice was positioned at the peak maximum in the ^{15}N dimensions and centered about the peak maximum in the ^{1}H dimension; the width of each slice was ∼3 times the peak line width. ^3J(HN−HA) coupling was included in the simulation of the ^{1}H line shapes, and experimental data points were fitted to a sum of two Lorentzians:

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

where $(\nu_0 + {}^3J/2)$ and $(\nu_0 - {}^3J/2)$ are ^{1}H frequencies (in hertz) of the maxima of a peak doublet separated by the ^3J(HN−HA) coupling constant, $\Delta\nu$ is the full width at half-height, and A_1 and A_2 are normalization constants. ^3J(HN−HA) constants for individual residues were fixed at values obtained as described below. Uncertainties of the $^1H^N$ line widths 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 ^3J(HN−HA) Constants. For small proteins in buffer solution and at low glycerol concentrations, ^3J(HN−HA) coupling contributes significantly to $^1H^N$ line shapes, resulting in resolved or unresolved peak doublets (Figure S1 of the Supporting Information). We used GB1 and NmerA lysate samples and a purified His-tagged ubiquitin sample to obtain ^3J(HN−HA) constants for individual residues. Experimental data from two or three different samples were fitted using eq 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(HN−HA) values, they were compared with the ^3J(HN−HA) values predicted from corresponding three-dimensional protein structures (Figure S1 of the Supporting Information).

Viscosity Dependence of $^1H^N$ Line Widths in GB1. To determine how the GB1 spectra change with viscosity, we obtained $^1H^N$ line widths for individual peaks in a corresponding ^{1}H−^{15}N HSQC spectrum (see above). The interpretation of absolute values for $^1H^N$ line widths, 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 $^1H^N$ line widths and solution viscosity. Three data sets were used for glycerol titrations: purified GB1 (35−60 wt % glycerol, data set 1), a GB1/NmerA mixed lysate (0−40% glycerol, data set 2), and dGB1 lysate (0−40% glycerol, data set 3) (Table S1 of the Supporting Information). For all titrations, we used d_8-glycerol (99% d_8, from Cambridge Isotope Laboratories) (Figure S2 of the Supporting Information) to prevent strong background signals in NMR spectra. The solution viscosity for each sample was calculated on the basis of known glycerol concentrations (see the 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 concentration (Table S1, Figure S3, and Figure S4 of the Supporting Information). For each sample from data sets 1−3, we obtained $^1H^N$ line widths for individual peaks, which corresponded to unambiguously assigned GB1 residues. For final analysis, we chose 26 GB1 residues that showed reliable data in at least two data sets (i.e., they have nonoverlapped peaks in HSQC spectra and $^1H^N$ line width errors of less than 20%, and each data set exhibited a linear dependence of $^1H^N$ line width on viscosity). To obtain $^1H^N$ line widths versus viscosity calibration slopes for each residue, $^1H^N$ line widths 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 $^1H^N$ line widths in the in-cell HSQC spectrum (see below).

Determination of the Intracellular Viscosity Using $^1H^N$ Line Width Analysis for GB1. $^1H^N$ line widths 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 $^1H^N$ line widths 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 $^1H^N$ Line Widths for In-Cell NmerA. First, we fitted $^1H^N$ line widths for 23 NmerA residues as a function of solution viscosity for the GB1/NmerA mixture lysate data set (0−40% glycerol, data set 2 from Table S1 of the Supporting Information) to obtain viscosity calibration slopes, and on the basis of the slopes, we calculated $^1H^N$ line widths expected for the intracellular viscosity obtained from the $^1H^N$ line width analysis of GB1 (see above).

BSA Titration of the GB1/NmerA Lysate Sample. To mimic intracellular molecular crowding with a protein crowding agent, $^1H^N$ line widths 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 BSA (ACROS Organics). To obtain an average line width for each protein, we averaged the resulting values 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 using the relationship $MW^{app} = MW \times \eta^{BSA}/\eta_0$, where MW is the protein molecular weight and 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 $^1H^N$ line width analysis for GB1 (see above).

