Physicochemical code for quinary protein interactions in *Escherichia coli*

Xin Mu[a], Seongil Choi[a], Lisa Lang[a], David Mowray[b], Nikolay V. Dokholyan[b], Jens Danielsson[a,1], and Mikael Oliveberg[a,1]

[a]Department of Biochemistry and Biophysics, Arrhenius Laboratories of Natural Sciences, Stockholm University, S-106 91 Stockholm, Sweden; and [b]Department of Biochemistry and Biophysics, University of North Carolina School of Medicine, Chapel Hill, NC 27599

How proteins sense and navigate the cellular interior to find their functional partners remains poorly understood. An intriguing aspect of this search is that it relies on diffusive encounters with the crowded cellular background, made up of protein surfaces that are largely nonconserved. The question is then if/how this protein search is amenable to selection and biological control. To shed light on this issue, we examined the motions of three evolutionary divergent proteins in the *Escherichia coli* cytoplasm by in-cell NMR. The results show that the diffusive in-cell motions, after all, follow simplistic physical–chemical rules: The proteins reveal a common dependence on (*i*) net charge density, (*ii*) surface hydrophobicity, and (*iii*) the electric dipole moment. The bacterial protein is here biased to move relatively freely in the bacterial interior, whereas the human counterparts more easily stick. Even so, the in-cell motions respond predictably to surface mutation, allowing us to tune and intermix the protein's behavior at will. The findings show how evolution can swiftly optimize the diffuse background of protein encounter complexes by just single-point mutations, and provide a rational framework for adjusting the cytoplasmic motions of individual proteins, e.g., for rescuing poor in-cell NMR signals and for optimizing protein therapeutics.

in-cell NMR | protein surface properties | intracellular diffusion

Despite considerable progress in mapping out how proteins interact functionally through structure and evolved interfaces (1–3), there is yet little known about how proteins interact nonspecifically upon random diffusive encounters (4–10). Although these nonspecific "quinary" (11) interactions are typically weak and short-lived, they are still expected to affect function because of their sheer numbers: Under crowded cellular conditions, they compete with specific binding (6–8), control diffusion (12), and skew structural stability (5, 13–19). The question is then to what extent this dynamic background of nonspecific interactions is biologically controlled and optimized. Part of the answer is hinted by the tendency of soluble proteins, nucleic acids, and membranes to carry a repulsive net-negative charge (20, 21). Such basal level of repulsion between the cellular components is further indicated by a matching surplus of positive counter ions like K^+, Mg^{2+}, Ca^{2+}, and Zn^{2+}, which greatly exceeds the concentration of negative dittos HCO_3^-, Cl^-, SO_4^{2-}, and PO_4^{2-} (22). However, proteins expose also positive, polar, and hydrophobic moieties that operate against the net-negative charge repulsion by engaging in attractive interactions upon diffusive encounters. The strength and duration of these attractive interactions depend on the proteins' detailed surface composition, relative orientations, and ability to adapt complementary shapes. Following Elcock's estimate for the *Escherichia coli* cytoplasm, each protein experiences at all times approximately five putative interaction partners in its immediate cellular environment (8). Sometimes, mutual fits enable strong functional binding (1, 2), but, most often, the proteins just separate after a brief tête-à-tête (3), in search of higher-affinity partners. A key detail is that the effect of this quinary interplay on protein stability varies with protein identity and type of host cell (5, 9, 13–19, 23–26). The "crowding" effect is thus not limited to steric exclusion (27) but has also a decisive dependence on sequence composition and the

details of the quinary interactions. In this study, we examine to what extent these ubiquitous background interactions are biologically tunable and can be accounted for (Fig. 1). The results show that the quinary interactions are critically sensitive to surface mutation and that proteins from divergent organisms respond very differently to the *E. coli* cytoplasm: Although the bacterial protein moves relatively freely in the bacterial interior, the human homolog tends to stick. Even so, the proteins can readily be tuned to any desired in-cell mobility by just a few structurally benign surface mutations, following a universal dependence on their macroscopic surface properties. As such, our findings present the physical–chemical code for quinary interactions in the *E. coli* cytoplasm with numerous implications for functional protein design and deciphering the functional evolution of proteomes.

