

# Connecting Longitudinal and Transverse Relaxation Rates in Live-Cell NMR

Sarah Leeb, Fan Yang, Mikael Oliveberg,\* and Jens Danielsson\*

Cite This: *J. Phys. Chem. B* 2020, 124, 10698–10707

Read Online

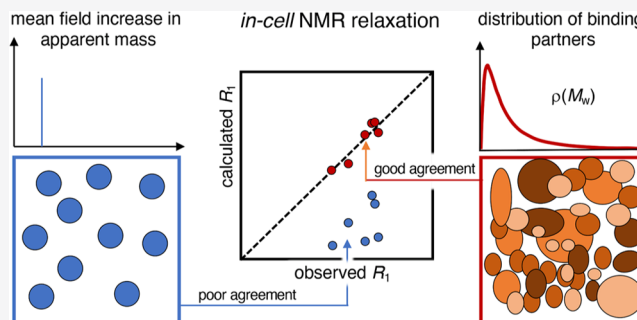
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** In the cytosolic environment, protein crowding and Brownian motions result in numerous transient encounters. Each such encounter event increases the apparent size of the interacting molecules, leading to slower rotational tumbling. The extent of transient protein complexes formed in live cells can conveniently be quantified by an apparent viscosity, based on NMR-detected spin-relaxation measurements, that is, the longitudinal ( $T_1$ ) and transverse ( $T_2$ ) relaxation. From combined analysis of three different proteins and surface mutations thereof, we find that  $T_2$  implies significantly higher apparent viscosity than  $T_1$ . At first sight, the effect on  $T_1$  and  $T_2$  seems thus nonunifiable, consistent with previous reports on other proteins. We show here that the  $T_1$  and  $T_2$  deviation is actually not a inconsistency but an expected feature of a system with fast exchange between free monomers and transient complexes. In this case, the deviation is basically reconciled by a model with fast exchange between the free-tumbling reporter protein and a transient complex with a uniform 143 kDa partner. The analysis is then taken one step further by accounting for the fact that the cytosolic content is by no means uniform but comprises a wide range of molecular sizes. Integrating over the complete size distribution of the cytosolic interaction ensemble enables us to predict both  $T_1$  and  $T_2$  from a single binding model. The result yields a bound population for each protein variant and provides a quantification of the transient interactions. We finally extend the approach to obtain a correction term for the shape of a database-derived mass distribution of the interactome in the mammalian cytosol, in good accord with the existing data of the cellular composition.



## INTRODUCTION

The cell interior is an immensely complex and crowded environment, both in prokaryotic and eukaryotic cells. Hence, a freely diffusing protein is bound to undergo countless collisions and transient interactions with the surrounding macromolecules.<sup>1–3</sup> Most frequently, the collisions are non-elastic, that is, a transient, so-called, “encounter complex” is formed, allowing for Brownian-surface diffusion and a close-range search for putative specific recognition and binding. Preferably, such weak interactions may result in colocalization of functionally related proteins, often called the quinary structure,<sup>4–6</sup> which can be described as a functional subset of all transient interactions. The duration of this surface search has been postulated to be under evolutionary optimization,<sup>1,7,8</sup> based on the argument that if the “hand shake” is too brief, the proteins may fail to recognize their partner and, conversely, if it is too long, they will simply waste their precious time. Consistent with this idea, the repulsive net-charge between cytosolic proteins across evolutionarily divergent organisms seems optimized for marginal colloidal stability.<sup>9</sup> For a given protein, each encounter-complex formation results in a transient reduction in rotational and translational diffusion.<sup>10–13</sup> The rotational component ( $D_{rot}$ ) of this retardation is well-suited to be quantified by in-cell nuclear magnetic

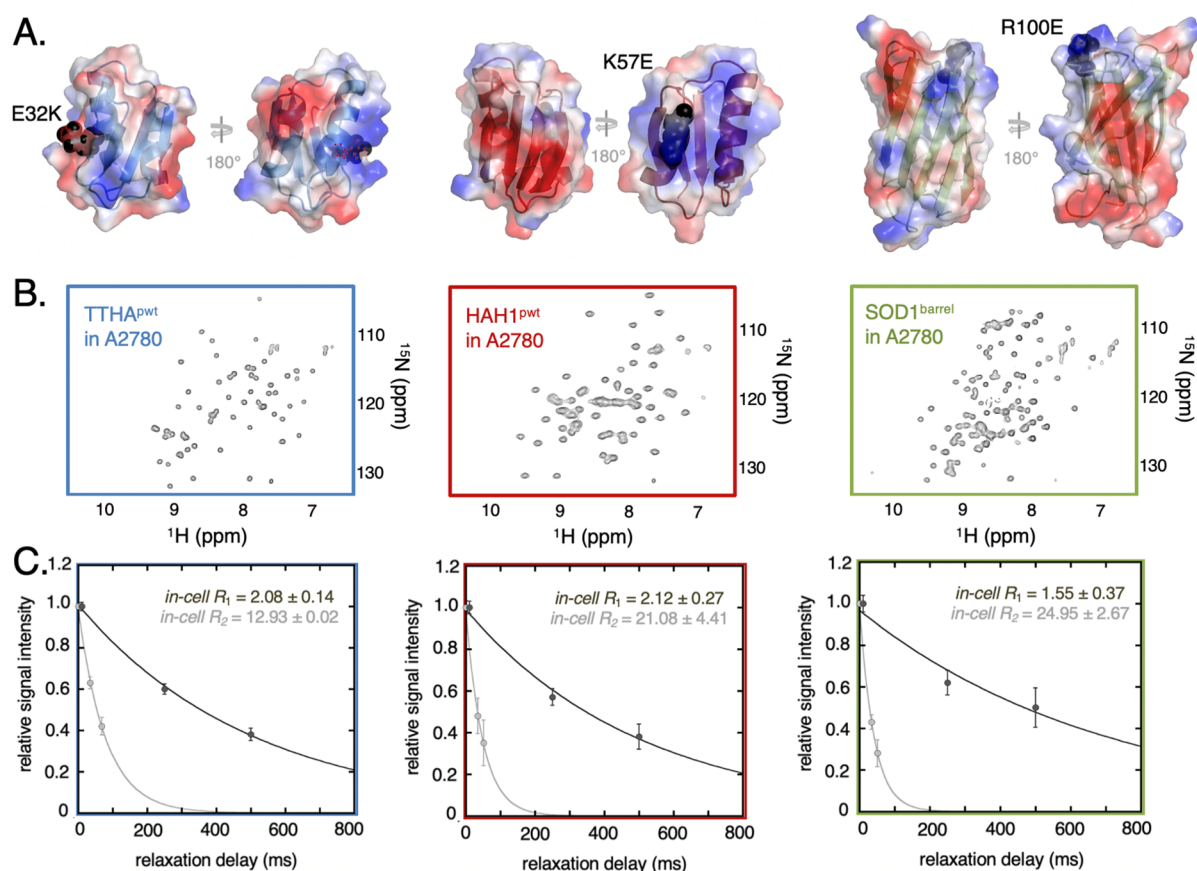
resonance (in-cell NMR). In essence, the rotational motions are here coupled to the longitudinal ( $T_1$ ) and transverse ( $T_2$ ) relaxation times<sup>14</sup> that, in turn, provide a measure of a particular protein’s propensity to interact with the neighboring intracellular components.<sup>6,10,11,15</sup> The quantitative link between the molecular motions and the relaxation time is described by the overall correlation time ( $\tau_c$ ), which, in the general case, is the harmonic mean of global rotation ( $\tau_r$ ), local motions ( $\tau_{loc}$ ), and additional components such as chemical exchange ( $\tau_{ex}$ ):  $\tau_c^{-1} = \tau_r^{-1} + \tau_{loc}^{-1} + \tau_{ex}^{-1}$ . The effects of the in-cell encounters on the molecular motions are expected to span several timescales and can be both global and local, but for folded proteins, the local motions are of less importance, that is,  $T_1$  and  $T_2$  mainly report on the rotational correlation time,  $\tau_r$ .<sup>16</sup>

Received: September 11, 2020

Revised: October 22, 2020

Published: November 12, 2020





**Figure 1.** Probe proteins in-cell NMR properties and relaxation. (A) shows the proteins TTHA<sup>PWT</sup> (blue), HAH1<sup>PWT</sup> (red), and SOD1<sup>BARREL</sup> (green) including secondary structure elements. The electrostatic surface of each protein is displayed, with blue-colored patches belonging to the basic, positively charged residues arginine or lysine and the red colored patches belonging to the acidic, negatively charged residues glutamate or aspartate. The surface charge mutations are highlighted as black spheres. (B) depicts HMQC spectra of the reporter protein electroporated into live A2780 cells. (C) In-cell NMR relaxation data of the three basis reporter proteins are shown. Signal intensity attenuation obtained from the  $R_1$  (dark gray) and  $R_2$  (light gray) experiments are shown as filled circles and the corresponding fitted single exponential fits are shown as solid lines. The error bars in the relaxation rates are estimated from the signal-to-noise-ratio in each experiment (Supporting Information Methods).