Average $^1H^N$ Line Width Analysis for Lysate Samples. To understand whether molecular weight is the primary determinant of protein $^1H^N$ line widths or whether other factors (e.g., transient interactions) significantly affect line widths, and how these contributions vary from protein to protein, we plotted the average $^1H^N$ line width for GB1, NmerA, ubiquitin, and Ubi3A as a function of protein molecular weight. $^1H^N$ line widths of individual residues for GB1, NmerA, ubiquitin, and Ubi3A were obtained from HSQC spectra of their lysates. To obtain an average line width 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 using the relationship $MW^{app} = MW \times \eta^{lysate}/\eta_0$, where MW is the protein molecular weight and 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 nonoverlapped 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 ∼3 times greater than the peak line width. Experimental data points were fitted to a Lorentzian function:

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

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

The difference in line widths between TROSY and anti-TROSY lines, $\Delta\Delta\nu^{TAT}$, was calculated using the relationship $\Delta\Delta\nu^{TAT} = \Delta\nu^{anti\text{-}TROSY} - \Delta\nu^{TROSY}$; uncertainties were estimated as $[(\sigma^{TROSY})^2 + (\sigma^{anti\text{-}TROSY})^2]^{1/2}$, 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, data set 1 from Table S1 of the Supporting Information). 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.

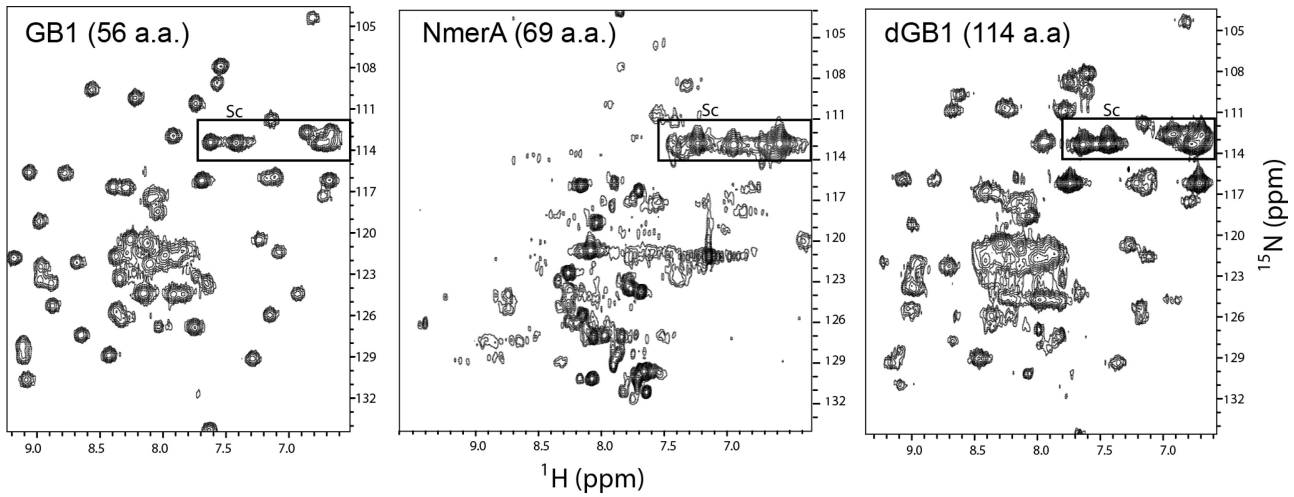


Figure 1. Effect of protein size on the quality of in-cell NMR spectra. In-cell $^{1}H-^{15}N$ HSQC spectra of GB1, NmerA, and GB1-L1-GB1 (dGB1). Spectra were measured with 8, 32, and 16 scans, corresponding to total acquisition times of 10, 40, and 20 min for GB1, NmerA, and dGB1, respectively. The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

All line width analysis was 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 the low stability of in-cell samples and leakage of protein into the extracellular medium[36] (see Figure S5 of the Supporting Information for more details). 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] and 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 of the Supporting Information). Despite the fact that all these proteins demonstrated high-quality NMR spectra in vitro, we found that only two of them, GB1 and NmerA, exhibited a reasonably good in-cell NMR spectrum (Figure 1 and Figure S7 of the Supporting Information), while ubiquitin is invisible in cells (Figures S6 and S7 of the Supporting Information). (Our results for ubiquitin agree with those of Li et al., who reported that they saw no signals for ubiquitin in *E. coli* cells,[39] but not with those of 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 recording 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 (Figure 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 in-cell and in vitro peak positions, 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-L_{15}-NmerA fusion protein (i.e., the two domains connected by a 15-amino acids linker) 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-L_{15}-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 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 NH_2 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 (Figure 1). Moreover, previously, it was shown that deuteration, bringing about a reduction in the level of $^{1}H^{N}$ 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