Experimental Protocol

Proteins. To assure that the proteins mainly report on the nonspecific in-cell encounters, possible specific interactions were mitigated by removal of the active sites and binding sites (*SI Materials and Methods*). For bacterial TTHA1718 [Protein Data Bank (PDB) 2ROE] (28) and human HAH1 (PDB 1TL4) (29), this process involved mutational substitution of the metal-binding ligands, and, for SOD1, truncation of the active-site loops that also ruptures the dimer interface (PDB 4BCZ) (30–32). The resulting pseudo-wild-type proteins are denoted $TTHA^{PWT}$, $HAH1^{PWT}$, and $SOD1^{barrel}$, and are listed in Table S1 and Fig. S1 together with 127 analyzed surface mutations.

Experimental Strategy. The proteins (Table S1 and Fig. S1) were produced and analyzed in *E. coli* BL21(DE3)pLysS cells as follows

---

Significance

This study shows that the diffusive motions of proteins in live cells are by no means without control but follow simplistic physical–chemical rules that can be quantified and optimized through surface composition. Most strikingly, human proteins are observed to stick to the "foreign" environment of bacterial cells, whereas the bacterial analogue moves around freely. Even so, the human proteins can predictably be transformed to bacterial behavior with a few structurally benign surface mutations, and, conversely, the bacterial protein can be made to stick. The findings have not only fundamental implications for how protein function is controlled at the physical–chemical level but can also be used to adjust protein motion in *Escherichia coli* at will.
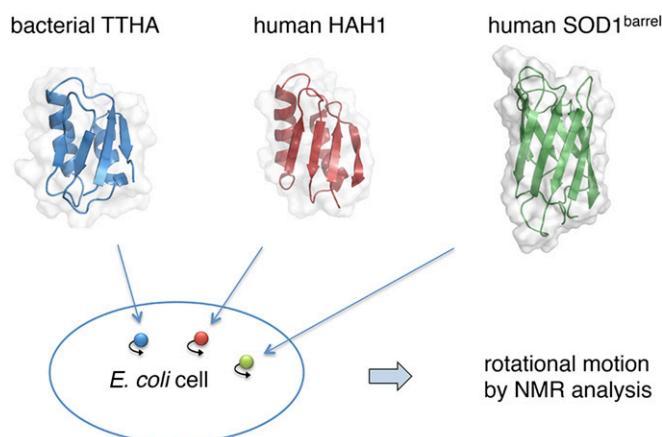

Fig. 1. Measurements of the cytoplasmic motions of three evolutionary divergent proteins by in-cell NMR: bacterial TTHA, human HAH1, and human SOD1[barrel]. Isotope-labeled proteins were generated directly in the *E. coli* cytoplasm by overexpression. The intact cells were then carefully washed and transferred to NMR tubes for detection of internal protein motions. Our observable is the protein's rotation correlation time, i.e., how freely it tumbles around, which is a sensitive measure of the interactions experienced in the crowded *E. coli* interior. In essence, NMR spectra with narrow cross-peaks show that the proteins tumble unrestricted, and broadened-out spectra show that the proteins get stuck. By point mutation, we gradually tune the three proteins between these extremes to map out the physical–chemical code for the cytoplasmic crosstalk (Table S1 and Fig. S1).