Previous studies have shown that the extent of  $D_{\text{rot}}$  retardation upon cell internalization depends not only on the protein's physicochemical surface properties<sup>6,10,11,17</sup> but also on the type of host cell.<sup>10,11,13</sup> This in-cell retardation is manifested as increased  $T_1$  and decreased  $T_2$  relaxation times<sup>11,14,18,19</sup> (Figure 1, Supporting Information 1), where most commonly  $T_2$  has been exploited to quantify the retardation effect, both directly by determination of the relaxation time<sup>10,11,19</sup> and indirectly through line-broadening analysis.<sup>6,20</sup>

To enable comparison of proteins with different mass, we translate here the observed relaxation times into “apparent viscosities” ( $\eta^{\text{app}}$ ),<sup>11,13,20,21</sup> that is, the microscopic viscosities that yield the same relaxation parameters as observed in cells.<sup>10</sup> The  $\eta^{\text{app}}$  values are derived from the reference curves, in which  $T_1$  and  $T_2$  are determined for each protein in increasing amounts of glycerol with well-defined viscosities. Even so, this mean-field approach results in an apparent disagreement between  $T_1$  and  $T_2$  in-cell relaxation times, where transverse relaxation reports on much higher  $\eta^{\text{app}}$  than longitudinal relaxation.

To pinpoint this  $T_1$  and  $T_2$  inconsistency, we expand here the mean-field retardation approximation to a binding model, where the reporter protein is set to be in fast exchange between a free monomeric state and a bound state with a cellular

partner. This partner is, in the simplest case, a protein of uniform size or a distribution of proteins with different sizes. As a result, the observed  $T_1$  and  $T_2$  values for our three different proteins and their mutants can be accounted for by a single model. Our basic observation is that the relaxation data features can be accurately described by the minimal assumption of a single partner of uniform size. This accuracy is retained upon accounting for more realistic partner-mass distributions. The best partner-mass distribution is that obtained from the naturally occurring cytosolic proteins but with an increased tail of higher mass species. Consistent with the cellular composition, this tail is suggested to report on transient interactions with higher order complexes and larger biomolecular structures, such as membranes and ribosomes. Thus, by accounting for the full distribution of putative interaction partners, we are allowed to connect the rotational retardation to the fraction of transiently bound reporter proteins,  $p_B$ . This bound population constitutes also a quantitative link between the reporter protein properties and their interaction pattern in cells. A final implication of this result is that information about the particular partner-size distributions under a given set of conditions can be obtained directly from  $T_1$  and  $T_2$ , providing a new handle for exploring the macromolecular machinery at work in live cells.

## ■ EXPERIMENTAL PROCEDURES

**Protein Mutagenesis, Expression, and Purification.** Plasmids were transformed into *Escherichia coli* BL21(DE3) expression strains, and point mutations were introduced through site-directed mutagenesis. For  $^{15}\text{N}$ -isotope-enriched protein production, minimal medium [0.02 M  $\text{KH}_2\text{PO}_4$ , 0.04 M  $\text{Na}_2\text{HPO}_4$ , 0.1 M NaCl, 2 mM  $\text{MgSO}_4$ , 0.4% (w/v) glucose, and trace metals (1 mM  $\text{MgCl}_2$ , 15  $\mu\text{M}$   $\text{CaCl}_2$ , 1  $\mu\text{M}$   $\text{FeSO}_4$ , 32 nM  $\text{AlCl}_3$ , 33 nM  $\text{CoCl}_2$ , 12 nM  $\text{CuSO}_4$ , 120 nM KI, 100 nM  $\text{MnSO}_4$ , 4 nM  $\text{NiSO}_4$ , 16 nM  $\text{Na}_2\text{MoO}_4$ , 15 nM  $\text{ZnSO}_4$ , 5 nM  $\text{KCr}(\text{SO}_4)_2$ , 15 nM  $\text{H}_3\text{BO}_3$ , and 100  $\mu\text{M}$  citric acid), pH 7.0] supplemented with carbenicillin and 0.1% (w/v)  $^{15}\text{NH}_4\text{Cl}$  was inoculated and grown at 37 °C until  $\text{OD}_{600} = 0.6\text{--}0.8$ . Isopropyl  $\beta$ -*D*-1-thiogalactopyranoside (IPTG) (0.5 mM) was added for 4 h overexpression. Cells were harvested at 5000g for 10 min at 4 °C (Supporting Information Methods). Protein purification of  $\text{SOD1}^{\text{barrel}}$  is described in detail by Danielsson *et al.*,<sup>22</sup> while detailed protocols for  $\text{TTHA}^{\text{PWT}}$  and  $\text{HAH1}^{\text{PWT}}$  are described by Mu *et al.*<sup>6</sup> and Leeb *et al.*<sup>10</sup>

**Protein Transfer into Mammalian Cells for In-Cell NMR.** Cell growth and protein transfer by electroporation were as described in the literature.<sup>10</sup> In short, human ovary adenocarcinoma A2780 cells were grown to 70–90% confluence. Approximately,  $60 \times 10^6$  cells were suspended and supplemented with the respective reporter proteins to 1.5 mM final protein concentration. Electroporation was conducted by 115 V, 14–16 ms poring pulses followed by  $5 \times 50$  ms transfer pulses at 20 V. After electroporation, the cells were washed, plated, and left for 5 h recovery. Then, the live cells were transferred to a 4 mm flat-bottomed NMR tube (BMS-004B, Shigemi Inc., Tokyo, Japan) (Supporting Information Methods).

**In-Cell Relaxation Measurements.** All NMR data were acquired on a Bruker AVANCE III 700 MHz spectrometer with a cryogenically cooled triple-resonance probe. All experiments were performed at 37 °C using an “interleaved” acquisition method.<sup>10</sup> Both  $R_2$  and  $R_1$  measurements were carried out using one-dimensional  $^{15}\text{N}$ -filtered heteronuclear single quantum coherence (HSQC)-based pulse sequences with three relaxation delays each, ranging between 0 and 68 ms in the case of  $R_2$  and 10 and 500 ms in the case of  $R_1$ .

**Quantification of Protein Leakage.** Leakage of the reporter protein into the interstitial fluid surrounding the cells was quantified by carefully removing the cell slurry from the NMR tube. After spinning 5 min at 200 g, the supernatant was transferred to a fresh NMR tube and 1D  $^1\text{H}$ -SOFAS–HMQC was recorded (Figure S2). Integrating over the same spectral regions in both samples and then calculating the ratio after correcting for dilution in the supernatant samples showed that protein leakage typically is less than 10% (Figure S2). In addition, there is evidence that most of the leakage is introduced during sample preparation of the supernatant.<sup>6</sup>

**Relaxation Measurements of In Vitro Glycerol Series.** 200  $\mu\text{M}$  protein, 10 mM MES pH 6.5, 10% (v/v)  $\text{D}_2\text{O}$ , and increasing amounts of deuterated glycerol- $d_8$  (98% D) from 0 to 50% (v/v) were used. Both  $R_1$  and  $R_2$  were determined with 6–10 relaxation delays (Supporting Information Methods). Data was finally analyzed using MATLAB (MathWorks, MA, USA) scripts as described in the literature.<sup>10</sup>

**Lysozyme Crowding.** 100  $\mu\text{M}$   $\text{TTHA}^{\text{PWT}}$ , 20 mM MES pH 6.5, and 10% (v/v)  $\text{D}_2\text{O}$  together with 50 or 150 mg/mL human lysozyme (Sigma-Aldrich) were used. Both  $R_1$  and  $R_2$

were determined using 5–6 relaxation delay times. Data was analyzed with in-house MATLAB (MathWorks, MA, USA) scripts. Since lysozyme is pH-active, the sample pH had decreased to 6.0 and 5.6 for 50 and 150 mg/mL, respectively. Fluorescence spectroscopy was conducted under similar conditions, where 0–20 mM  $\text{TTHA}^{\text{PWT}}$  was titrated onto 100 mM human lysosome, upon which the induced change in intrinsic tryptophan fluorescence from lysozyme was detected (Figure S5).

