

Diffusive protein interactions in human versus bacterial cells

SARAH LEEB[§], THERESE SÖRENSEN[§], FAN YANG, XIN MU, MIKAEL OLIVEBERG* AND JENS DANIELSSON*

Department of Biochemistry and Biophysics, Arrhenius Laboratories of Natural Sciences, Stockholm University, S-106 91 Stockholm, Sweden.

Supporting Information

S1. Supporting methods

S1.1 Reporter proteins.

TTHA^{pwt} is derived from the Cu-binding protein TTHA1718 from *Thermus thermophilus*, by replacing the metal-coordinating cysteines with polar serines [1] [2]. HAH1^{pwt} is the analogously altered copper chaperone from humans [3] [2], and SOD1^{barrel} is the barrel scaffold of human SOD1, where the long catalytic loops have been truncated to short Gly-Ala-Gly linkers [4]. The active-site removals serve to eliminate specific *in-cell* interactions and bias the system towards the diffusive background interactions that are the focus of this study. Removal of the catalytic loops disrupts the SOD1 dimer interface and renders SOD1^{barrel} to be monomeric, also at high concentrations [4].

S1.2 Protein mutagenesis, production and purification.

Design and mutagenesis of the model proteins TTHA^{pwt}, HAH1^{pwt}, HAH1^{K57E} and SOD1^{barrel} have been extensively described in Mu et al. [5] and Danielsson et al. [6]. Expression was as follows: Plasmids carrying the protein-encoding genes and a carbenicillin resistance marker were transformed into *E. coli* BL21(DE3) cells and grown overnight on LB/Agar-plates, supplemented with carbenicillin. The colonies were then used to seed overexpression cultures grown in M9 minimal media at 37 °C with ¹⁵NH₄Cl as sole nitrogen source. Protein overexpression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at OD₆₀₀ ≈ 0.8, and left to progress for 18 h at 23 °C. Harvested *E. coli* cells were finally lysed by sonication and cell debris was removed by centrifugation at 39 000 g for 20 min in a JA 25.50 rotor (Beckman Coulter, California, USA) at 4 °C. Purification of TTHA^{pwt} was as described in Mu et al. [5] and that of SOD1^{barrel} as described in Danielsson et al. [6]. The latter purification protocol was also used for HAH1^{pwt} and HAH1^{K57E}, with the difference that the fraction precipitating between 30 % and 90 % (NH₄)₂SO₄ was collected and further purified by a S-100 gel filtration step. Protein purity was verified by 4 - 20 % precast SDS-PAGE gels (BioRad, California, USA).

S1.3 In-cell NMR sample preparation and in-vitro experimental conditions.

Human ovarian adenocarcinoma (A2780) cells purchased from Sigma-Aldrich (Missouri, USA) were grown in RPMI 1640 growth medium (Life Technologies, California, USA) supplemented with 10 % Fetal Bovine Serum (Life Technologies), 1 % Antibiotics-Antimycotics mixture (Life Technologies) and 0.45 $\mu\text{g/ml}$ Plasmocin (InvivoGen, California, USA) at 37 °C with 5 % CO₂ in a Hepa Class 100 steri-cycle CO₂ incubator (Thermo Fisher, Massachusetts, USA). Culture dishes with a confluence of ~ 60 - 90 % were used for all *in-cell* NMR samples. Prior to harvest, all cells were washed twice with 5 ml 1 x DPBS (Life Technologies). Cells were detached by trypsination (Trypsin/EDTA in DPBS, Life Technologies) for 3 min at 37 °C. Harvested cells were washed twice in 50 ml OptiMEM (Life Technologies). To remove the medium, the cell suspension was centrifuged at 200 g for 5 min at 25 °C and the supernatant discarded. ¹⁵N-labeled protein was added to a final concentration of 1.5 mM and the electroporation solution was divided into 100 μl aliquots in electroporation cuvettes corresponding to ~ 3×10^6 cells/cuvette.

Electroporation was carried out with a single 14 ms poring pulse at 115 V followed by five 50 ms 20 V transfer pulses separated by 50 ms delays (NEPA21 Super Electroporator, Nepa Gene Co., Ichikawa, Japan). The pulse parameters were slightly modulated to keep the energy in the samples in the range of 4 - 4.5 J. After electroporation, growth medium was added to the cuvettes, the suspensions were carefully transferred to culture dishes with 10 ml growth medium and left to recover at 37 °C for 4 - 5 hours, after which dead cells and excess (extracellular) protein were carefully washed away twice with 5 ml 1 x DPBS, and viable attached cells were harvested. The harvested cells were washed twice with 50 ml OptiMEM. Cell viability after electroporation was 34 - 75 %. The cells were re-suspended in medium (~300-500 μl OptiMEM), and supplemented with D₂O (Sigma-Aldrich) to a final concentration of 10 %. The sample was transferred to a 4 mm (< 30×10^6 cells) or 5 mm (> 30×10^6 cells) NMR tube (BMS-004B/BMS-005B, Shigemi Inc., Tokyo, Japan). OptiMEM was carefully added on top of the cell suspension to a final volume of 700 or 900 μl , depending on the size of the NMR tube. In order to gather all cells

at the bottom of the NMR tube, i.e. within the detection volume of the NMR spectrometer, the sample was gently, manually centrifuged prior to starting the NMR experiment. *In-vitro* control NMR samples were in 10 mM MES pH 6.4 complemented by 150 mM NaCl and 10 % D₂O, unless stated otherwise.

S1.3.1 In-cell NMR sample preparation in E.coli.

Plasmids carrying the target protein gene were transformed into the *E. coli* strain BL21(DE3)pLysS. The cells were grown in 100 mL LB including the relevant antibiotics until they reached an OD₆₀₀ of 0.6 – 0.8. The cells were then harvested by gentle centrifugation at 800 x g for 8 min at 25 °C. After removing the supernatant, the cell pellet was resuspended and transferred to 100 ml fresh M9 medium supplemented with ¹⁵N-enriched ammonium chloride and relevant antibiotics. After 30 min incubation at 37 °C on a shaking table, 0.5 mM IPTG was added to induce the over-expression of the target protein. The cells were further incubated for 2.5 hours and subsequently harvested by centrifugation at 800 x g for 8 min at 25 °C. The cell pellet was then resuspended in fresh M9 buffer (0.02 M KH₂PO₄, 0.04 M Na₂HPO₄, 0.1 M NaCl, 1 mM MgSO₄) and transferred into an NMR tube. The cell slurry was compacted into a dense pellet by spinning the tube manually and gently for 5 to 10 min.

S1.4 NMR spectroscopy.

All *in-cell* and lysate experiments were carried out on a Bruker Avance III 700 MHz spectrometer with a triple-resonance cryoprobe. Two-dimensional spectra were recorded using a ¹H-¹⁵N-band-selective optimized flip-angle short-transient heteronuclear multiple quantum coherence (SOFAST-HMQC) pulse sequence including excitation sculpting-based water suppression. The number of scans was put to 2048, the number of increments in the indirect ¹⁵N-dimension to 64 at a spectral width of 34 ppm (13 ms acquisition time). The acquisition time in the ¹H-dimension was set to 50 ms and the inter-scan delay to 40 ms amounting to a total measurement time of ~ 4.5 hours. All experiments were flanked by one 1D ¹⁵N-filtered ¹H-SOFAST-HMQCs immediately prior and after the experiments, with 8192 scans and

40 ms acquisition time taking ~ 13 min to run. Combined the measurements took less than 5 hours to perform.