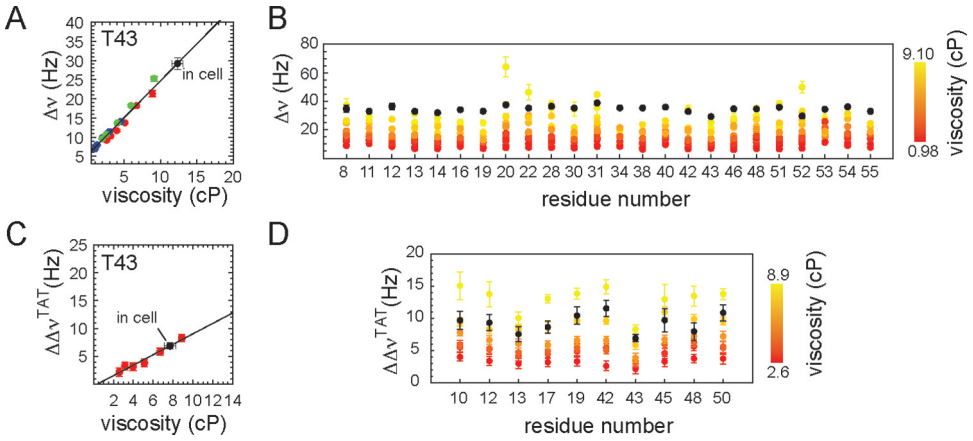


Figure 2. Determination of the apparent intracellular viscosity. (A and C) Representative example of a measured (A) $^1H^N$ line width, $\Delta\nu$, and (C) difference between ^{15}N TROSY and anti-TROSY line widths, $\Delta\Delta\nu^{TAT}$, as a function of the solution viscosity for residue T43 in GB1. Red, blue, and green correspond to data sets 1−3 (Table S1 of the Supporting Information), respectively. Solid lines show the best linear fit obtained using all titration data (A) or only data set 1 (C). (B and D) Histograms showing $\Delta\nu$ (B) and $\Delta\Delta\nu^{TAT}$ (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 colored black and labeled.

to broadening from protein rotational diffusion and other possible contributing factors that affect protein NMR line widths. 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 surface properties similar to those of GB1, but a rotational correlation time that is approximately twice that of GB1 (Figure S4 of the Supporting Information). As expected, with the increased rotational correlation time, the in-cell NMR spectrum of dGB1 presented a lower signal-to-noise ratio (S) and broader peaks than the GB1 spectrum (Figure 1). However, in cells, this 12.5 kDa dGB1 protein construct exhibited much better resolved spectra than 6.91 kDa NmerA (Figure 1). Moreover, ubiquitin, which is significantly smaller than dGB1 (Table S2 of the Supporting Information), exhibited no backbone amide signals in cells (Figures S6 and S7 of the Supporting Information). All these findings indicate that the high viscosity of the intracellular environment and consequent reduction in mobility of proteins, which increases with molecular size, is an important factor but not necessarily the limiting factor determining their NMR visibility in cells.

Determination of the Apparent Intracellular Viscosity Using $^1H^N$ Line Widths 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 to protein diffusion in the *E. coli* cell. There are several NMR techniques such as pulsed field gradient measurements and relaxation experiments that allow accurate and detailed characterization of protein diffusion. Unfortunately, most of them are not suitable for in-cell samples because of poor sample stability and the low signal-to-noise ratio of in-cell NMR spectra. By contrast, $^1H^N$ line widths, which for many rigid protein systems are strongly correlated with overall rotational correlation time,[33] can be accurately estimated from a $^1H-^{15}N$ HSQC spectrum and, consequently, can be used to probe sample viscosity. The interpretation of absolute values for $^1H^N$ line widths, 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 $^1H^N$ line widths and solution viscosity, using three data sets: purified GB1, a cell lysate containing a mixture of GB1 and NmerA, and a cell lysate containing the dGB1 protein construct (see Materials and Methods and Table S1 of the Supporting Information). The viscosity dependencies of $^1H^N$ line widths obtained were used to estimate the apparent intracellular viscosity (Figure 2A,B). We found that the $^1H^N$ line widths obtained for GB1 in the *E. coli* cell corresponded to a viscosity ~11 ± 2 times that of water (Figure 3). This estimate of in-cell viscosity represents an upper

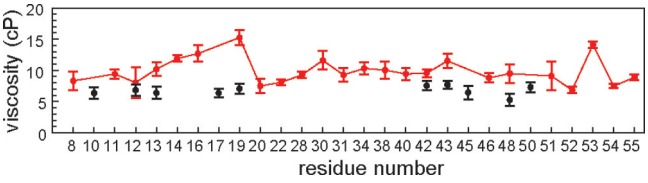


Figure 3. Apparent intracellular viscosity as a function of residue number, obtained from $^1H^N$ line width (red) and TROSY/anti-TROSY (black) analysis on GB1.