**Curating and Analyzing the Cytosolic Proteome Database.** The proteomic composition of the mammalian cytosol was estimated from the database by Geiger *et al.*,<sup>23</sup> where lysate proteins of eleven human cancer cell lines were quantified by mass spectrometry. Of the 11,731 proteins in the Geiger list, 4.2% were given ambiguous UniProt-IDs, and these sequences were omitted. The calculations of charge density of the proteins are described in the Supporting Information Methods.

## ■ RESULTS AND DISCUSSION

**Model System.** Quantification of NMR relaxation in live mammalian cells requires good signal to noise data. This means that the proteins used must exhibit only weak interactions with the surroundings as extensive interactions lead to severe line-broadening.<sup>5</sup> We use here three well-characterized model proteins: (i) the nonmetal binding variant of the putative heavy-metal binding protein  $\text{TTHA1718}$  from *Thermus thermophilus* ( $\text{TTHA}^{\text{PWT}}$ ),<sup>6</sup> (ii) the corresponding demetalated variant of the copper chaperone  $\text{HAH1}$  ( $\text{HAH1}^{\text{PWT}}$ ),<sup>6</sup> and (iii) the loop-truncated variant of the human superoxide dismutase  $\text{SOD1}$  ( $\text{SOD1}^{\text{barrel}}$ )<sup>22</sup> (Figure 1).  $\text{TTHA}^{\text{PWT}}$  and  $\text{HAH1}^{\text{PWT}}$  are structural homologues, but with distinct surface properties, making them well suited as reporter proteins for surface-mediated interactions in cells. All three proteins exhibit good NMR-relaxation properties in live mammalian cells,<sup>10</sup> where well-resolved HMQC spectra (Figure 1) indicate that they only diffusely interact with the cell interior.<sup>10,24</sup> That is, they mainly probe nonspecific transient encounters. Nonetheless, in previous studies, we have shown that these three proteins exhibit different degrees of transient interactions related to their surface properties. Of particular importance here is the surface net-charge density and surface hydrophobicity.<sup>6,10</sup> To expand the surface-property space, we included further three surface variants of the model proteins above, all of which alter the surface net charge by 2 units: a Glu to Lys substitution on the  $\text{TTHA}^{\text{PWT}}$  background ( $\text{TTHA}^{\text{E32K}}$ ), a Lys to Glu substitution on the  $\text{HAH1}^{\text{PWT}}$  background ( $\text{HAH1}^{\text{K57E}}$ ), and an Arg to Glu substitution on  $\text{SOD1}^{\text{barrel}}$  ( $\text{SOD1}^{\text{R100E}}$ ) (Figure 1). This set of six proteins then serves as reporters on the transient interactions between the protein surfaces and the cellular surroundings. For the in-cell NMR measurement, the proteins were electroporated into A2780 cells, obtaining approximately 20  $\mu\text{M}$  intracellular concentration,<sup>10</sup> corresponding to approximately 10  $\mu\text{M}$  total sample concentration.

**Longitudinal and Transverse Relaxation Report Different Apparent Viscosity in Human Cells.** To quantify the mean-field parameter,  $\eta^{\text{app}}$ , we determined the global  $^{15}\text{N}$   $R_1 = T_1^{-1}$  and  $^{15}\text{N}$   $R_2 = T_2^{-1}$  of all six proteins in live A2780 cells at 37 °C<sup>10</sup> (Figure 1, Table 1, Figures S1 and S2). Hereinafter, we will use the relaxation rate ( $R_i$ ) in parallel to the relaxation time ( $T_i$ ), for clarity in the analysis. Due to the relatively low total concentration of isotope-labeled reporter

**Table 1.** Collected Physicochemical Properties and In-Cell Relaxation Rates of the Reporter Proteins in Human A2780 Cells

Protein	$M_w$ (Da)	net charge <sup>a</sup>	$R_1$	$R_2$ <sup>b</sup>
TTHA <sup>PWT</sup>	7009	-1.47	$2.08 \pm 0.14$	$12.93 \pm 0.02$
TTHA <sup>E32K</sup>	7008	0.50	$1.96 \pm 0.50$	$22.62 \pm 2.26$
HAH1 <sup>PWT</sup>	7353	0.94	$2.12 \pm 0.27$	$21.08 \pm 4.41$
HAH1 <sup>K57E</sup>	7354	-1.08	$2.04 \pm 0.29$	$11.76 \pm 1.56$
SOD1 <sup>barrel</sup>	10962	-0.70	$1.55 \pm 0.37$	$24.95 \pm 2.67$
SOD1 <sup>R100E</sup>	10949	-2.50	$1.74 \pm 0.31$	$16.63 \pm 2.26$

<sup>a</sup>Net charge calculated using propKa 3.0.<sup>50</sup> <sup>b</sup>Data from Leeb *et al.*<sup>10</sup>

protein in the cell sample and the limited life time of the cells in the NMR tube,<sup>25</sup> we were only able to record three relaxation delay times (Figure 1), where we integrate over a set of amide signals (Supporting Information Methods). The interscan delay was set to 1 s, which is bordering on being too short for the  $R_1$  relaxation rates of  $1.5\text{--}2\text{ s}^{-1}$ . As a possible consequence of this trade-off between interscan relaxation delay and the total experimental time, the  $R_1$  data shows some deviation from ideal mono-exponential behavior, most evidently for the shortest delay (Figure 1). In addition, a contributing factor to the deviation from single exponentiality could stem from the presence of two populations of the reporter protein—cytosolic and leaked protein. However, the leakage is in most cases negligible, no systematic deviations are found in the  $R_2$  attenuation curves (Figure 1), and no correlation between the magnitude of deviation and the amount of leakage can be observed. Furthermore, fitting the data to a biexponential yielded nonphysiological relaxation values (Figure S3). Taken together, we conclude that leakage is not a major contributing factor to the deviation from monoexponentiality.

To test the reproducibility, we determined  $R_1$  of HAH1<sup>PWT</sup> twice, with  $R_1 = 2.12 \pm 0.27$  and  $2.09 \pm 0.18\text{ s}^{-1}$ , indicating the precision of the relaxation measurements.  $R_2$  precision has previously been shown to be similar.<sup>10</sup> We compared the  $R_1$

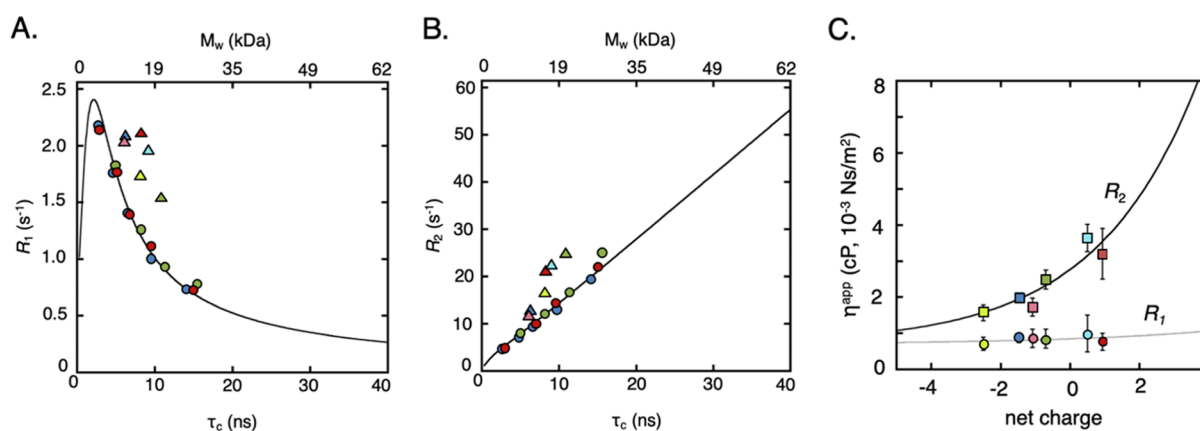
and  $R_2$  values to reference values obtained in water–glycerol mixtures to get  $\eta^{\text{app}}$  (Supporting Information 2, Figure S4, Table S1). As expected from previous work,<sup>6,10,11</sup> the results show that the  $R_2$ -derived  $\eta^{\text{app}}$  values increase with increased protein net-charge (Figure 2, Table 2, Figure S4). This