For the transverse relaxation measurements, a one-dimensional ^{15}N -filtered heteronuclear single quantum coherence (HSQC)-based pulse sequence was used that employs Carr-Purcell-Meiboom-Gill (CPMG) pulse trains during the delay times to refocus any line broadening caused by field inhomogeneities or chemical exchange processes. For each relaxation experiment, data points at 3 different relaxation delays were collected. The inter-scan delay was set to 1 s and the acquisition time to 125 ms. In mammalian cells 5120 scans were recorded amounting to a total measurement time of ~ 5 hours, while in *E. coli* – due to the large concentrations of over-expressed protein – 512 scans (~ 30 min acquisition) were enough to get sufficient signal. The experiments were once again flanked by one-dimensional ^{15}N -filtered ^1H -SOFAST-HMQCs with 8192 scans and an acquisition time of 40 ms (~ 13 min) in the case of mammalian cells and 2048 (~ 3 min) in the case of *E. coli*. The total measurement time amounted to ~ 5.5 hours and to ~ 40 min for the two organisms respectively.

All *in-vitro* reference spectra including the viscosity and pH series were performed on the same spectrometer. All *in-cell* and *in-vitro* experiments were measured at 310 K. The *in-vitro* control experiments regarding the legitimacy of using a spectral fraction of a one-dimensional data set to obtain transverse relaxation rates were performed on a Bruker Avance III 600 MHz spectrometer without cryoprobe and at 298 K. All NMR data were processed using either the TopSpin (Bruker, Massachusetts, USA) or NMRPipe software[7].

SI.5 NMR line width determination.

The line widths were quantified by fitting a Gaussian function to a 1D slice in the ^1H -dimension of the *in-cell* and *in-vitro* control SOFAST-HMQC spectra for each well-resolved cross peak. The number of cross peaks that met this criterium was 44 of a total of 66 residues for TTHA^{PWT}, 41 of a total of 68 residues for HAH1^{PWT} and 65 of a total of 110 residues for SOD1^{barrel}. The fitting was performed using the built-in line shape tool in the NMRFAM-SPARKY software [8], and each fit was manually inspected and evaluated for fit quality. No window functions were employed for spectra that were subsequently used for line broadening analysis as they would have distorted the peak widths.

S1.6 Determining transverse relaxation rates.

In the case of one-dimensional relaxation data, signal intensities were determined by integrating over the spectral regions 9.4-8.6 ppm for TTHA^{pwt} as well as HAH1^{pwt} and 9.7-8.9 ppm for SOD1^{barrel}, ensuring no background signal, neither in *E. coli* nor in A2780 cells. The attenuated intensities were subsequently plotted as a function of relaxation delay time and fitted to a single exponential decay. The analysis was performed using in house Matlab (MathWorks, MA, USA) scripts. In the case of two-dimensional data, signal intensities and relaxation rates were extracted using the CcpNmr [9] software and its built-in 'Data Analysis' tool.

S1.7 Determining the protein net charge.

Net charge densities of the various proteins were taken from Mu et al. [2] where they had been determined from the migration speed on a native gel. To obtain the relative net charge, the net charge density was scaled with the surface area of each protein, where the surface area, in turn, was estimated from the hydrodynamic radii of the proteins [2]. To correlate the relative net charge of each protein to a nominal charge, the theoretical charge was estimated from the primary sequence and a linear correlation between the relative net charge and the theoretical charge was found. This linear relationship was used for the assignment of a net charge to each protein variant from the migration on the native gel.

*S1.8 Converting *E. coli* mobility^{in-cell} to η^{app} .*

The quinary interactions in *E. coli* were quantified by the retardation of the rotational tumbling rate, observed as line broadening with subsequent intensity loss [2]. To be able to compare the effects in mammalian cells and *E. coli*, we had to convert this retardation to η^{app} . Similar to the relaxation case, changes in microscopic viscosity affect the protein tumbling rate in a size- and shape-dependent manner [10] [11], and hence a reference curve must be determined for each studied protein. Here, we use water/glycerol mixtures as reference solutions for the viscosity series, as the viscosity

of these mixtures is well understood [12]. We find that in the region 0 - 50 % glycerol, the relative intensity decreased non-linearly with viscosity for all three proteins (Fig. S10). The response upon increasing the viscosity differs however between the proteins, i.e. the larger the protein, the stronger the relative retardation of the tumbling rate (Fig. S10). From these reference curves, any determined relative intensity loss under equi-concentration conditions [2] can be correlated to an apparent viscosity, η^{app} .

*S1.9 Database derived proteome analysis of *E. coli* and human cells.*

Protein sequences of human and *E. coli* cytosolic proteins were collected from the Uniprot database [13]. The human proteins were from the full human genome and the sub-ensemble used for further analysis was from proteins annotated as cytoplasmic (N= 5216). The *E. coli* proteins were selected from the K12 genome with the cytosolic annotation as selection criterium (N= 1073). No further curating was performed. Protein net charge was estimated by simply adding the basic residues (Arg and Lys) and subtracting the acidic residues (Glu and Asp). This relies on the assumption that all basic residues are positively charged, all acidic negatively charged and the histidine residues are on average neutral at intracellular conditions. The estimated net charge under these assumptions has been shown to agree well with PROPKA3 [14] calculations based on the protein structures [15]. Molecular mass was calculated from the sum of the tabulated mass of each amino acid. The distributions of masses and net charges are shown as histograms in Figure 1.

S1.10 Estimation of transiently bound population from relaxation data.

Any encounter complex with a lifetime longer than τ_r would result in a slower D_{rot} . The spectral quality and the similarities of in-cell and *in-vitro* spectra (Fig. 3) indicates that the interactions are short lived on the NMR timescale, and the lifetimes can be estimated to 10 ns – 0.5 ms. The observed transverse relaxation rate (*in-cell* R_2) is, assuming that the binding is transient and in fast exchange on the chemical shift difference timescale, the population weighted average of the relaxation rates of the free state (R_2^{free}) and the distribution of the bound states ($\tilde{R}_2^{\text{bound}}$):

$$R_2^{\text{OBS}} = p_{\text{free}} R_2^{\text{free}} + (1 - p_{\text{free}}) \tilde{R}_2^{\text{bound}} \quad \text{Eq. S1}$$

Consequently, the population of the bound state is given by:

$$p_{\text{bound}} = 1 - \frac{R_2^{\text{OBS}} - \tilde{R}_2^{\text{bound}}}{R_2^{\text{free}} - \tilde{R}_2^{\text{bound}}} \quad \text{Eq. S2}$$

The relaxation rate of the free state is here assumed to correspond to the relaxation rate of the protein in dilute buffer. The relaxation rate of the distribution of the bound states, on the other hand, can be estimated from (i) the approximatively linear relationship between the transverse relaxation rate and the molecular weight in the mass interval relevant for cytosolic proteins [16] [17], and (ii) the distribution of protein molecular weights in the cytosols of human and bacterial cells [13] (Fig. 1).

The apparent mass of a transient complex between a probe protein with molecular weight M^P and a cytosolic protein with mass M_w is simply the sum of the masses: $M^P + M_w$. The relaxation rate of the transient complex is then given by:

$$\tilde{R}_2(M_w) = \alpha_0 + \alpha_1(M^P + M_w) \quad \text{Eq. S3}$$

Where α_0 and α_1 are empirical parameters derived from the experimental relationship between protein mass and transversal relaxation [17]. The molecular mass of a transient complex partner can be any from the distribution of masses of the cytosolic proteins (Fig 1). The mass distribution can relatively well be described as a Γ -distribution (Fig. 1) [18] [19], with the probability density function:

$$P(M_w) = a \cdot b^k \frac{1}{\Gamma(k)} M_w^{k-1} e^{-b \cdot M_w} \quad \text{Eq. S4}$$

Where k is a shape parameter, b is a scale parameter and a is here a normalization constant.