limit, because other factors might also affect in-cell $^1H^N$ line widths, including contributions linked to rotational diffusion, such as a change in 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 on the microsecond to millisecond time scale, as well as sample and magnetic field inhomogeneity. To exclude contributions not linked to molecular tumbling, we performed a line width analysis of GB1 using ^{15}N TROSY and anti-TROSY spectra. The difference between ^{15}N TROSY and anti-TROSY relaxation and, consequently, their line widths 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 line widths, $\Delta\Delta\nu^{TAT}$, is unaffected by contributions from chemical exchange and sample or magnetic field inhomogeneity. To avoid interpreting absolute $\Delta\Delta\nu^{TAT}$ values, we examined the viscosity dependence of $\Delta\Delta\nu^{TAT}$. Specifically, we performed glycerol titrations of purified GB1 (data set 1, Table S1 of the Supporting Information) and found

that $\Delta\Delta\nu^{TAT}$ values for individual GB1 residues can be fitted as a linear function of sample viscosity (Figure 2C,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 ~30% lower than the one estimated from the $^1H^N$ line width analysis (Figure 3), indicating that ~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 ~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 the change in 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 masses of ≤13 kDa will be detectable via in-cell NMR spectroscopy. By contrast, even though the molecular mass 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 line widths in the cell must be considered. Recently, it was shown that transient nonspecific interactions with other proteins significantly broaden resonances, suggesting that this may be an important factor for in-cell protein NMR visibility.[23] Indeed, these interactions slow protein diffusion by increasing protein molecular size and/or changing protein shape due to binding. In addition, line broadening can be caused by microsecond to millisecond 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 $^1H^N$ line widths in the cell lysate, which had approximately the same soluble protein components as the cell cytoplasm but was ~8−10 times more dilute. As a result, the apparent viscosity of the cell lysate, obtained by $^1H^N$ line width analysis, was much lower than in cells and only slightly higher than in water (Table S1 of the Supporting Information). On the basis of 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 (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 $^1H^N$ line widths 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 line widths on apparent molecular weight, which are linked to molecular tumbling and linearly proportional to molecular size and viscosity (Figure 4A, black). Transient interactions in the BSA-spiked lysates should significantly affect $^1H^N$ line widths, resulting in positive deviations from this linear dependence, i.e., when the higher bulk viscosity or/and molecular interactions increase the average $^1H^N$ line width more than predicted.

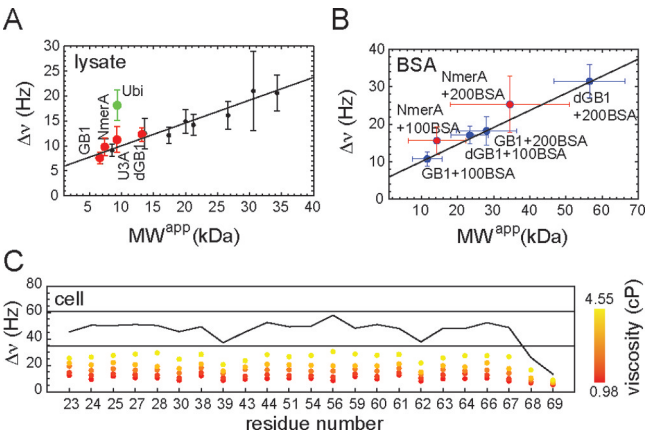


Figure 4. Effect of intracellular environments and molecular crowding on protein mobility. (A) Average $^1H^N$ line widths in lysate samples of GB1, NmerA, Ubi3A (U3A in the figure), dGB1 (red), and ubiquitin (green) and average $^1H^N$ line widths (black) obtained from glycerol titrations of GB1 and dGB1 lysates from data sets 2 and 3 (Table S1 of the Supporting Information) as a function of an apparent molecular weight; $MW^{app} = MW \times \eta/\eta_0$, 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 $^1H^N$ line widths 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; $MW^{app} = MW \times \eta/\eta_0$, where MW, η, and η_0 are the 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 panel A). (C) Histograms showing experimental $^1H^N$ line widths, $\Delta\nu$, as a function of NmerA residue number. Colors indicate viscosity in solution from lowest (red) to highest (yellow). Horizontal lines show the average $^1H^N$ line widths for in-cell GB1 and dGB1 samples. A black solid line shows theoretically predicted NmerA line widths expected for an intracellular environment 11 times as viscous as water.