**Table 2.** Apparent Viscosities  $\eta^{\text{app}}$  Derived from  $F_{R_1}$  and  $F_{R_2}$ 

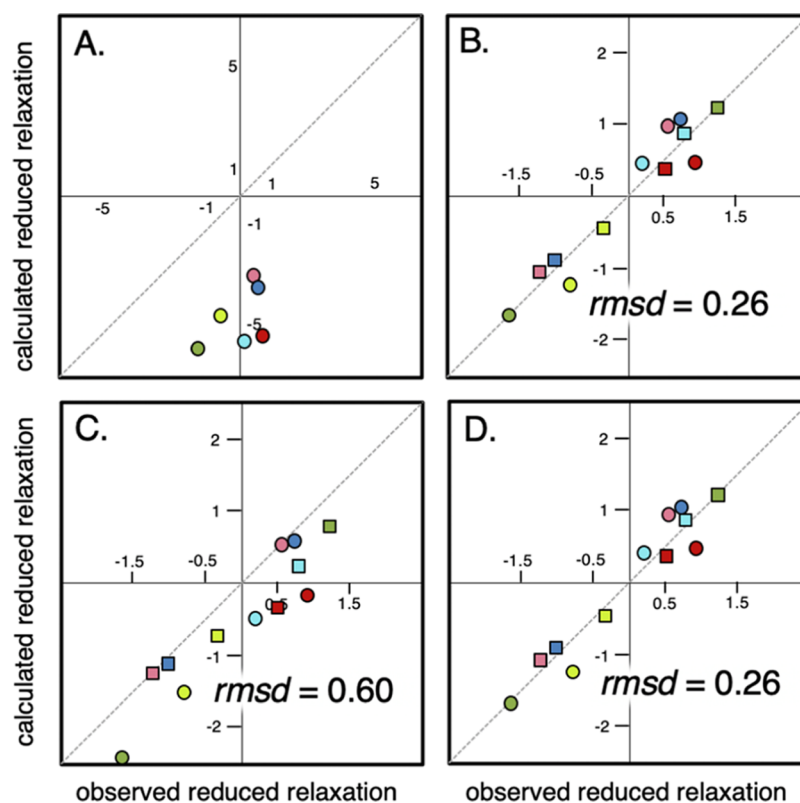
protein	$\eta^{\text{app},R_1}$ (cP) <sup>a</sup>	$\eta^{\text{app},R_2}$ (cP)
TTHA <sup>PWT</sup>	0.86 (−0.12, +0.10)	$2.02 \pm 0.00$
TTHA <sup>E32K</sup>	0.96 (−0.48, +0.54)	$3.64 \pm 0.38$
HAH1 <sup>PWT</sup>	0.78 (−0.26, +0.21)	$3.20 \pm 0.70$
HAH1 <sup>K57E</sup>	0.84 (−0.24, +0.26)	$1.72 \pm 0.25$
SOD1 <sup>barrel</sup>	0.81 (−0.23, +0.29)	$2.48 \pm 0.26$
SOD1 <sup>R100E</sup>	0.68 (−0.16, +0.21)	$1.56 \pm 0.22$

<sup>a</sup>In the case of  $R_1$ -derived  $\eta^{\text{app}}$ , the error is asymmetric and are shown in parentheses.

complies with the notion that the more positively charged the proteins are, the more strongly they interact with the intracellular environment.<sup>6,10,11,17,26–29</sup> It is further apparent that the  $R_1$ -derived  $\eta^{\text{app}}$  values are much less affected (Figure 2) by showing lower apparent viscosity as well as nearly no charge dependence. Using the  $R_2$ -derived  $\eta^{\text{app}}$  to predict the corresponding  $R_1$  values results thus in poor agreement with the observed values (Figure 3), with high rmsd = 4.86 and low Pearson correlation coefficient  $r^2 = 0.23$  between the measured and calculated relaxation rates. Nonetheless, the observation that  $R_1$  and  $R_2$  yield different  $\eta^{\text{app}}$  is in good agreement with previous findings.<sup>11,19</sup> From <sup>19</sup>F NMR relaxation analysis, this discrepancy has been suggested to stem from <sup>19</sup>F  $R_2$  being more sensitive to transient interactions, while <sup>19</sup>F  $R_1$  reports mainly on local motions.<sup>19</sup> Basically, this shows that this mean-field approach at some level fails to fully describe the in-cell effect. For comparison, <sup>15</sup>N  $R_1$  and <sup>15</sup>N  $R_2$  of the same proteins in glycerol–water mixtures fall well within the theoretical predictions (Figure 2, Supporting Information 1, Table S1).



**Figure 2.** NMR relaxation data and apparent viscosity derived therefrom. (A,B) NMR relaxation rates as functions of rotational correlation time,  $\tau_r$ , and molecular weight  $M_w$ , at 700 MHz (16.5 T) field strength. The colored circles are the measured relaxation rates [(A)  $R_1$ , (B)  $R_2$ ] for the three proteins (HAH1<sup>PWT</sup>: red, SOD1<sup>barrel</sup>: green and TTHA<sup>PWT</sup>: blue) in increasingly viscous glycerol solutions. While  $\tau_r$  values calculated from the two relaxation rates (Supporting Information 1) fit the predicted theoretical values (black line) in the case of the *in vitro* glycerol data, clear deviations from the theory are found for the in-cell relaxation data (triangles), where the brighter colors correspond to the respective surface mutation. (C) Apparent viscosities  $\eta^{\text{app}}$  derived from transverse,  $R_2$  (squares), and from longitudinal in-cell relaxation,  $R_1$  (circles), plotted against protein net charge, where the colors are as in (A,B). The lines are empirically fitted exponential curves, with an offset corresponding to the intrinsic viscosity of water.<sup>10</sup> The marked discrepancy between the obtained  $\eta^{\text{app}}$  values suggests that the mean-field  $\eta^{\text{app}}$  model is insufficient for explaining changes in NMR relaxation due to intracellular encounters.



**Figure 3.** Agreement between observed and calculated reduced relaxation rates for the different models. The calculated  $R_1 = T_1^{-1}$  values in (A) are from the mean field  $\eta^{\text{pp}}$  approach, note the different axis scale in this figure. (B) Corresponds to fast exchange to a binding partner with an optimized mass of 143 kDa. In (C, D), lognormal mass distributions are used, where the shape factors are optimized in (D). The dashed line corresponds to a 1:1 correlation. Reduced  $R_1$  values are shown as circles, while reduced  $R_2$  values are depicted as squares. The color coding is the same as in Figures 1 and 2.

One explanation for the observed discrepancy is that the intracellular environment, in contrast to glycerol, adds chemical-exchange contributions from local surface interactions.  $R_2$  is here known to be most affected through exchange broadening ( $R_{\text{ex}}$ ),<sup>30,31</sup> while  $R_1$  is left relatively unaffected. Challenging this idea, we have previously found that with our reporter proteins, the cytosolic enhancement of  $R_2$  is largely due to changes in global rotation,  $\tau_r$ .<sup>10</sup> In essence, our conclusion is based on the finding that the line-broadening effect in the in-cell NMR spectra is uniform over all residues,<sup>10</sup> suggesting that the effect is global retardation rather than localized exchange effects. The effect on  $R_2$  is, moreover, independent of the nuclei type, which would not be the case if  $R_{\text{ex}}$  would be the dominating factor. Finally, removal of the refocusing CPMG train in the pulse sequence of the  $R_2$  experiment<sup>10</sup> renders very similar relaxation rates, which suggests relatively small  $R_{\text{ex}}$  contributions in the ms regime. Taken together, these observations indicate that exchange processes are not the primary cause for discrepancy in the observed in-cell relaxation effects and that an alternative explanation is to be found.