Assuming that the probe protein encounters any cytosolic protein with the same frequency, or in other words with equal probability, and further assuming – for simplicity – that the probe protein has only one binding partner at any given time point: by combining Eqs. S2-S4 we can estimate the fraction of probe protein that is bound to a surrounding protein at any given time point.

$$p_{\text{bound}} = 1 - \int_0^{\infty} \frac{R_2^{\text{OBS}} - \bar{R}_2(M_w)}{R_2^{\text{free}} - \bar{R}_2(M_w)} P(M_w) dM_w \quad \text{Eq. S5}$$

Equation 5 is thus a function of the observed intracellular relaxation rate (R_2^{OBS}), the relaxation rate in dilute buffer (R_2^{free}) and the molecular weight of the probe protein (M^p). This analysis shows that although TTHA^{pwt} and SOD1^{barrel} exhibit similar η^{app} in mammalian cells, the magnitude of transient interactions differs significantly (*c.f. main text*). These results are thus in apparent conflict with the findings from the η^{app} analysis, and this disagreement stems from the relative change in mass upon transient binding being larger for HAH1^{pwt} than for SOD1^{barrel}, i.e. although the bound population is similar, the relative change in relaxation rate is then larger for HAH1^{pwt} than for the SOD1^{barrel}, resulting in a larger apparent viscosity increase. The effective mass of the binding partner in human cells is approximately 34.5 kDa (Fig 1), which for HAH1^{pwt} ($M_w = 7.4$ kDa) gives $(34.5+7.4)/7.4 = 5.7$ in relative mass change between the ensemble of bound states and free HAH1^{pwt}. For SOD1^{barrel} the corresponding relative mass change is 4.1. The difference in relative mass change between the proteins ($5.7/4.1 = 1.4$) fully accounts for the difference in relative change in relaxation rates. For HAH1^{pwt} $R_2^{\text{buffer}}/R_2^{\text{in-cell}}$ is $4.9/21.1 = 4.3$ and for SOD1^{barrel} 3.0, which gives the same ratio $4.3/3.0 = 1.4$. Hence, the apparent viscosity describes the proteins' response to the cell interior phenomenologically. However, to obtain a semi-quantitative measure on the transient interactions with the cytosolic proteome, it's size distribution needs to be taken into account.

S2. Supporting Controls

S2.1 Control of cell internalisation and leakage.

To confirm that the analysed protein was located inside the cells during the experiments, two control samples were prepared and run after each acquisition of *in-cell* data: (i) the interstitial fluid in the cell sample was controlled for leaked protein, named the *supernatant sample*, and (ii) the internalised protein was released from the cells by destruction of the cell integrity to yield a *lysate sample*.

The supernatant sample was prepared directly after the *in-cell* data acquisition by first removing the medium located above the packed cells and then carefully resuspending them in the remaining medium. The cell suspension was transferred to a 1.5 ml Eppendorf tube and gently centrifuged (300 g, 5 min, 25 °C) using a table-top centrifuge. The supernatant was carefully transferred into a NMR tube and a 1D-¹⁵N-SOFAST-HMQC was recorded to detect the presence of leaked isotope-labelled protein. Typically, the signal intensity was less than 10 % of the *in-cell* signal (Fig. S3).

The lysate sample was prepared by adding 200 µl fresh OptiMEM to the cell pellet from the preparation of the supernatant sample. The cells were lysed by sonication: 5 rounds of 1 s pulse/1 s pause cycles were executed with a 10 s cycle duration. Cell debris was removed by centrifugation (17 000 g, 10 min, 4 °C), and the supernatant was transferred to an NMR tube and an 1D-¹⁵N-SOFAST-HMQC was recorded (Fig. S1).

S2.2. Estimating the protein concentrations in the *in-cell* samples.

Varying protein concentrations were added to $\sim 10^7$ cells/ml cell suspensions in 500 µL samples. Protein concentrations of 0, 1, 4, 10, 20, 30 & 40 µM were used for HAH1^{pwt}, and concentrations of 0, 2, 4, 6, 8, 10 & 20 µM were used for TTHA^{pwt} and SOD1^{barrel}. For each sample a 1D-¹⁵N-SOFAST-HMQC spectrum was recorded. Reference curves were constructed by integrating over a signal region of 9.5-7.1 ppm (HAH1^{pwt}, SOD1^{barrel}) or 9.3-7.1 ppm (TTHA^{pwt}) in the acquired NMR spectra. To account for background cell signals we used relative intensities per million cells (I_{rel}) determined by taking the difference between x µM and 0 µM protein according to Eq. S6, where the subscripts x and 0 refer to protein concentrations.

$$I_{rel} = I_x - I_0 = \left[\frac{I_{abs}}{50} \right]_x - \left[\frac{I_{abs}}{50} \right]_0 \quad \text{Eq. S6}$$

I_{rel} was plotted against the protein concentration and a linear curve was fitted to the data.

For each *in-cell* spectrum, the integral of the signal was taken over the same interval as for the reference curves, and the apparent protein concentration was determined. We found that the average *in-cell* concentrations of the three proteins were; $11.2 \pm 5.6 \mu\text{M}$ for HAH1^{pwt}, $13.6 \pm 5.6 \mu\text{M}$ for SOD1^{barrel} and $17.1 \pm 7.2 \mu\text{M}$ for TTHA^{pwt} (Fig. S2).

S2.3 Determination of cell volume fraction in the detection volume.

The cell volume fraction was quantified by comparing the volume occupied by all cells to that of the entire detection volume. The assumption was that all cells were in the detection volume. The diameter and therewith the detection volume, vary depending on whether a 4 mm or 5 mm NMR tube is used. The average cell volumes in the literature was used for the calculations ($2500 \mu\text{m}^3$) [20] [21]. This was multiplied with the number of cells in each sample to get the total cell volume. The average occupancy of the detection volume varied slightly for the three proteins; 0.46 ± 0.07 for HAH1^{pwt}, 0.44 ± 0.10 for SOD1^{barrel} and 0.53 ± 0.21 for TTHA^{pwt}, and this yields 0.48 ± 0.13 over all samples.

S2.4 Chemical shift perturbations in the mammalian cytosol.

The SOFAST-HMQC spectra do not only report on line width but also on chemical shifts, which in turn report on the electrochemical environment the spins reside in. Comparing the chemical shifts in cells with *in-vitro* data could then provide information on changes in structure, dynamics, interactions and local environment. However, in order to separate specific localised effects from more trivial ionic strength, pH and viscosity effects a ‘good’ reference spectrum is required. Here, a good reference spectrum is one where the pH and ionic properties mimic the *in-cell* environment. We have previously determined the *in-cell* pH in A2780 cells to be approximately 6.5 [22], and we confirmed this by comparing the chemical shifts from *in-cell* spectra with *in-vitro* spectra (Fig. S5). It is noteworthy that the ionic

environment in cells is very different from the typical ‘physiological buffer’ [15], which will also have consequences for accurate pH determination. Despite this, we find that the chemical shifts in mammalian cells can be remarkably well reproduced in an isotonic PBS buffer at pH 6.5 (*c.f. below*). Yet, some small but significant chemical shift differences can still be detected (Fig. 3 and Fig. S6). The localisation of the chemical shift differences for TTHA^{pwt} and HAH1^{pwt} show no clear structural clustering. In the case of SOD1^{barrel}, on the other hand, the induced chemical shift changes from the *in-cell* environment are mainly located on the sheet normally harbouring the catalytically active loops. This region coincides both with the localisation of induced chemical shifts due to histidine sidechain protonation/deprotonation, as well as the most dynamic region of SOD1^{barrel} [23] [4]. The observed chemical shift changes may report on dynamical effects, altered pK_a properties of histidine protonation in the more complex ionic environment of the cell interior and – perhaps as a result of that – more prominent transient interactions in certain positions. Although there are some small sequence variations in both line broadening and induced chemical shifts (Fig. S4 and S6), we conclude that the major effect on the NMR spectrum upon transfer of the proteins into the cells is a general line broadening due to a global rotational retardation.

S2.5. Confirmation of *in-cell* pH.