Figure 4A shows the average $^1H^N$ line widths for GB1 and NmerA as a function of the apparent molecular weight and clearly demonstrates that the average NmerA and GB1 line width agrees very well with their predicted values. This result indicates the absence of significant broadening in the NmerA lysate spectrum compared with that of 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 $^1H^N$ line widths of GB1 indicated that the apparent viscosities of the 100 and 200 g/L BSA samples were ~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, ~1.4 and ~2.2 cP for 100 and 200 g/

dx.doi.org/10.1021/bi201287e | *Biochemistry* 2011, 50, 9225–9236

L BSA, respectively. Because the viscosity of our lysate samples was only slightly higher (0.98 cP) than that of water (0.92 cP) (see Table S1 and Figure S2 of the Supporting Information), we expected our lysate−BSA samples to have approximately 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 chymotrypsin inhibitor 2 (CI2).[23] In line with this previous result, the positive deviation of the apparent viscosity obtained from the $^1H^N$ line width 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 line width 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 difference in size between the protein observed and an interacting crowding agent that leads to changes in the protein's NMR line width (assuming fast exchange on the NMR time scale between the free and bound states). In the case of GB1−BSA interactions, the participation of even a very small fraction of GB1 (6.18 kDa) in the formation of a complex with BSA (69.3 kDa) should result in significant increases in GB1 line widths. For example, even if only 10% of the GB1 formed a transient complex with BSA in solution, the GB1 line widths 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 of whether weak, transient interactions between NmerA and BSA are more favored and, consequently, result in more significant line broadening arises.

To answer this question, we compared how GB1, dGB1, and NmerA line widths change in the presence of BSA. We suggested that the bulk viscosity, which is roughly equal to the apparent viscosity obtained from the GB1 $^1H^N$ line width analysis, depends only on the BSA concentration. Consequently, for each protein, we calculated an apparent molecular weight: $MW^{app} = MW \times \eta^{BSA}/\eta_0$, where MW is the protein molecular weight, η^{BSA} is the viscosity of the lysate solution in the presence of 100 or 200 g/L BSA, and η_0 is the viscosity of water. Figure 4B demonstrates that, like those of GB1, the average line widths of dGB1 increased as predicted from the bulk viscosity of BSA samples. The average line width for NmerA is slightly larger than that predicted from its molecular weight. However, this difference is smaller than experimental errors and contributes no more than 30% to NmerA line widths, 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 line widths differently compared to lysate or BSA solutions. To estimate how broad the in-cell NmerA spectrum should be on the basis of molecular crowding and viscosity considerations only, we used data set 2 (Table S1 of the Supporting Information) to obtain per residue dependencies of NmerA $^1H^N$ line widths for viscosity, which were used to calculate the expected line widths in cells (Figure 4C). To allow for changes in molecular tumbling and possible sample inhomogeneity, we used the apparent intracellular viscosity, obtained from the $^1H^N$ line width analysis of GB1, i.e., 11 times the viscosity of water (Figure 3). In contrast with experimental observations (Figure 1), such $^1H^N$ line width analysis predicted that in-cell NmerA should have a spectrum less broad and more intense than that of the in-cell dGB1 sample, and only slightly broader than that of the GB1 sample. This disagreement between experimental and predicted line widths 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 a significantly stronger influence on NmerA than GB1, while in the cell lysate contributions of these interactions are approximately 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 ranges of several hundred micromolar and a few millimolar, respectively, we estimate that NmerA interacts with cytoplasmic macromolecules with a K_d in the tens of millimolar range, which agrees with a 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 with alanine, viz., L8A, I44A, and V70A (in Figure S9 of the Supporting Information, 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 (Figure S6 of the Supporting Information). 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 ∼10-fold diluted relative to the in-cell sample, ubiquitin exhibited a reasonably good HSQC spectrum. However, $^1H^N$ peak line widths for the lysate sample were ∼2 times larger than those predicted from its molecular size (Figure 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 line width 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 are 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 (Figure 4A). The residues substituted with alanine in this mutant are located on a solvent-exposed hydrophobic surface of ubiquitin [the I44-containing surface (Figure S9 of the Supporting Information)] that is responsible for substrate recognition and binding.[45] Consequently, our observation that transient interactions become significantly weaker in Ubi3A compared to those in 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