**Transient Binding to a Single Large Partner Reconciles  $R_1$  and  $R_2$  Data.** To seek a model that predicts both  $R_1$  and  $R_2$ , we first extended the mean-field assumption where the tumbling of the entire protein ensemble is affected homogeneously to a model where free monomers, with mass  $M_j$ , are in rapid exchange with transient clusters of an average molecular weight  $M^{\text{av}}$  (Supporting Information 3). This

extension yields a population-weighted average of the relaxation rates of the bound and free state according to

$$R_i^{\text{calc}} = p_B F_{R_i}(M_j + M^{\text{av}}) + (1 - p_B) F_{R_i}(M_j) \quad (1)$$

where  $p_B$  is the fraction of reporter protein bound to the cluster at any given time and  $F_{R_i}$  is the closed-form expression of the relaxation rate as a function of mass (Supporting Information 1). The good in-cell NMR properties of the proteins<sup>10</sup> indicate that the fast-exchange criterion holds since long-lived binding would yield line-broadening beyond the detection limit.<sup>5</sup> This model constitutes a simplified description of the transiently bound state: the encounter complex is pictured as a rigid body with mass  $M_j + M^{\text{av}}$ . However, a formed encounter complex involves Brownian motion along the complex surface, which also affects relaxation. The assumption of a rigid complex may thus result in a slight underestimation of  $p_B$ . In this analysis, we assume that the correlation time is independent of the direction and that the diffusion tensor is symmetric. Although the former simplification breaks down in highly concentrated heterogeneous environments, the relatively small effects on the relaxation properties of the reporter proteins in A2780<sup>10</sup> suggest a comparably diluted environment. Use of a set of relaxation rates of individual spins located over the structure would allow the determination of the diffusion tensor.<sup>32</sup> However, here, we determine an average relaxation rate from many amide spins, spread out over the structure, justifying the use of an isotropic diffusion tensor. The near-linear dependence between  $R_2$  and molecular mass (Figure 2) means that any encounter complex mass can be accounted for by adjusting

$p_B$ , while, in contrast, the nonlinear relationship between  $R_1$  and mass (Figure 2) strongly confines the possible masses of the complex, fixating  $p_B$ . To benchmark this approach for quantifying weak transient interactions, we determined the effect on the relaxation rates of TTHA<sup>pwt</sup>, in presence of 50 and 150 mg/mL human lysozyme-concentrations comparable with the total protein levels in human cells.<sup>10,33</sup> Here, we find that  $R_1$  and  $R_2$  agrees well with a 1:1 weak transient complex for  $p_B = 0.25$  and  $0.46$  respectively, in good accordance with fluorescence-detected binding affinity (Supporting Information 4, Figure S5).

As shown in Figure 3, the introduction of a bound species with a common interaction partner of  $M_{av} = 143$  kDa in addition to the free monomer accounts well for the in-cell  $R_1$  and  $R_2$  values of all six protein variants (Table S2). To assure that we get the same statistical weight for both  $R_1$  and  $R_2$  data in the fitting procedure, we use reduced relaxation parameters according to

$$R_i^j = (R_i^{j,\text{OBS}} - R_i^{\text{av}}) / \sigma_i \quad (2)$$

where  $j$  denotes reporter protein  $j$ ,  $i = 1 \parallel 2$ , brackets denote the average over the full dataset (e.g. all  $R_2$  values), and  $\sigma$  is the standard deviation of the dataset. The correlation between the observed and calculated  $R_1$  now approaches a unit line with  $r^2 = 0.93$  and  $\text{rmsd} = 0.26$  (Figure 3). That is, the in-cell effect can indeed be described by alterations of  $\tau_r$  alone, without the need to invoke additional chemical-exchange contributions. Furthermore, this shows that both  $R_1$  and  $R_2$  can be reconciled in a model where all proteins “feel” the same interaction environment and where the sole difference is the bound fraction. Intriguingly, and somewhat surprisingly, all reporter proteins show relatively low  $p_B$  values (Table S2) despite being in the crowded cytosol. Still, the inert, soluble, and, mainly, negatively charged reporter proteins are expected to be kept soluble by charge repulsion with the negative surroundings, reducing the amount of potential complex-forming contacts.<sup>9</sup> Further, also in the presence of high concentration of the positively charged lysozyme, the negatively charged TTHA<sup>pwt</sup> shows low  $p_B$ , confirmed by fluorescence experiments (Figure S5), which underlines the low interactivity of this protein. Furthermore, the excellent quality of all the reporter proteins' in-cell NMR spectra indicates a low  $p_B$ , as a highly populated complex of high molecular weight would yield a significant increase in  $R_2$  if substantially populated (Figure 2).

To further test this result, we examined next how well the value of  $M^{\text{av}} = 143$  kDa actually agrees with the various sizes of transient complexes expected to be formed in the cell. As a base for comparison, we used all human proteins in the UniProt database<sup>34</sup> annotated as cytosolic (Supporting Information Methods). The protein-mass distribution of this data set complies with both  $\Gamma$ - and lognormal distributions,<sup>35</sup> where the latter gives a somewhat better fit (Figure 3).

Notably, the employed protein-data set yields no information of abundance, which might bias the distribution. To test for such bias, we used an alternative data set from Geiger *et al.*,<sup>23</sup> listing the individual sequences and relative abundance of lysate proteins from several mammalian cancer cell lines. The result shows that the abundance-weighted distribution agrees well with the original cytosolic subset from the UniProt database (Supporting Information Methods, Figure S6). On this basis, we stick to the lognormal distribution from the cytosolic subset as an estimate for the interaction partner protein mass distribution, yielding an average protein mass of

$M^{\text{av}} = 73$  kDa. Upon lowering the mass in eq 1 from  $M_{av} = 143$  kDa to  $M^{\text{av}} = 73$  kDa, however, the agreement with the observed relaxation data becomes compromised (Supporting Information 3, Table S2). The reason why the 143 to 73 kDa decrease cannot simply be compensated for by a higher population of bound species is due to the distinct mass dependence of  $R_1$  and  $R_2$  (Figure 1): for a given molecular mass of the transient complex, a change in bound population to fit one type of in-cell relaxation will inevitably lead to a coupled change in the other and concomitantly to a mismatch in most cases.

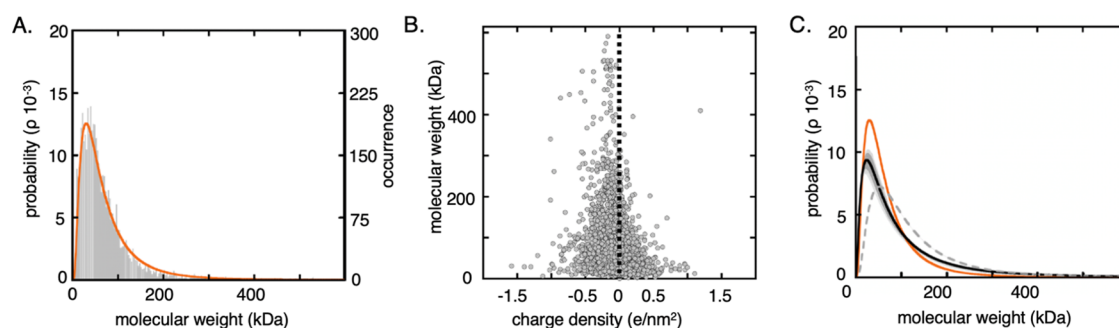
#### Approaching a Physiologically Relevant Situation.

Even if we can reconcile both types of relaxation with a two-state binding model, a uniform protein mass does not realistically describe the in-cell situation. To better account for the natural protein-mass heterogeneity, we developed the model to include not only the encounters with a single average partner but with a distribution of species matching the cytosolic protein sizes. A set of putative binding partners were obtained by integrating over the full mass distribution of cytosolic proteins. This allows us to obtain the mass-weighted relaxation rates of the protein-encounter complexes (eq 2). Notably, the fast-exchange model in eq 1 relies on the assumption that the reporter protein is equally likely to collide and interact with all molecules in the distribution. Hence, as the surface properties are crucial for interaction and encounter formation,<sup>6,10,17</sup> the net-charge density of different intervals in the mass distribution must overlap. As a control, we divided the distribution into three size regimes, that is, <70, 70–140, and >140 kDa, and calculated the net-charge density<sup>9</sup> (Supporting Information 5). The result confirms that the surface-charge distribution of the three subensembles indeed overlaps nicely (Figure S7). This uniform mass–charge relation allows us to include the full distribution of binding partners in the fast exchange model described in eq 1, where the relaxation rate of the bound state now is given by

$$R_i^{\text{bound}} = \frac{\int_0^\infty \rho(M_w) F_{R_i}(M_j + M_w) dM_w}{\int_0^\infty \rho(M_w) dM_w} \quad (3)$$

where  $\rho(M_w)$  is the size distribution of interaction partners and  $F_{R_i}(M_j + M_w)$  is the closed-form expression for the relaxation rate for a complex between protein  $j$  and a binding partner with mass  $M_w$  (Supporting Information 3). Optimization of the bound population ( $p_B$ ) for each reporter protein using eqs 1 and 3 and integration over the database-derived lognormal  $\rho(M_w)$  (Figure 3, Table S2) improve the agreement between observed and calculated  $R_1$  values, compared to the single average mass of the cytosolic proteins. Yet, the lower  $r^2 = 0.87$  and  $\text{rmsd} = 0.60$ , together with the systematic underestimate of calculated  $R_1$ , indicate that the distribution of masses of the cytosolic proteins remains shifted towards too small proteins to account for the observed in-cell retardation.