The effective pH found in the interior of the A2780 mammalian cell line was determined by comparing the *in-cell* 2D ¹H¹⁵N-SOFAST-HMQC spectrum of HAH1^{pwt} to a series of *in-vitro* 2D ¹H¹⁵N-HSQC spectra acquired for a range of pHs between 5.5 and 7.5. From those the spectrum at pH 6.5 is overlapping best with the *in-cell* spectrum. However, in some parts of the protein chain – particularly in the C-terminal region – several resonances are shifted from those in the reference spectrum at pH 6.5. Adding 150 mM NaCl to the protein sample eradicates most of these shift differences (Fig. S5), allowing us to draw the conclusion that the initial shifts arise from the higher ionic strength found in the interior of cells and confirming that the effective pH indeed is around 6.5. *In-vitro* reference SOFAST-HMQCs were recorded – with and without 150 mM NaCl – for all three proteins (Fig. S5), again confirming that the cytosol’s effective pH is approximately 6.5. Hence, for the sake of comparability and to eliminate any contributions from protonation effects to

relaxation rates and line broadening in the viscosity standard curves, all *in-vitro* reference spectra were henceforth performed at pH \sim 6.5.

S2.6 Linear time dependent correction yields similar R_2 values to those acquired in 'interleaved' mode.

Initially, the relaxation data was acquired by running three experiments with varying relaxation delays sequentially. However, this resulted in somewhat scattered decay profiles (Fig. S7). On a closer look, it was apparent that the deviation from the single exponential curve showed a time-dependent trend, most likely originating from the cell pellet slowly compacting during the course of the measurement. Comparing the 1D-SOFAST-HMQC spectra before and after the *in-cell* relaxation measurements confirmed this observation, as the signal intensities increased with time. To account for this time-dependent change in signal amplitude, we introduced a linear correction factor C_F , determined from the signal intensity of the two control experiments immediately before (A_{before}) and after (A_{after}) the relaxation measurements, separated by the time required for acquiring the T_2 data (Δt): $C_F(t) = 1 - \frac{A_{\text{after}}}{A_{\text{before}}\Delta t} \cdot t$. The amplitude of each relaxation experiment recorded at time t_i after the *before*-experiment was normalised by a factor $C_F(t_i)$. Using this correction indeed reduced the scatter (Fig. S7). With this in mind, a new R_2 data set was acquired. This time the three experiments were run with various delays in parallel and not as sequential blocks, i.e. interleaved acquisition. The spectra and signal intensities obtained in this way are an average over the total acquisition time and any time-dependent trends in the signal decay curves are consequently averaged out. Despite the very crude assumption that these trends behave linearly with recording time, the obtained relaxation rates after the correction gave fairly similar results to those from the interleaved experiments with $14.1 \pm 3.9 \text{ s}^{-1}$ with linear correction and $12.9 \pm 1.4 \text{ s}^{-1}$ with interleaved acquisition for TTHA^{pwT}, $19.6 \pm 5.9 \text{ s}^{-1}$ and $21.1 \pm 1.4 \text{ s}^{-1}$ for HAH1^{pwT} and $23.6 \pm 5.9 \text{ s}^{-1}$ and $25.0 \pm 1.5 \text{ s}^{-1}$ for SOD1^{barrel}. Nonetheless, the interleaved acquisition resulted in better data and a more accurate determination of R_2 .

S2.7 Comparison of 1D and 2D R_2 measurements.

Due to the limited life span of mammalian cells in the NMR tube, it is paramount for data acquisition to be as short as possible. However, due to the low signal of *in-cell* samples, a large number of scans needs to be recorded. In addition to that, techniques exploiting relaxation-enhancing pulse elements to reduce measurement time, such as SOFAST [24] or BEST [25] – are not applicable for single coherence relaxation measurements. Hence, to be able to acquire the necessary data within a reasonable amount of time, we chose to reduce the dimensionality of the conventional HSQC-based ^{15}N -transverse relaxation experiment. To verify that relaxation rates obtained from this projection are the same as those obtained from the conventional 2D-experiment, we split the ^1H -spectrum into a number of sections equal to the number of peaks in the corresponding two-dimensional spectrum of the respective protein and compared the obtained relaxation rate distributions between 1D and 2D data (Fig. S8). The results show clearly that the obtained relaxation rates differ only marginally from each other. As a next step, we tested if taking a portion of the spectrum for calculating relaxation rates was representative for the whole protein. This is crucial, as naturally abundant ^{15}N -amide protons in the context of the extremely high protein concentration inside cells cause background signals too high to be neglected. To avoid the regions of high background signal, we chose 9.4-8.6 ppm in the case of TTHA^{pwt} and HAH1^{pwt} and 9.7-8.9 ppm in the case of SOD1^{barrel}. Again, sections in this spectral portion were split into the number of peaks in the corresponding two-dimensional spectrum and compared to each other in form of histograms (Fig. S8). Although, the average values of the thus obtained rates deviate more from the values obtained from the 2D spectrum than if the whole spectrum was used, the values are all within one standard deviation.

S2.8 Measuring R_2 without refocusing CPMG pulse trains.

Quinary interactions can affect the transverse relaxation rate, R_2 , two-fold: On the one hand the interaction with other proteins will slow down the tumbling rate and therefore increase the rotational correlation time, leading to elevated relaxation rates [11]. On the other hand, transient interactions change the local chemical environment at the interaction site, resulting in a change of Larmor frequency in the affected spins. However, for this to noticeably manifest in an increased relaxation rate, the exchange time of the interaction needs to be in the order of the frequency difference, typically

in the μs - ms -time regime [26] [27]. To reduce contributions of these μs - ms -exchange processes, a high-frequency Carr-Purcell-Meiboom-Gill (CPMG) pulse train is run during the incrementally increased relaxation delay times [28, 29]. Any exchange processes with a rate between $100\text{-}3000\text{ s}^{-1}$ will therefore be refocused in a typical ^{15}N - R_2 -experiment.

The uniformity of the cross peak line-broadening and the unperturbed chemical shifts in the *in-cell* 2D-SOFAST-HMQC spectra, suggest that the increase in relaxation rate is less due to chemical exchange in the μs - ms time range and rather due to a retarded D_{rot} . To put this to the test, we performed a set of modified R_2 measurements where the CPMG pulse train was replaced by a single refocusing pulse (to refocus phase shifts due to dipolar coupling and chemical shift evolution). In this way, the chemical exchange processes are allowed to take full effect on the relaxation rate. Since CPMG pulse trains can cause sample heating, a compensatory CPMG block was placed in the very beginning of the pulse sequence. This compensatory block was adjusted for each protein to be of the exact same length as the maximum delay of the corresponding conventional relaxation experiment. The results from these experiments are summarized in Figure S11. The relaxation rates of all three proteins are unchanged upon removal of the CPMG pulse train during the relaxation measurements, confirming that chemical exchange processes in the μs - ms time range seem to play only a minor role in modulating the *in-cell* R_2 values.

S2.9 Double mutations in the HAH1-MNK1 recognition interface results in larger effects on η^{app} than corresponding mutations outside the interface.

Introducing a charged carboxyl group in the HAH1-MNK interface by single point mutations reduces the η^{app} more than expected from an introduction of a carboxyl group anywhere else on the protein surface. This pattern continues when pairs of positive side chains on HAH1^{PWT} are replaced by negative carboxylates: all double mutations involving one residue in the binding cluster show a large effect ($\text{mobility}^{\text{in-cell}} = 0.55 \pm 0.06$), while double mutations involving two residues in the cluster yield an even larger increase in mobility ($\text{mobility}^{\text{in-cell}} = 0.66 \pm 0.01$). On the other hand, double mutations only involving K and/or R residues outside of the binding cluster show a significantly smaller increase in mobility ($\text{mobility}^{\text{in-cell}} = 0.30 \pm 0.05$),

compared to those in the charge cluster. All three positively charged residues involved in the MNK1 recognition site [30] show very high response upon mutation, although the native target protein is absent in *E. coli* (Fig. S12). This indicates that, on top of the overall global surface properties, specific regions may have evolved to facilitate transient non-specific interactions, presenting the specificity site to the encounter partner.