(Figure S5B of the Supporting Information). 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 (Figure S6 of the Supporting Information), 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 millimolar 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) (Figure 3). This result predicts that the slow molecular tumbling for small proteins with a molecular mass of up to 13 kDa is not a limiting factor for observing them inside the cell. Indeed, 12.4 kDa dGB1, which tumbles ∼2 times more slowly than GB1, exhibits a high-quality in-cell NMR spectrum (Figure 1). By contrast, NMR signals for many other smaller proteins are too broad to be observed in our experience and that of others, indicating the presence of attractive interactions between a protein of interest and other intracellular macromolecules. Indeed, intermolecular interactions (including transient binding and/or 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 two to six GFP molecules that are covalently and linearly linked; net charge of GFP at neutral pH is −5.3) in the *E. coli* cytoplasm was 1 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 by the 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 the level of cyt *c* diffusion in lysate. Strong electrostatic interactions of the highly basic cyt *c* (with a net charge of +7) with the negatively charged *E. coli* cytosolic proteins result in formation of a complex with a molecular mass greater than 150 kDa, which could not be detected by NMR; however, charge-inverted mutants in cyt *c* or elevated salt concentrations led to disruption of cyt *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 cyt *c*, NmerA and ubiquitin have nearly no net charge (Table S2 and Figure S8 of the Supporting Information), 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 the "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 approximately 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 (Figure S8 of the Supporting Information). 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 millimolar, such that complexes are observed only under overexpression 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 ∼10 times compared to the concentrations in cells, and therefore, the weak interactions for ubiquitin are estimated to be ∼100 times stronger than those for NmerA (i.e., they have K_d of hundreds of micromolar). 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] (Figure S9 of the Supporting Information)], 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 a K_d of 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 20 different families.[45] The question of what mechanisms facilitate promiscuous hydrophobic interactions for ubiquitin arises. It has been shown that conformational dynamics on the hydrophobic patch on the I44-containing surface play an essential role in 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 a strong 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 anticorrelates 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, a loss of 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 values of hundreds of micromolar to tens of millimolar. 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 in ensuring that it forms 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 the overall charge, the distribution of hydrophobic residues, and the conformational flexibility.

■ ASSOCIATED CONTENT

Supporting Information

Details about in-cell NMR experiments and leakage controls, amino acid sequences of the fusion constructs, experimental data sets (Table S1), biophysical properties of the proteins used in this study (Table S2), determination of ^3J(HN−HA) constants (Figure S1), viscosity as a function of glycerol concentration (Figure S2), viscosity corrections for lysate samples (Figure S3), molecular tumbling and surface properties of dGB1 versus GB1 (Figure S4), protein leakage control for the GB1 in-cell spectrum (Figure S5), controls for the ubiquitin in-cell spectrum (Figure S6), lysate and in-cell NMR spectra (Figure S7), surface properties of GB1, NmerA, and ubiquitin (Figure S8), and surface changes for the Ubi3A mutant (Figure S9). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Telephone: (413) 545-6094. Fax: (413) 545-1289. E-mail: gierasch@biochem.umass.edu.

Author Contributions

Q.W. and A.Z. contributed equally to this work.

Funding

This work was supported by a National Institutes of Health Director's Pioneer Award (OD000945).

■ 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; NMR, nuclear magnetic resonance; NmerA, N-terminal domain of mercuric ion reductase (MerA); HSQC, heteronuclear single-quantum coherence; *S*, signal-to-noise ratio; TROSY, transverse relaxation optimized spectroscopy.

■ REFERENCES

(1) Verkman, A. S. (2002) Solute and macromolecule diffusion in cellular aqueous compartments. *Trends Biochem. Sci. 27*, 27−33.

(2) Mika, J. T., and Poolman, B. (2011) Macromolecule diffusion and confinement in prokaryotic cells. *Curr. Opin. Biotechnol. 22*, 117−126.

(3) Elcock, A. H. (2010) Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Curr. Opin. Struct. Biol. 20*, 196−206.

(4) Gershenson, A., and Gierasch, L. M. (2011) Protein folding in the cell: Challenges and progress. *Curr. Opin. Struct. Biol. 21*, 32−41.

(5) Zhou, H.-X., Rivas, G., and Minton, A. P. (2008) Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys. 37*, 375−397.

(6) Ellis, R. J. (2001) Macromolecular crowding: An important but neglected aspect of the intracellular environment. *Curr. Opin. Struct. Biol. 11*, 114−119.

(7) Heo, M. Y., Maslov, S., and Shakhnovich, E. (2011) Topology of protein interaction network shapes protein abundances and strengths of their functional and nonspecific interactions. *Proc. Natl. Acad. Sci. U.S.A. 108*, 4258−4263.

(8) Yadav, S., Shire, S. J., and Kalonia, D. S. (2010) Factors affecting the viscosity in high concentration solutions of different monoclonal antibodies. *J. Pharm. Sci. 99*, 4812−4829.

(9) Zorrilla, S., Hink, M. A., Visser, A. J., and Lillo, M. P. (2007) Translational and rotational motions of proteins in a protein crowded environment. *Biophys. Chem. 125*, 298−305.