(*SI Materials and Methods*). Overexpression was induced by 1-thio-β-D-galactopyranoside (IPTG) in isotope-enriched medium and sustained for 4 h, yielding intracellular concentrations of 1.6 ± 0.8 mM N[15]-labeled protein (*SI Materials and Methods* and Fig. S1). The culture was then nondisruptively centrifuged at $800 \times g$ for 8 min to obtain a pellet of intact, viable cells. One gram of this cell pellet was carefully dissolved in 1 mL of M9 buffer, transferred to an NMR tube, and subjected to in-cell spectral analysis, i.e., sample[in cell]. For control of protein leakage, the sample[in cell] was extracted from the NMR tube and centrifuged at $2,400 \times g$ for 2 min to obtain a cell-free supernatant of the intervening M9 buffer, i.e., sample[supernatant]. The pelleted cells were then resuspended to the original sample volume with M9 buffer, lysed to set free the intracellular target protein, and centrifuged at $17,000 \times g$ for 10 min to remove cell debris and larger macromolecular species. This decrowded lysate, which represents a much more dilute background than the cytoplasm (*SI Controls*, Fig. S1, and Table S2), was finally used as a control for internalized protein, i.e., lysate[supernatant]. An advantage of this protocol is that the amount of labeled protein in the NMR detection volume is kept the same in sample[in cell] and lysate[supernatant], allowing estimates of dynamic changes upon cell rupture from changes in peak heights alone (*Results* and *SI Controls*).

Results

Bacterial and Human Homologs Respond Differently to the *E. coli* Cytoplasm.

For reference, we acquired first the in-cell NMR spectra of the pseudo-wild-type variants of bacterial TTHA (TTHA[PWT]), human HAH1 (HAH1[PWT]), and the human SOD1 barrel (SOD1[barrel]) (Fig. 1). Analysis shows that *E. coli* cells expressing bacterial TTHA[PWT] (sample[in cell]) display a highly dispersed Heteronuclear Multiple Quantum Coherence (HMQC) spectrum with well-resolved cross-peaks, characteristic for a rapidly tumbling and fully folded structure (Fig. 2). Subsequent tests of leakage (sample[supernatant]) reveal typically very low amounts of labeled material in the intervening medium (*SI Controls*, Table S1, and Fig. S2), indicating that the signal indeed stems from protein in the cytoplasm (Fig. 2). Finally, the supernatant of the lysed cells (lysate[supernatant]) yields a spectrum analogous to that in sample[in cell], but with a slight sharpening of the HMQC cross-peaks (Fig. 2).

The lysate[supernatant] spectrum is further indistinguishable from that obtained in pure buffer, showing that the low levels of endogenous *E. coli* proteins remaining in the lysate supernatant (~10 mg/mL; *SI Controls* and Fig. S1) have negligible impact on the analysis: lysate[supernatant] mimics a pure-buffer control. Taken together, these results show that the *E. coli* cytoplasm has limited, but yet measurable, impact on the dynamic behavior of bacterial TTHA[PWT] and that these restrictions ease upon cell lysis. With the human analog HAH1[PWT], however, the situation is distinctly different. The in-cell spectrum of this protein is broadened out to the extent that it completely disappears, using the same data contour level settings as for TTHA[PWT]. What remain visible in sample[in cell] are only the distinct cross-peaks of the background of labeled metabolites that always accumulate during overexpression in labeled growth medium (Fig. 2). The cause of such global line broadening for a globular protein is normally decreased rotational mobility (33–35). Conspicuously, the folded HAH1[PWT] spectrum subsequently restores to high resolution upon cell lysis in lysate[supernatant], indicating that the protein is just reversibly restricted by the cytoplasm and readily sets free upon cell lysis (Fig. 2). We observe the very same behavior for the human SOD1[barrel] (Fig. 2): Although the protein is previously found to move freely in human cells (5, 31), it seems to stick in the *E. coli* cytoplasm.

In-Cell NMR Spectra Can Be Made to Appear and Vanish by Single-Point Mutations.