Supporting References

1. Sakakibara, D., Sasaki, A., Ikeya, T., Hamatsu, J., Hanashima, T., Mishima, M., Yoshimasu, M., Hayashi, N., Mikawa, T., Walchli, M., et al. (2009). Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* *458*, 102-105.
2. Mu, X., Choi, S., Lang, L., Mowray, D., Dokholyan, N.V., Danielsson, J., and Oliveberg, M. (2017). Physicochemical code for quinary protein interactions in *Escherichia coli*. *Proc Natl Acad Sci U S A* *114*, E4556-E4563.
3. Anastassopoulou, I., Banci, L., Bertini, I., Cantini, F., Katsari, E., and Rosato, A. (2004). Solution structure of the apo and copper(I)-loaded human metallochaperone HAH1. *Biochemistry* *43*, 13046-13053.
4. Danielsson, J., Kurnik, M., Lang, L., and Oliveberg, M. (2011). Cutting off functional loops from homodimeric enzyme superoxide dismutase 1 (SOD1) leaves monomeric beta-barrels. *J Biol Chem* *286*, 33070-33083.
5. Mu, X., Choi, S., Lang, L., Mowray, D., Dokholyan, N.V., Danielsson, J., and Oliveberg, M. (2017). Physicochemical code for quinary protein interactions in *Escherichia coli*. *Proceedings of the National Academy of Sciences* *114*, 4556-4563.
6. Danielsson, J., Kurnik, M., Lang, L., and Oliveberg, M. (2011). Cutting off functional loops from homodimeric enzyme superoxide dismutase 1 (SOD1) leaves monomeric beta-barrels. *Journal of Biological Chemistry* *286*, 33070-33083.
7. 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.
8. Lee, W., Tonelli, M., and Markley, J.L. (2015). NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* *31*, 1325-1327.
9. Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* *59*, 687-696.
10. Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik* *322*, 549-560.
11. Kowalewski, J., and Maler, L. (2017). *Nuclear Spin Relaxation in Liquids: Theory, Experiments, and Applications*, Second Edition, (CRC Press).
12. Cheng, N.-S. (2008). Formula for the Viscosity of a Glycerol–Water Mixture. *Industrial and Engineering Chemistry Research* *47*, 3285-3288.
13. Consortium, T.U. (2018). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* *47*, D506-D515.
14. Olsson, M.H., Sondergaard, C.R., Rostkowski, M., and Jensen, J.H. (2011). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput* *7*, 525-537.
15. Wennerstrom, H., Vallina Estrada, E., Danielsson, J., and Oliveberg, M. (2020). Colloidal stability of the living cell. In *To be published*.

16. Farrow, N.A., Zhang, O., Szabo, A., Torchia, D.A., and Kay, L.E. (1995). Spectral density function mapping using ¹⁵N relaxation data exclusively. *J Biomol NMR* *6*, 153-162.
17. Rossi, P., Swapna, G.V., Huang, Y.J., Aramini, J.M., Anklin, C., Conover, K., Hamilton, K., Xiao, R., Acton, T.B., Ertekin, A., et al. (2010). A microscale protein NMR sample screening pipeline. *J Biomol NMR* *46*, 11-22.
18. Tiessen, A., Perez-Rodriguez, P., and Delaye-Arredondo, L.J. (2012). 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* *5*, 85.
19. White, S.H., and Jacobs, R.E. (1993). The evolution of proteins from random amino acid sequences. I. Evidence from the lengthwise distribution of amino acids in modern protein sequences. *J Mol Evol* *36*, 79-95.
20. Luby-Phelps, K. (2000). Cytoarchitecture and physical properties of cytoplasm: volume, viscosity, diffusion, intracellular surface area. *International review of cytology* *192*, 189-221.
21. Zhang, Z., Krylov, S., Arriaga, E.A., Polakowski, R., and Dovichi, N.J. (2000). One-dimensional protein analysis of an HT29 human colon adenocarcinoma cell. *Analytical chemistry* *72*, 318-322.
22. Danielsson, J., Mu, X., Lang, L., Wang, H., Binolfi, A., Theillet, F.X., Bekei, B., Logan, D.T., Selenko, P., Wennerstrom, H., et al. (2015). Thermodynamics of protein destabilization in live cells. *Proc Natl Acad Sci U S A* *112*, 12402-12407.
23. Danielsson, J., Awad, W., Saraboji, K., Kurnik, M., Lang, L., Leinartaite, L., Marklund, S.L., Logan, D.T., and Oliveberg, M. (2013). Global structural motions from the strain of a single hydrogen bond. *Proc Natl Acad Sci U S A* *110*, 3829-3834.
24. Schanda, P., Kupce, E., and Brutscher, B. (2005). SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR* *33*, 199-211.
25. Schanda, P., Van Melckebeke, H., and Brutscher, B. (2006). Speeding up three-dimensional protein NMR experiments to a few minutes. *J Am Chem Soc* *128*, 9042-9043.
26. Kleckner, I.R., and Foster, M.P. (2011). An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta* *1814*, 942-968.
27. Korzhnev, D.M., and Kay, L.E. (2008). Probing invisible, low-populated States of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc Chem Res* *41*, 442-451.
28. Carr, H.Y., and Purcell, E.M. (1954). Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *American Physical Society* *94*, 630-638.
29. Meiboom, S., and Gill, D. (1958). Modified Spin - Echo Method for Measuring Nuclear Relaxation, Volume 29.
30. Banci, L., Bertini, I., Calderone, V., Della-Malva, N., Felli, I.C., Neri, S., Pavelkova, A., and Rosato, A. (2009). Copper(I)-mediated protein-protein interactions result from suboptimal interaction surfaces. *Biochem J* *422*, 37-42.

Supporting figures and legends

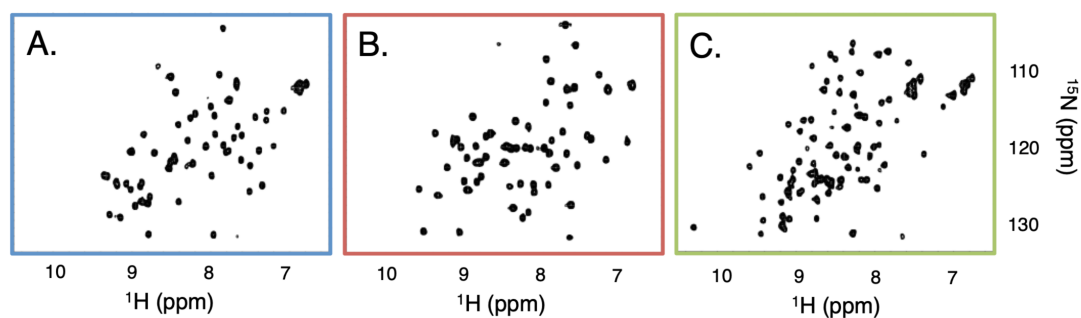


Figure S1. SOFAST-HMQC spectra of TTHA^{pwt} (A), HAH1^{pwt} (B) and SOD1^{barrel} (C) in the dilute lysate sample after *in-cell* NMR acquisition. In all three cases the spectra show well resolved signals with narrow line-widths, reassuring that the line broadening effect seen in the *in-cell* NMR spectra (Fig. 2) is due to transient reversible interactions with the cellular environment. No indications on any permanent changes on the proteins can be found.

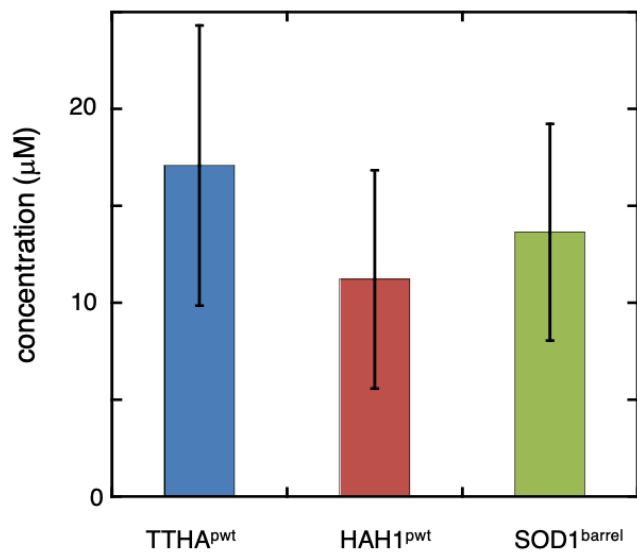


Figure S2. Protein concentrations in the *In-cell* NMR samples were estimated by comparing the signal intensity of the *in-cell* 1D-SOFAST-HMQC spectrum of each protein to standard curves. The concentration reported here is the total concentration in the NMR sample, and the actual *in-cell* concentration is approximately a factor of 2 higher as the effective cell volume packing is close to 50%.