(10) Elowitz, M. B., Surette, M. G., Wolf, P. E., Stock, J. B., and Leibler, S. (1999) Protein mobility in the cytoplasm of *Escherichia coli*. *J. Bacteriol. 181*, 197−203.
(11) Spitzer, J., and Poolman, B. (2009) The role of biomacromolecular crowding, ionic strength, and physicochemical gradients in the complexities of life's emergence. *Microbiol. Mol. Biol. Rev. 73*, 371−388.
(12) McConkey, E. H. (1982) Molecular evolution, intracellular organization, and the quinary structure of proteins. *Proc. Natl. Acad. Sci. U.S.A. 79*, 3236−3240.
(13) Rowe, A. J. (2011) Ultra-weak reversible protein-protein interactions. *Methods 54*, 157−166.
(14) Vaynberg, J., and Qin, J. (2006) Weak protein-protein interactions as probed by NMR spectroscopy. *Trends Biotechnol. 24*, 22−27.
(15) Deeds, E. J., Ashenberg, O., and Shakhnovich, E. I. (2006) A simple physical model for scaling in protein-protein interaction networks. *Proc. Natl. Acad. Sci. U.S.A. 103*, 311−316.
(16) McGuffee, S. R., and Elcock, A. H. (2010) Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput. Biol. 6*, e1000694.
(17) Miklos, A. C., Sarkar, M., Wang, Y. Q., and Pielak, G. J. (2011) Protein crowding tunes protein stability. *J. Am. Chem. Soc. 133*, 7116−7120.
(18) Burz, D. S., Dutta, K., Cowburn, D., and Shekhtman, A. (2006) Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). *Nat. Methods 3*, 91−93.
(19) Sakakibara, D., Sasaki, A., Ikeya, T., Hamatsu, J., Hanashima, T., Mishima, M., Yoshimasu, M., Hayashi, N., Mikawa, T., Walchli, M., Smith, B. O., Shirakawa, M., Guntert, P., and Ito, Y. (2009) Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature 458*, 102−105.
(20) Selenko, P., and Wagner, G. (2007) Looking into live cells with in-cell NMR spectroscopy. *J. Struct. Biol. 158*, 244−253.
(21) Serber, Z., Selenko, P., Hansel, R., Reckel, S., Lohr, F., Ferrell, J. E., Wagner, G., and Dotsch, V. (2006) Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. *Nat. Protoc. 1*, 2701−2709.
(22) Serber, Z., Ledwidge, R., Miller, S. M., and Dotsch, V. (2001) Evaluation of parameters critical to observing proteins inside living *Escherichia coli* by in-cell NMR spectroscopy. *J. Am. Chem. Soc. 123*, 8895−8901.
(23) Wang, Y., Li, C., and Pielak, G. J. (2010) Effects of proteins on protein diffusion. *J. Am. Chem. Soc. 132*, 9392−9397.
(24) Barnes, C. O., Monteith, W. B., and Pielak, G. J. (2011) Internal and global protein motion assessed with a fusion construct and in-cell NMR spectroscopy. *ChemBioChem 12*, 390−391.
(25) Huth, J. R., Bewley, C. A., Jackson, B. M., Hinnebusch, A. G., Clore, G. M., and Gronenborn, A. M. (1997) Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. *Protein Sci. 6*, 2359−2364.
(26) Zhang, N., Wang, Q., Ehlinger, A., Randles, L., Lary, J. W., Kang, Y., Haririnia, A., Storaska, A. J., Cole, J. L., Fushman, D., and Walters, K. J. (2009) Structure of the s5a:k48-linked diubiquitin complex and its interactions with rpn13. *Mol. Cell 35*, 280−290.
(27) Piotrowski, J., Beal, R., Hoffman, L., Wilkinson, K. D., Cohen, R. E., and Pickart, C. M. (1997) Inhibition of the 26 S proteasome by polyubiquitin chains synthesized to have defined lengths. *J. Biol. Chem. 272*, 23712−23721.
(28) Inomata, K., Ohno, A., Tochio, H., Isogai, S., Tenno, T., Nakase, I., Takeuchi, T., Futaki, S., Ito, Y., Hiroaki, H., and Shirakawa, M. (2009) High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature 458*, 106−109.
(29) Byeon, I.-J. L., Louis, J. M., and Gronenborn, A. M. (2004) A captured folding intermediate involved in dimerization and domain-swapping of GB1. *J. Mol. Biol. 340*, 615−625.
(30) Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997) 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. 94*, 12366−12371.
(31) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277−293.
(32) Keller, R. L. J. (2004) *The computer aided resonance assignment tutorial*, 1st ed., Cantina-Verlag, Goldau, Switzerland.