The question is then, what allows TTHA[PWT] to tumble relatively freely in the *E. coli* interior whereas HAH1[PWT] and SOD1[barrel] seem to get stuck? Because the effect unlikely arises from specific binding to biological partners but rather from diffuse quinary interactions with the molecular background (5, 9), we start by investigating the role of protein charge (21). As illustrated in Fig. 3, the electrostatic potentials of the TTHA[PWT], HAH1[PWT], and SOD1[barrel] surfaces are quite different. First out was the mutation HAH1[K57E], which replaces the positive surface charge K57 of the human protein with the negative counterpart E57 of the bacterial TTHA[PWT]. The mutation has negligible impact on protein folding and stability (*SI Controls* and Fig. S2), but it increases the formal net charge of HAH1 by −2. As a result, the in-cell cross-peaks of HAH1 sharpen up radically and reveal a high-resolution HMQC spectrum of the folded state, analogous to that of TTHA[PWT] (Fig. 2). Cytoplasmic localization was verified by the sample[supernatant], which revealed no trace of leakage (Table S1). Whatever dynamic restrictions are experienced by the pseudo-wild-type protein in the *E. coli* cytoplasm appear to be largely counteracted by the point mutation K57E. Similarly, the mutation R100E induces distinct cross-peak sharpening of SOD1[barrel], albeit that the spectral restoration is not as prominent as observed for HAH1[K57E] (Fig. 2). Boosting protein's net-negative charge seems thus to be an efficient means to increase cytoplasmic mobility and rescue poor in-cell spectra. As proof of principle, we finally designed a protein to stick to the *E. coli* interior by decreasing the formal net charge of TTHA[PWT] with the construct TTHA[E58K]. The result is a nearly complete loss of the in-cell NMR signal (Table S1 and Fig. 2).

Strategy for Quantification of In-Cell Mobility.

To establish whether the ordered response to net charge alterations indeed relates to unspecific electrostatic repulsion, or stems from serendipitous obstruction of functional binding sites, we set out to quantify the effect. Under the present conditions, the NMR line width ($\Delta \nu_{1/2}$) scales with the protein correlation time (τ_c), which, in turn, is mainly determined by three components,