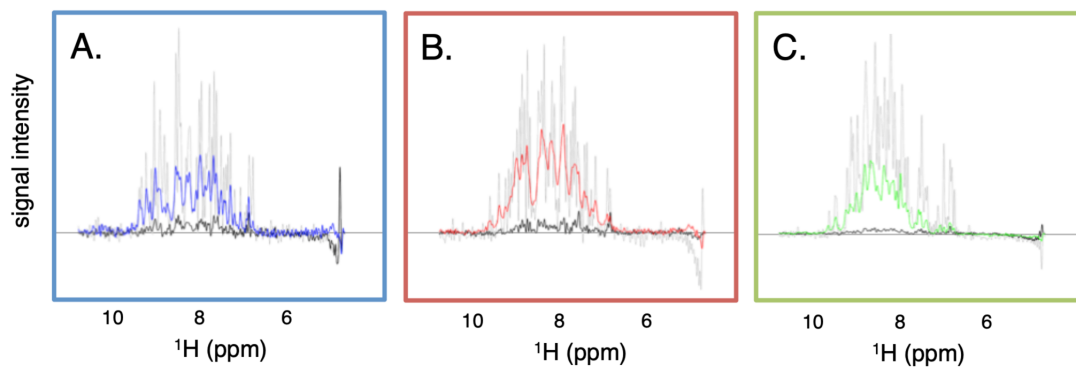


Figure S3. Control of leakage of cytosolic protein to the interstitial fluid during the *in-cell* NMR acquisition. Each panel depicts the *in-cell* (coloured line), the supernatant (black line) and lysate (grey line) 1D-SOFAST-HMQC spectra of TTHA^{pwt} (A), HAH1^{pwt} (B) and SOD1^{barrel} (C). The low supernatant signal shows that the amount of leakage is very small. Since producing the supernatant sample induces some leakage, the actual leakage is even less than shown in the figures.

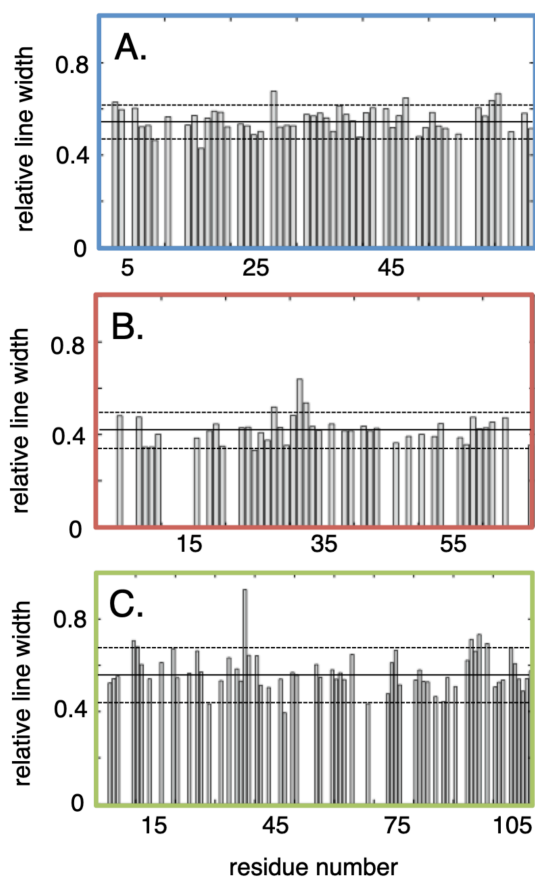


Figure S4. Relative line broadening ($\text{line-width}^{\text{buffer}} / \text{line-width}^{\text{in-cell}}$) per residue for the three probe proteins, TTHA^{PWT} (A), HAH1^{PWT} (B) and SOD1^{barrel} (C). Data is shown for residues with non-overlapping cross-peaks with signal strength good enough for quantification. All three proteins show uniform line broadening, without any obvious topological hotspots, suggesting that the main contributor to line broadening is a global retardation of the rotational tumbling rate. The average relative broadening is indicated as a solid line and the dashed lines correspond to \pm one standard deviation.

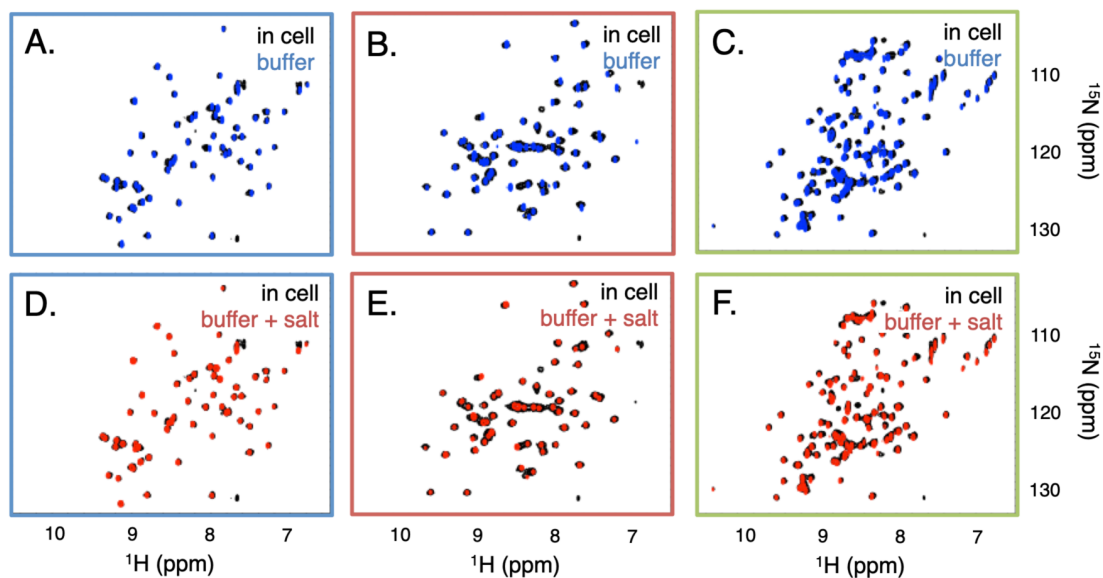


Figure S5. The effects of salt on the *in-vitro* reference spectra. A comparison of *in-cell* spectra (black cross-peaks) of TTHA^{pwt} (A and D), HAH1^{pwt} (B and E) and SOD1^{barrel} (C and F) with *in-vitro* references in weak MES-buffer pH 6.5 (blue cross peaks in A-C) shows significant chemical shift differences of a subset of cross peaks. However, addition of 150 mM NaCl (red cross peaks in D-F) results in spectra with significantly smaller chemical shift differences.