(33) Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III, Skelton, N. J.,Rance,M. (2007) *Protein NMR spectroscopy: Principles and practice*, 2nd ed., Elsevier Academic Press, Amsterdam.
(34) Markus, M. A., Dayie, K. T., Matsudaira, P., and Wagner, G. (1994) Effect of deuteration on the amide proton relaxation rates in proteins: Heteronuclear NMR experiments on villin 14T. *J. Magn. Reson., Ser. B 105*, 192−195.
(35) Walsh, J. D., Meier, K., Ishima, R., and Gronenborn, A. M. (2010) NMR studies on domain diffusion and alignment in modular GB1 repeats. *Biophys. J. 99*, 2636−2646.
(36) Barnes, C. O., and Pielak, G. J. (2011) In-cell protein NMR and protein leakage. *Proteins 79*, 347−351.
(37) Miklos, A. C., Li, C. G., Sharaf, N. G., and Pielak, G. J. (2010) Volume exclusion and soft interaction effects on protein stability under crowded conditions. *Biochemistry 49*, 6984−6991.
(38) Schlesinger, A. P., Wang, Y., Tadeo, X., Millet, O., and Pielak, G. J. (2011) Macromolecular crowding fails to fold a globular protein in cells. *J. Am. Chem. Soc. 133*, 8082−8085.
(39) Li, C., Wang, G. F., Wang, Y., Creager-Allen, R., Lutz, E. A., Scronce, H., Slade, K. M., Ruf, R. A., Mehl, R. A., and Pielak, G. J. (2010) Protein ^{19}F NMR in *Escherichia coli*. *J. Am. Chem. Soc. 132*, 321−327.
(40) Crowley, P. B., Chow, E., and Papkovskaia, T. (2011) Protein interactions in the *Escherichia coli* cytosol: An impediment to in-cell NMR spectroscopy. *ChemBioChem 12*, 1043−1048.
(41) Palmer, A. G., Kroenke, C. D., and Loria, J. P. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol. 339*, 204−238.
(42) Crowley, P. B., Brett, K., and Muldoon, J. (2008) NMR spectroscopy reveals cytochrome c-poly(ethylene glycol) interactions. *ChemBioChem 9*, 685−688.
(43) Li, C., and Pielak, G. J. (2009) Using NMR to distinguish viscosity effects from nonspecific protein binding under crowded conditions. *J. Am. Chem. Soc. 131*, 1368−1369.
(44) Li, C., Charlton, L. M., Lakkavaram, A., Seagle, C., Wang, G., Young, G. B., Macdonald, J. M., and Pielak, G. J. (2008) Differential dynamical effects of macromolecular crowding on an intrinsically disordered protein and a globular protein: Implications for in-cell NMR spectroscopy. *J. Am. Chem. Soc. 130*, 6310−6311.
(45) Dikic, I., Wakatsuki, S., and Walters, K. J. (2009) Ubiquitin-binding domains: From structures to functions. *Nat. Rev. Mol. Cell Biol. 10*, 659−671.
(46) Link, A. J., Robison, K., and Church, G. M. (1997) Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. *Electrophoresis 18*, 1259−1313.
(47) Nenninger, A., Mastroianni, G., and Mullineaux, C. W. (2010) Size Dependence of Protein Diffusion in the Cytoplasm of *Escherichia coli*. *J. Bacteriol. 192*, 4535−4540.
(48) Kumar, M., Mommer, M. S., and Sourjik, V. (2010) Mobility of Cytoplasmic, Membrane, and DNA-Binding Proteins in *Escherichia coli*. *Biophys. J. 98*, 552−559.
(49) Hoskins, J., Lovell, S., and Blundell, T. L. (2006) An algorithm for predicting protein-protein interaction sites: Abnormally exposed amino acid residues and secondary structure elements. *Protein Sci. 15*, 1017−1029.
(50) Fisher, R. D., Wang, B., Alam, S. L., Higginson, D. S., Robinson, H., Sundquist, W. I., and Hill, C. P. (2003) Structure and ubiquitin binding of the ubiquitin-interacting motif. *J. Biol. Chem. 278*, 28976−28984.

(51) Van Horn, W. D., Ogilvie, M. E., and Flynn, P. F. (2009) Reverse micelle encapsulation as a model for intracellular crowding. *J. Am. Chem. Soc. 131*, 8030−8039.
(52) Boehr, D. D., Nussinov, R., and Wright, P. E. (2009) The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol. 5*, 789−796.
(53) Wlodarski, T., and Zagrovic, B. (2009) Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proc. Natl. Acad. Sci. U.S.A. 106*, 19346−19351.
(54) Lindman, S., Xue, W. F., Szczepankiewicz, O., Bauer, M. C., Nilsson, H., and Linse, S. (2006) Salting the charged surface: pH and salt dependence of protein G B1 stability. *Biophys. J. 90*, 2911−2921.