$$\tau_c^{-1} = \tau_r^{-1} + \tau_l^{-1} + \tau_{exch}^{-1}, \qquad [1]$$

where τ_r is the rotational correlation time, τ_l is the local dynamics correlation time, and τ_{exch} is the correlation time of any putative chemical exchange. Because the protein variants in this study are fully folded and maintain, in all cases, fixed 3D structures in the *E. coli* cytoplasm (*SI Controls*, Table S3, and Fig. S2), the
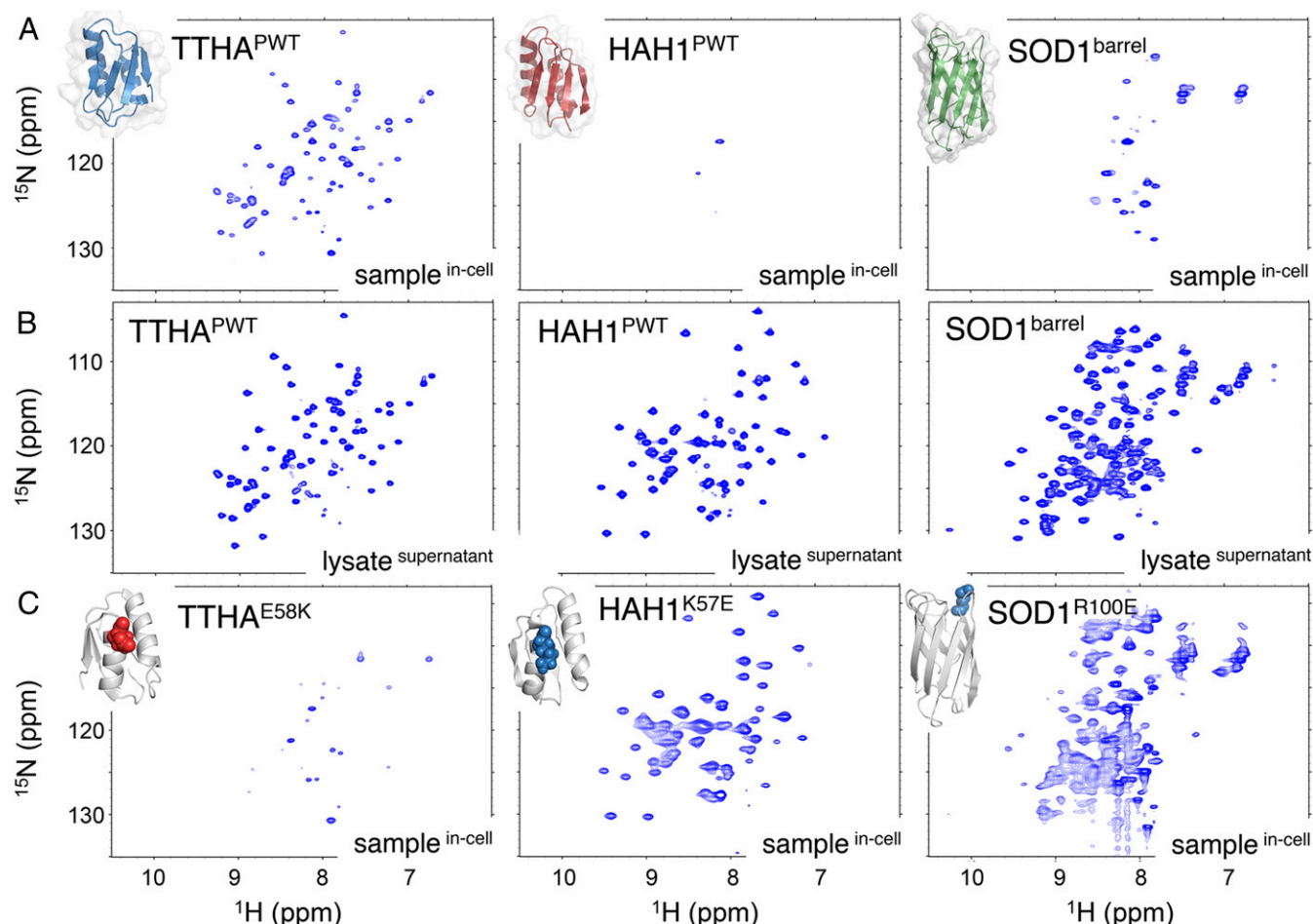
Fig. 2. In-cell HMQC spectra showing that the bacterial TTHA^{PWT} moves relatively unrestricted in the *E. coli* cytoplasm, whereas the human HAH1^{PWT} and SOD1^{barrel} get restricted by intracellular interactions, i.e., the protein motions become significantly retarded. The behavior, however, is readily reversed by point mutation of surface charges. (*A*) The well-resolved in-cell spectrum of TTHA^{PWT} indicates rapid motion in the cytoplasm, whereas the broadened-out spectra of HAH1^{PWT} and SOD1^{barrel} show that the in-cell motions are restricted (sample$^{in cell}$, intact cells). The cross-peaks of the latter spectra are largely from the background of small metabolites. (*B*) Corresponding spectra after cell lysis, showing that all three proteins move freely upon removal of the cytoplasmic restrictions (lysate^{supernatant}, decrowded lysates). (*C*) The in-cell motions respond readily to mutation of surface charge, allowing reversal of the pseudo-wild-type behavior. Reduction of negative charge causes the TTHA^{E58K} spectrum to broaden out. Conversely, addition of negative charge brings the spectra of HAH1^{K58E} and SOD1^{R100E} to high resolution.

contribution from local motions (τ_l) is negligible (36). Likewise, the uniform distribution of the line-broadening effect seen for all amino acids of the proteins (*SI Controls* and Fig. S2) indicates that the change in chemical exchange (τ_{exch}) upon transfer into the cells is small, as this would preferentially affect moieties on the protein surfaces. Hence, the line-broadening effect is global, i.e., the relative change in line width is similar for all cross-peaks. Consequently, the main modulator of NMR line width ($\Delta\nu_{1/2}$) can here be ascribed to the rotation correlation time (τ_r), in accordance with previous in-cell NMR studies (37, 38),

$$\tau_c \approx \tau_r = \frac{4\pi R_H^3 \eta}{3k_B T} = \frac{V_H \eta}{k_B T}, \qquad [2]$$