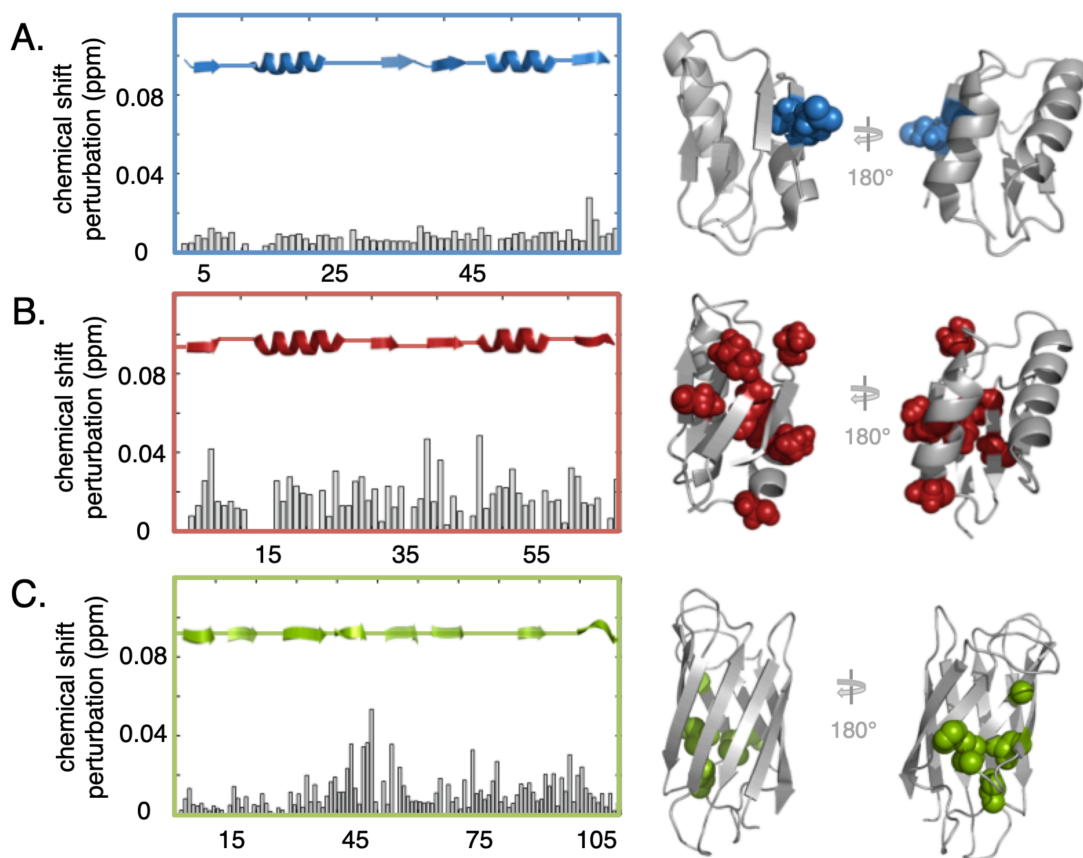


Figure S6. The combined ^1H and ^{15}N chemical shift perturbation upon transfer of TTHA^{pwt} (A), HAH1^{pwt} (B) and SOD1^{barrel} (C) into live human cells. The chemical shifts are compared to *in-vitro* spectra in isotonic PBS. For all proteins the most affected residues are projected onto the structures in the right panel. The secondary structure elements of the proteins are shown as cartoon inset in the panels. TTHA^{pwt} and HAH1^{pwt} show no obvious clustering of the induced effects, while SOD1^{barrel} shows indications of clustering of induced shifts on strands 4, 5 and 7. This region both corresponds to the cluster of titratable histidine side chains, but also to the dynamic region of SOD1^{barrel} [23].

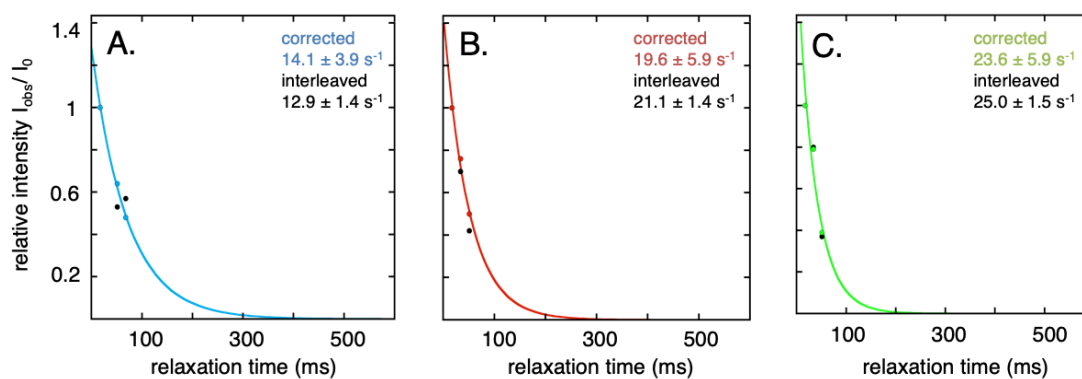


Figure S7. Time-corrected signal decay curves for transverse relaxation rate determination for TTHA^{PWT} (A), HAH1^{PWT} (B) and SOD1^{barrel} (C). Before running the relaxation measurements in ‘interleaved mode’, the individual data points were collected sequentially allowing gradual time-dependent processes to influence the signal intensities (black markers). Correcting linearly for this time-dependence (coloured markers) resulted in decay profiles that much more resembled the exponential signal loss expected for relaxation processes. The obtained relaxation rates were very close to those obtained from the measurements acquired in the ‘interleaved mode’ confirming the overall reproducibility.

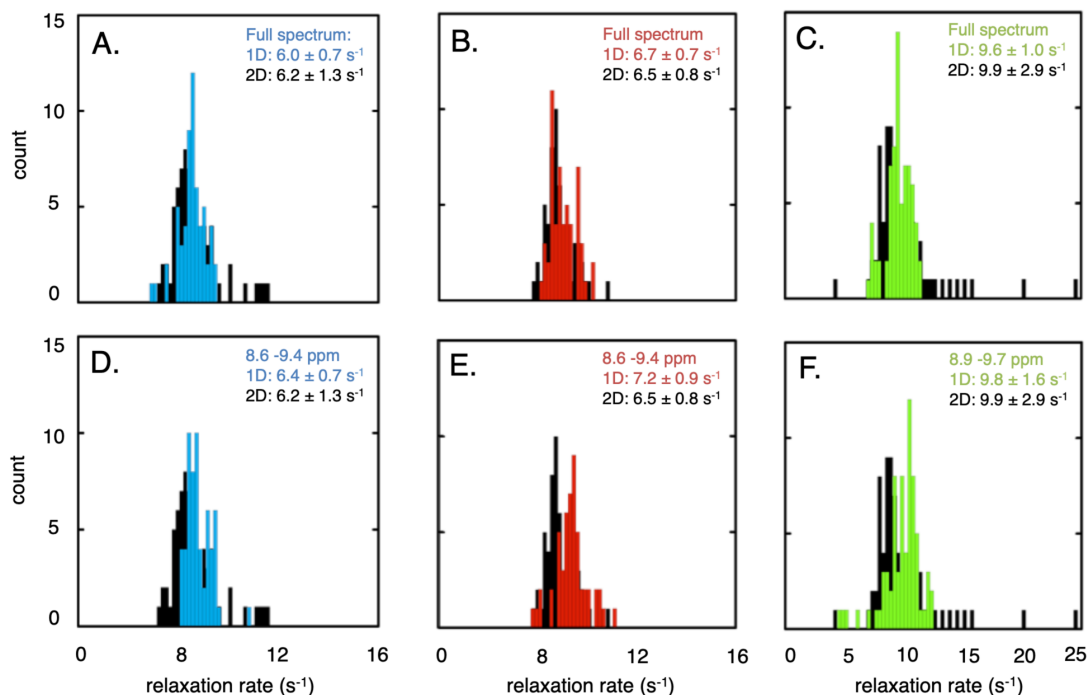


Figure S8. Relaxation rate distributions of 1D and 2D data in comparison. The transverse relaxation rates obtained from each peak in the 2D spectrum are depicted in black, while those obtained from 1D data sets are coloured: blue for TTHA^{PWT} (A and D), red for HAH1^{PWT} (B and E) and green for SOD1^{barrel} (C and F). A-C shows the distribution of rates obtained from the total signal of the 1D data set, while D-F depicts those obtained from a spectral sub-portion (0.8 ppm starting from the most downfield shifted signals). Both datasets show that the 1D and 2D data sets give similar results with the differences being within the error of the measurements and that a relaxation rate determined from this fraction of the spectrum is indeed representative for the whole protein.