**Accounting for Larger Cellular Complexes in the Size Distribution Finally Reconnects the  $T_1$  and  $T_2$  Relaxations.** Internalized proteins encounter not only other monomeric proteins but interact also with larger cellular components. Although a diffusing protein is most likely to collide with other proteins, simply because of their large proportion of the cellular dry weight,<sup>36,37</sup> it will also encounter other macromolecular structures and surfaces of larger dimensions. Examples of such structures are the cytoskeleton,



**Figure 4.** Size distribution estimated from in-cell relaxation data agrees well with the database-derived size distribution of cytosolic proteins. (A) Normalized histogram representation of the mass distribution of the cytosolic proteome of human proteins from the Uniprot database,<sup>34</sup> with 5217 proteins. A lognormal probability density function was fitted to the histogram (orange line). (B) Net charge density shows an approximately normal distribution centred at  $-0.065 \text{ e/nm}^2$ . The dashed line corresponds to zero charge. (C) Optimized size distributions combining  $R_1$  and  $R_2$  data from the six reporter proteins into a single binding model. The optimized lognormal distribution (black) compared to the database-derived distribution (orange), corresponding to the fitted distribution in (A). The gray shaded area depicts the variation upon random removal of one (dark gray) or two (bright gray) relaxation pairs. The dashed gray line is the maximum entropy distribution from the family of solutions (Supporting Information 6, Figures S8, S9).

membranes, and ribosomes. The amount of transient interactions is mainly determined by the overall macromolecular concentration and surface properties of the interacting molecules,<sup>6,10</sup> where the surface charge seems to be a key determinant. To a first approximation, the surface-charge density of the larger cellular structures can be assumed to follow the same distribution as the cytosolic proteins (Figure 4). Most clearly, the cytoskeleton is primarily composed of tubulin, actin, and lamin, all of which show similar net negative surface charge density as soluble proteins. The surface architectures of membranes and ribosomes, however, are partly distinct from proteins and need extra consideration. Mammalian membranes are dynamic bilayers, where the fraction of anionic lipids is between 10 and 30%.<sup>38,39</sup> This translates to a surface net charge of  $-0.15$ – $-0.45 \text{ e/nm}^2$ , which is somewhat more negative than for the average protein ( $-0.065 \text{ e/nm}^2$ ). Considering also the presence of  $\sim 1/10 \text{ nm}^2$  integral membrane proteins<sup>40</sup> with a positive-inside orientation<sup>41,42</sup> and that the lipid bilayer also consists of nonlipid, noncharged alcohols such as cholesterol, the effective surface charge density can still be assumed to fall in the same range as protein surfaces. Ribosomes are highly negative, abundant megadalton entities,<sup>43,44</sup> where the high surface charge density places ribosomes in the negative tail of the net charge distribution (Figure 4). Nonetheless, even in this case transient interactions seem to follow a similar net charge dependence as protein–protein interactions,<sup>6,10,12</sup> where positively charged proteins even form semistable complexes with ribosomes.<sup>12,45</sup> Indeed, for several proteins, the observed line broadening in in-cell NMR experiments has been assigned to mainly stem from transient ribosome interactions.<sup>46,47</sup>

An additional high-mass contribution is the simultaneous interaction between multiple partners,<sup>48</sup> albeit that this possibility seems here disfavored by the low population of complexes formed by our current reporter proteins. Since the relaxation effect upon interaction with more than one partner is, nevertheless, indistinguishable from that with a single large entity, it is reasonable to include all transient complexes as a part of the same size and net-charge density distribution. To allow for such larger complexes to be included, we optimized  $p_B$  for each reporter protein while simultaneously optimizing a common mass distribution of interaction partners. For this purpose, we employed a generalized lognormal distribution

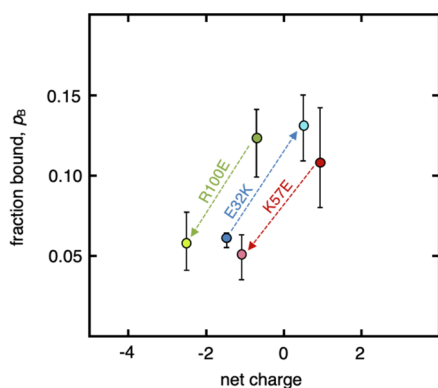
with free optimization of the distribution parameters (Supporting Information 3). The fit optimizes  $n + 2$  parameters for  $n$  pairs of relaxation rates ( $p_B$  for each reporter protein and 2 global distribution parameters), which means that at least two different data pairs are needed for a robust fit. We tested this approach to analyze transient binding of TTHA<sup>PWT</sup> to lysozyme *in vitro* and used the two  $R_1/R_2$  pairs from the 50 and 150 mg/mL lysosome samples to fit a distribution of  $M_w$ . Reassuringly, a narrow distribution centered around 15 kDa was obtained (Figure S5).

Next, the six pairs of in-cell relaxation rates were used in a global fit, where  $R_1$  or  $R_2$  alone are not enough to determine the distribution, as for a single relaxation rate, any distribution can be accommodated by a shift in  $p_B$ . We find that the distribution that best reproduces the relaxation data indeed resembles the one predicted from the natural intracellular environment: the distribution not only fits the sizes of the soluble proteins but also includes a high-mass tail that accounts for the larger cytosolic components (Figure 4, Table S2). The fitted relaxation rates are virtually identical to those from the fast-exchange model with a uniform partner of  $M^v = 143 \text{ kDa}$  (Figure 3), with good agreement between calculated and observed  $R_1$  values, where  $r^2 = 0.93$  and  $\text{rmsd} = 0.26$  (Figure 3). The distribution optimization from in-cell data yields a family of solutions for a given  $p_B$  (Figure S8), which, in turn, allows the maximum-entropy distribution to be deducted (Figure S9, Supporting Information 6). This is a distribution that explains the data, while still carrying minimal “a priori” information. As the relaxation data can be explained by a single average binding partner with mass  $M^{av}$ , a symmetric distribution centered at  $M^{av}$  would provide the distribution with less information. However, the obtained maximum entropy solution resembles the asymmetrical database-derived distribution (Figure 4) with the characteristic high mass tail.

Although our extended distribution analysis does not lead to higher precision, it strengthens the approach by demonstrating that the NMR data also complies with a physiologically realistic cell composition. Moreover, it validates the ansatz of fast exchange in transient encounters and provides a method for estimating the size distribution of macromolecules at work in live cells. To determine the robustness of the obtained distribution, we randomly removed first one and then two data pairs and repeated the fit. The test shows, reassuringly, that the

distribution only exhibits small changes (Figure 4). As an additional control, we examined if the obtained distribution still reflects the intrinsic physical coupling between translational and transverse relaxation. In other words, can we accommodate any set of “inconsistent”  $R_1/R_2$  pairs with a distribution, simply by rendering our conclusions about transient encounters nonconclusive? To test this, we prepared a set of relaxation pairs with  $R_1$  values systematically offset to lower values, while keeping  $R_2$  unchanged. The results show that we cannot obtain any distribution that reproduces these relaxation pairs with the same precision as the real data (Supporting Information 7, Figure S10).

**Bound Population as Quantification of Transient Interactions in the Cell.** Thus, use of the full distribution as possible binding partners explains both relaxation parameters using a single physically relevant model: the assumption of fast exchange between a bound and a free state (eq 1). Additionally, it provides an intuitively apprehensible parameter in the form of the population of bound protein species ( $p_B$ ). The latter is here the sole specific parameter for the reporter protein and constitutes a direct measure of the actual interactivity of the protein inside the studied cell type. Upon comparing the obtained  $p_B$  values with the apparent viscosity (determined from  $R_2$  alone), there is a common relationship with surface charge: less negative net charge results in both higher  $\eta^{\text{app}}$  and  $p_B$ . This relationship is further underlined by the clear response on surface charge mutations (Figure 5,



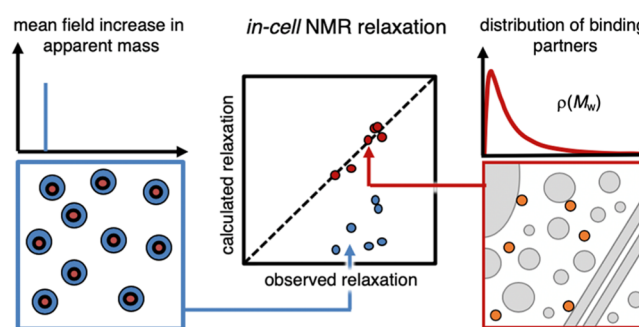
**Figure 5.** Determined bound population as a function of net charge. The optimized population from a model with transient interaction of the reporter proteins with a distribution of interaction partners, blue marker: TTHA<sup>PWT</sup>; red marker: HAH1<sup>PWT</sup>; and green marker: SOD1<sup>barrel</sup>. The brighter markers correspond to the surface mutation variants of the reporter proteins and highlight the importance of surface net charge on in-cell transient encounter formations.

Table S2). Notably, the response on the surface-charge perturbations seems similar for all three reporter proteins, where  $p_B$  approximately doubles with a net-charge change by two units. The reporter proteins, however, follow distinct trajectories in the net charge- $p_B$  plane (Figure 5), probably reporting on the differences in general inter-reactivity. At the same net charge, SOD1<sup>barrel</sup> stands out as the most interactive of the reporter proteins, while HAH1<sup>PWT</sup> tumbles most freely (Figure 5). This result is in contrast to the more rudimentary  $\eta^{\text{app}}$  analysis, where SOD1<sup>barrel</sup> emerges as less retarded (Figure 2). The reason for this underestimate of SOD1<sup>barrel</sup> interactivity in the  $\eta^{\text{app}}$  analysis is that the relative change in the apparent size upon transient interactions is less for a larger protein than for a smaller one,<sup>10</sup> while the  $p_B$  analysis takes this explicitly

into account. These results serve as a good example of the advantage of using  $p_B$  determined from both longitudinal and transverse relaxation when characterizing transient interactions in live cells. Since the other readout from the global analysis is the effective size distribution, this method can also be used to observe cell-type differences and perturbations of the macromolecular size distribution in the cytosol. Accordingly, the approach can relatively simply provide new insights into cell function, cellular composition, and the formation of higher-order interactomes in live cells.

## CONCLUDING REMARKS

To summarize, our study shows that the mean-field approaches based on apparent viscosity fail to accurately describe the in-cell effect on NMR relaxation data since they do not explicitly account for fast exchange between the monomer and monomer-partner complexes that is bound to accompany transient-binding events (Figure 6). Taking this fast exchange



**Figure 6.** Comparing the models. The mean field approach, where a reporter protein (orange) is assigned an apparent mass (blue panel), cannot simultaneously describe in-cell  $R_1$  and  $R_2$  data. However, a model with the reporter protein in a single free state and a population in a distribution of bound states fully reconciles the  $R_1$  and  $R_2$  data. The distribution of monomeric cytosolic proteins is not sufficient by itself but transient interactions with larger components, such as protein assemblies, membranes, ribosomes, and cytoskeleton have to be accounted for.

into account not only reconciles the previous issue of seemingly inconsistent relaxation parameters, but also sheds light on the intracellular interactions through pinpointing the population of bound species ( $p_B$ ). Of particular interest, the results show that the disparate effect of transient in-cell interactions on  $T_1$  and  $T_2$  can be explained by mass-altering binding alone without introducing chemical-exchange contributions (Figure 6). Although this simplifying result by no means rules out contributions from chemical-exchange effects, it is notable that they are not required to explain the observed data.

Another interesting detail is that to accurately reproduce the observed relaxation rates, larger interaction partners than the soluble cytosolic proteins need to be included (Figure 6). This observation complies with the view that both soluble proteins and higher-order complexes, like for example, ribosomes and cytoskeletons, play important roles in modulating the rotational diffusion in live cells.<sup>5,12,28,46,47,49</sup> The influence of the full distribution of intracellular interaction partners on the rotational diffusion suggests also that the NMR relaxation analysis can be used for exploring more intricate aspects of cellular function. One such example is how the interplay between the intracellular components responds to physio-



logical or genetic perturbations, where the interaction-size distributions for some key players are expected to undergo significant changes. Although it remains to establish how far this type of analysis can be taken given the signal-to-noise and many degrees of freedom, the results in this study show that, in principle, it is doable.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.0c08274>.

Closed-form expression for the mass dependence of the relaxation rates, determination of  $\eta^{\text{app}}$ , the binding models, validation of the models from lysozyme binding, surface net charge estimations, maximum entropy calculations, and validation of  $R_1/R_2$  self-consistency; materials and methods; in-cell transverse and longitudinal relaxation; protein leakage quantification; correlation between leakage and deviation from single exponentiality; apparent viscosity reference curves for longitudinal and transverse relaxation; benchmarking the method using weak interaction between TTHA<sup>Pwt</sup> and human lysozyme; comparison of the cytosolic proteome size-distributions; histogram of surface charge density for proteins of three different size ranges; family of solutions for possible size distributions of the cytosolic interactome; selected representatives for the cytosolic interactome; relaxation rate parameters need to be globally consistent to give reasonable results; longitudinal and transverse relaxation rates; and collected results for the calculations of both relaxation rates (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Mikael Oliveberg** – Department of Biochemistry and Biophysics, Arrhenius Laboratories of Natural Sciences, Stockholm University, Stockholm 106 91, Sweden; [orcid.org/0000-0003-1919-7520](https://orcid.org/0000-0003-1919-7520); Email: [mikael@dbb.su.se](mailto:mikael@dbb.su.se)

**Jens Danielsson** – Department of Biochemistry and Biophysics, Arrhenius Laboratories of Natural Sciences, Stockholm University, Stockholm 106 91, Sweden; [orcid.org/0000-0002-6048-6896](https://orcid.org/0000-0002-6048-6896); Email: [jens.danielsson@dbb.su.se](mailto:jens.danielsson@dbb.su.se)

### Authors

**Sarah Leeb** – Department of Biochemistry and Biophysics, Arrhenius Laboratories of Natural Sciences, Stockholm University, Stockholm 106 91, Sweden

**Fan Yang** – Department of Biochemistry and Biophysics, Arrhenius Laboratories of Natural Sciences, Stockholm University, Stockholm 106 91, Sweden; [orcid.org/0000-0002-9616-6552](https://orcid.org/0000-0002-9616-6552)

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jpcb.0c08274>

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The research was funded by the Knut and Alice Wallenberg Foundation (2017-0041), the Swedish Research Council (2017-01517), and the Magnus Bergvall foundation (2017-02228).

## ■ REFERENCES

- (1) Berg, O. G.; von Hippel, P. H. Diffusion-controlled macromolecular interactions. *Annu. Rev. Biophys. Biophys. Chem.* **1985**, *14*, 131–160.
- (2) McGuffee, S. R.; Elcock, A. H. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput. Biol.* **2010**, *6*, e1000694.
- (3) Sarkar, M.; Li, C.; Pielak, G. J. Soft interactions and crowding. *Biophys. Rev.* **2013**, *5*, 187–194.
- (4) McConkey, E. H. Molecular evolution, intracellular organization, and the quinary structure of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1982**, *79*, 3236–3240.
- (5) Majumder, S.; Xue, J.; DeMott, C. M.; Reverdatto, S.; Burz, D. S.; Shekhtman, A. Probing protein quinary interactions by in-cell nuclear magnetic resonance spectroscopy. *Biochemistry* **2015**, *54*, 2727–2738.
- (6) Mu, X.; Choi, S.; Lang, L.; Mowray, D.; Dokholyan, N. V.; Danielsson, J.; Oliveberg, M. Physicochemical code for quinary protein interactions in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E4556–E4563.
- (7) Schreiber, G.; Fersht, A. R. Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol.* **1996**, *3*, 427–431.
- (8) Camacho, C. J.; Weng, Z.; Vajda, S.; DeLisi, C. Free energy landscapes of encounter complexes in protein-protein association. *Biophys. J.* **1999**, *76*, 1166–1178.
- (9) Wennerström, H.; Vallina Estrada, E.; Danielsson, J.; Oliveberg, M. Colloidal stability of the living cell. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 10113–10121.
- (10) Leeb, S.; Sørensen, T.; Yang, F.; Mu, X.; Oliveberg, M.; Danielsson, J. Diffusive protein interactions in human versus bacterial cells. *Curr Opin Struct Biol* **2020**, *2*, 68–78.
- (11) Ye, Y.; Wu, Q.; Zheng, W.; Jiang, B.; Pielak, G. J.; Liu, M.; Li, C. Positively Charged Tags Impede Protein Mobility in Cells as Quantified by (19)F NMR. *J. Phys. Chem. B* **2019**, *123*, 4527–4533.
- (12) Schavemaker, P. E.; Smigiel, W. M.; Poolman, B. Ribosome surface properties may impose limits on the nature of the cytoplasmic proteome. *Elife* **2017**, *6*, e30084.
- (13) Ye, Y.; Wu, Q.; Zheng, W.; Jiang, B.; Pielak, G. J.; Liu, M.; Li, C. Quantification of size effect on protein rotational mobility in cells by (19)F NMR spectroscopy. *Anal. Bioanal. Chem.* **2018**, *410*, 869–874.
- (14) Palmer, A. G., 3rd. Nmr probes of molecular dynamics: overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 129–155.
- (15) Theillet, F.-X.; Binolfi, A.; Bekei, B.; Martorana, A.; Rose, H. M.; Stuver, M.; Verzini, S.; Lorenz, D.; van Rossum, M.; Goldfarb, D.; Selenko, P. Structural disorder of monomeric alpha-synuclein persists in mammalian cells. *Nature* **2016**, *530*, 45–50.
- (16) Lipari, G.; Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* **1982**, *104*, 4559–4570.
- (17) Barbieri, L.; Luchinat, E.; Banci, L. Protein interaction patterns in different cellular environments are revealed by in-cell NMR. *Sci. Rep.* **2015**, *5*, 14456.
- (18) Wall, K. P.; Hough, L. E. In-Cell NMR within Budding Yeast Reveals Cytoplasmic Masking of Hydrophobic Residues of FG Repeats. *Biophys. J.* **2018**, *115*, 1690–1695.
- (19) Ye, Y.; Liu, X.; Zhang, Z.; Wu, Q.; Jiang, B.; Jiang, L.; Zhang, X.; Liu, M.; Pielak, G. J.; Li, C. <sup>19</sup>F NMR spectroscopy as a probe of cytoplasmic viscosity and weak protein interactions in living cells. *Chemistry* **2013**, *19*, 12705–12710.