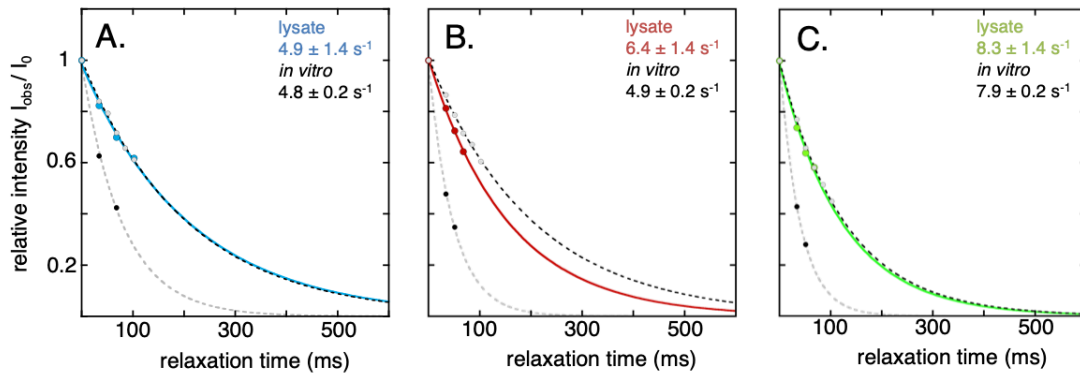


Figure S9. Transverse relaxation rates of TTHA^{pwt} (A), HAH1^{pwt} (B) and SOD1^{barrel} (C) in cells (black markers, dashed line), in buffer (open markers, dashed line) and in lysate (coloured marker and coloured line). For TTHA^{pwt} and SOD1^{barrel} the lysate signal is very similar to *in-vitro* references in 10 mM MES buffer, but for HAH1^{pwt} we also find a significant retardation of the rotational tumbling rate in the dilute lysate. This indicates that HAH1^{pwt} takes part in transient interactions with the soluble proteins even in the lysate, while the more inert TTHA^{pwt} and SOD1^{barrel} are relatively unaffected by the co-solutes in the lysate.

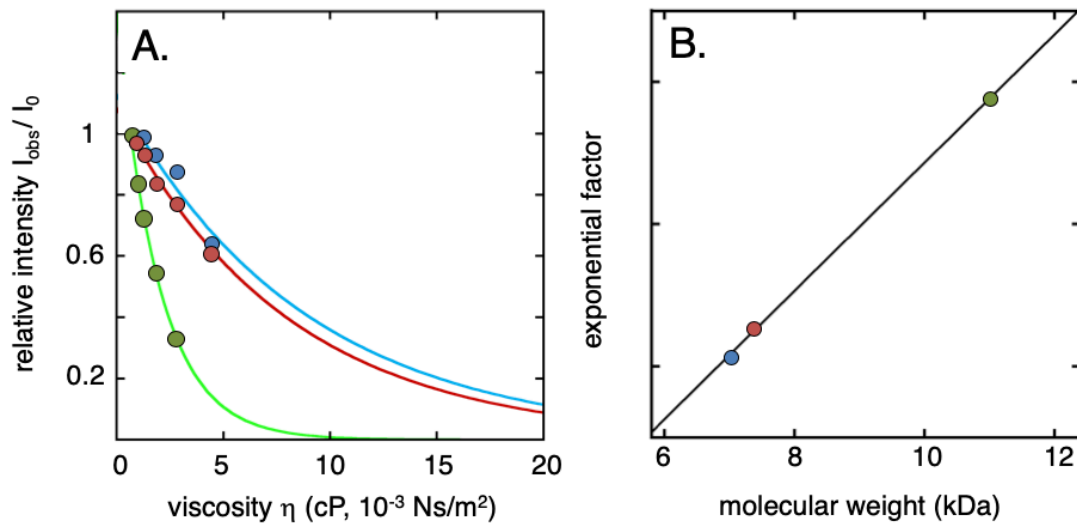


Figure S10. Standard curves for determination of apparent viscosity from relative signal intensity reduction. A: Reference curves regarding the change in intensity ratios upon increasing microscopic viscosity for all three probe proteins; the reference intensity (I_0) is the intensity in pure buffer, assumed to have water viscosity. The exponential factor of the dependence is directly linked to the size of the protein, as highlighted in the correlation plot in B.

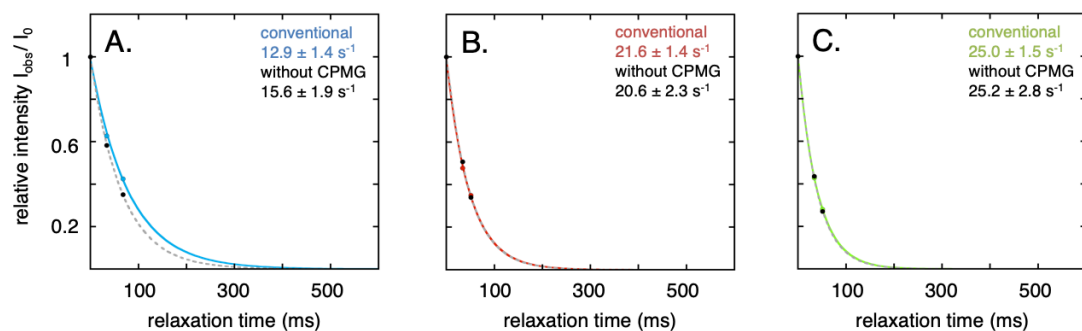


Figure S11. Transverse relaxation rates of TTHA^{pwt} (A), HAH1^{pwt} (B) and SOD1^{barrel} (C) in the absence of the refocusing CPMG pulse train (coloured lines) and with refocusing (grey dashed line). Even when chemical exchange processes in the μ s-ms time range are allowed to contribute to the signal decay, they do not have a significant impact on the R_2 values. Hence, the increased relaxation rates observed inside cells seem to stem mainly from the modulation of the proteins' rotational correlation time.

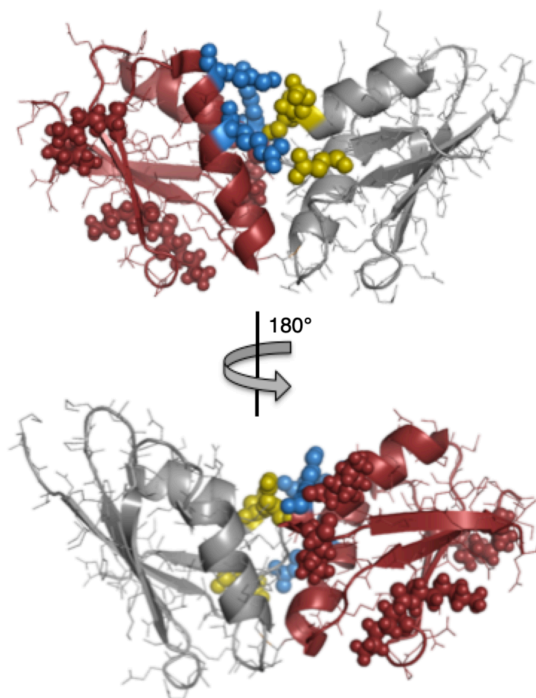


Figure S12. Structure of the native complex between HAH1 (red) and its native binding partner MNK1 (grey) (PDB-ID: 2K1R), highlighting the electrostatic binding interface (blue (positive side chains) and yellow spheres (negative side chains)) separated from the active copper binding site. Replacing any of the positive charges on HAH1^{pwt} involved in the MNK1 binding results in larger than expected effects on rotational freedom [2], suggesting that this recognition site also modulates the transient explorative quinary interactions with the cell surroundings.