- (20) Wang, Q.; Zhuravleva, A.; Gierasch, L. M. Exploring weak, transient protein–protein interactions in crowded *in vivo* environments by in-cell nuclear magnetic resonance spectroscopy. *Biochemistry* **2011**, *50*, 9225–9236.
- (21) Zorrilla, S.; Hink, M. A.; Visser, A. J. W. G.; Lillo, M. P. Translational and rotational motions of proteins in a protein crowded environment. *Biophys. Chem.* **2007**, *125*, 298–305.
- (22) Danielsson, J.; Kurnik, M.; Lang, L.; Oliveberg, M. Cutting off functional loops from homodimeric enzyme superoxide dismutase 1 (SOD1) leaves monomeric beta-barrels. *J. Biol. Chem.* **2011**, *286*, 33070–33083.
- (23) Geiger, T.; Wehner, A.; Schaab, C.; Cox, J.; Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* **2012**, *11*, M111 014050.
- (24) Danielsson, J.; Mu, X.; Lang, L.; Wang, H.; Binolfi, A.; Theillet, F.-X.; Bekei, B.; Logan, D. T.; Selenko, P.; Wennerström, H.; Oliveberg, M. Thermodynamics of protein destabilization in live cells. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 12402–12407.
- (25) Inomata, K.; Kamoshida, H.; Ikari, M.; Ito, Y.; Kigawa, T. Impact of cellular health conditions on the protein folding state in mammalian cells. *Chem. Commun.* **2017**, *53*, 11245–11248.
- (26) Gnutt, D.; Timr, S.; Ahlers, J.; König, B.; Manderfeld, E.; Heyden, M.; Sterpone, F.; Ebbinghaus, S. Stability Effect of Quinary Interactions Reversed by Single Point Mutations. *J. Am. Chem. Soc.* **2019**, *141*, 4660–4669.
- (27) Monteith, W. B.; Cohen, R. D.; Smith, A. E.; Guzman-Cisneros, E.; Pielak, G. J. Quinary structure modulates protein stability in cells. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 1739–1742.
- (28) Crowley, P. B.; Chow, E.; Papkovskaia, T. Protein interactions in the *Escherichia coli* cytosol: an impediment to in-cell NMR spectroscopy. *Chembiochem* **2011**, *12*, 1043–1048.
- (29) Cohen, R. D.; Pielak, G. J. Electrostatic Contributions to Protein Quinary Structure. *J. Am. Chem. Soc.* **2016**, *138*, 13139–13142.
- (30) McConnell, H. M. Reaction Rates by Nuclear Magnetic Resonance. *J. Chem. Phys.* **1958**, *28*, 430–431.
- (31) Palmer, A. G., 3rd. Chemical exchange in biomacromolecules: past, present, and future. *J. Magn. Reson.* **2014**, *241*, 3–17.
- (32) Berlin, K.; Longhini, A.; Dayie, T. K.; Fushman, D. Deriving quantitative dynamics information for proteins and RNAs using ROTDIF with a graphical user interface. *J. Biomol. NMR* **2013**, *57*, 333–352.
- (33) Mourão, M. A.; Hakim, J. B.; Schnell, S. Connecting the dots: the effects of macromolecular crowding on cell physiology. *Biophys. J.* **2014**, *107*, 2761–2766.
- (34) Consortium, U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515.
- (35) Tiessen, A.; Perez-Rodriguez, P.; Delaye-Arredondo, L. J. Mathematical modeling and comparison of protein size distribution in different plant, animal, fungal and microbial species reveals a negative correlation between protein size and protein number, thus providing insight into the evolution of proteomes. *BMC Res. Notes* **2012**, *5*, 85.
- (36) Yu, I.; Mori, T.; Ando, T.; Harada, R.; Jung, J.; Sugita, Y.; Feig, M. Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *Elife* **2016**, *5*, e19274.
- (37) Cheung, M. C.; LaCroix, R.; McKenna, B. K.; Liu, L.; Winkelman, J.; Ehrlich, D. J. Intracellular protein and nucleic acid measured in eight cell types using deep-ultraviolet mass mapping. *Cytometry, Part A* **2013**, *83*, 540–551.
- (38) (a) Nagle, J. F.; Tristram-Nagle, S. Structure of lipid bilayers. *Biochim. Biophys. Acta* **2000**, *1469*, 159–195. (b) Platre, M. P.; Jaillais, Y. Anionic lipids and the maintenance of membrane electrostatics in eukaryotes. *Plant Signaling Behav.* **2017**, *12*, e1282022.
- (39) Lemmon, M. A. Membrane recognition by phospholipid-binding domains. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 99–111.
- (40) Takamori, S.; Holt, M.; Stenius, K.; Lemke, E. A.; Grønborg, M.; Riedel, D.; Urlaub, H.; Schenck, S.; Brügger, B.; Ringler, P.; Müller, S. A.; Rammner, B.; Gräter, F.; Hub, J. S.; De Groot, B. L.; Mieskes, G.; Moriyama, Y.; Klingauf, J.; Grubmüller, H.; Heuser, J.; Wieland, F.; Jahn, R. Molecular anatomy of a trafficking organelle. *Cell* **2006**, *127*, 831–846.
- (41) Heijne, G. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the transmembrane topology. *EMBO J.* **1986**, *5*, 3021–3027.
- (42) von Heijne, G.; Gavel, Y. Topogenic signals in integral membrane proteins. *Eur. J. Biochem.* **1988**, *174*, 671–678.
- (43) Milo, R.; Jorgensen, P.; Moran, U.; Weber, G.; Springer, M. BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **2010**, *38*, D750–D753.
- (44) Fedyukina, D. V.; Jennaro, T. S.; Cavagnero, S. Charge segregation and low hydrophobicity are key features of ribosomal proteins from different organisms. *J. Biol. Chem.* **2014**, *289*, 6740–6750.
- (45) Burz, D. S.; Breindel, L.; Shekhtman, A. The Inescapable Effects of Ribosomes on In-Cell NMR Spectroscopy and the Implications for Regulation of Biological Activity. *Int* **2019**, *20*, 1297.
- (46) DeMott, C. M.; Majumder, S.; Burz, D. S.; Reverdatto, S.; Shekhtman, A. Ribosome Mediated Quinary Interactions Modulate In-Cell Protein Activities. *Biochemistry* **2017**, *56*, 4117–4126.
- (47) Breindel, L.; Yu, J.; Burz, D. S.; Shekhtman, A. Intact ribosomes drive the formation of protein quinary structure. *PLoS One* **2020**, *15*, e0232015.
- (48) von Bülow, S.; Siggel, M.; Linke, M.; Hummer, G. Dynamic cluster formation determines viscosity and diffusion in dense protein solutions. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 9843–9852.
- (49) Kyne, C.; Ruhle, B.; Gautier, V. W.; Crowley, P. B. Specific ion effects on macromolecular interactions in *Escherichia coli* extracts. *Protein Sci.* **2015**, *24*, 310–318.
- (50) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.