where V_H is the hydrodynamic volume and η is the apparent viscosity. Experimentally, we determined the change in τ_r (Eqs. **1** and **2**) upon transfer into the cytoplasm from the ratio of the cross-peak heights (h) in the lysate^{supernatant} and the sample$^{in cell}$ spectra according to

$$\text{mobility}^{in cell} = \tau_r^{lysate}/\tau_r^{in cell} = h^{in cell}/h^{lysate}, \qquad [3]$$

using the property that $h^{in cell}/h^{lysate} = \Delta\nu_{1/2}^{lysate}/\Delta\nu_{1/2}^{in cell}$ when protein concentration is kept constant. For simplicity, we refer tentatively to

mobility$^{in cell}$ as change in intracellular 'mobility' (37). A value of mobility$^{in cell}$ = 1 means that the in-cell mobility is similar to that in pure water, i.e., $\tau_r = 4$ ns for TTHA (28) and HAH1 (29) and 7 ns for SOD1^{barrel} (32), whereas mobility$^{in cell}$ = 0 means that the in-cell cross-peaks have broadened beyond detection (i.e., $\tau_r \gtrsim 60$ ns).

Quantification of Protein Net Charge. Because estimates of residue charge from model compound values fail to account for shifts in protonation state of titratable groups (i.e., shifted pKa values), induced by folding or mutation (39), we opted for analysis by electrophoretic mobility (40). Following standard protocols, each protein construct *i* (Table S1) was run on a native gel along with the TTHA^{PWT} reference (*SI Materials and Methods* and Fig. S4). The measured mobility ratio (R_f^i/R_f^{ref}) relates here to the protein net charge ratio (Z^i/Z^{ref}) according to (40)

$$R_f^i/R_f^{ref} = \left(Z^i E/f^i\right)/\left(Z^{ref} E/f^{ref}\right) = \left(Z^i/r_H^i\right)/\left(Z^{ref}/r_H^{ref}\right), \qquad [4]$$

where Z is the protein net charge, E is the applied electrical field, f is the friction coefficient, and r_H is the hydrodynamic radius = $f/6\pi\eta$, where η is the gel density. To obtain proportionality with in-cell interaction, r_H was converted to solvent-accessible surface
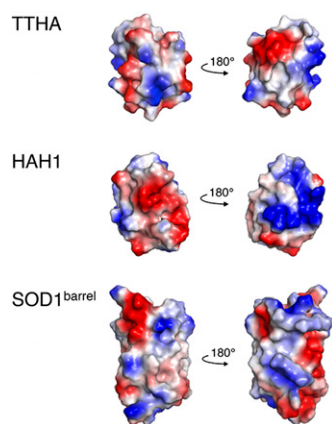
Fig. 3. The electrostatic potentials of the TTHA, HAH1, and SOD1^{barrel} surfaces as calculated by the software PyMol. Despite the bacterial TTHA and human HAH1 being homologous, their electrostatic surface patterns are very different. The projections of the structures to the left follow those in Fig. 1.

area (SASA $= 4\pi r_H^2$), yielding a dimensionless measure of the surface charge density ratio (charge^{density}) between construct i and the TTHA^{PWT} reference,

$$\text{charge}^{density} = (Z^i/\text{SASA}^i)/(Z^{ref}/\text{SASA}^{ref}) = \left(R_f^i/R_f^{ref}\right)\left(r_H^{ref}/r_H^i\right), \quad [5]$$

where R_f^i/R_f^{ref} is the observed gel mobility ratio (Eq. **5**) and r_H^{ref}/r_H^i is the radius-to-area conversion factor. The latter was determined by diffusion NMR, yielding r_H^{ref}/r_H^i = 1, 0.98, and 0.79 for TTHA^{PWT}, HAH1^{PWT}, and SOD1^{barrel}, respectively (Fig. S4 and Table S4). Because some proteins in Table S1 tend to lose net negative charge at pH values below 7, thus halting or reversing their electrophoretic mobility, the native gels had to be run according to standard protocols at pH 8.3 (*SI Materials and Methods*), where all constructs remain net negative. As a control, the eletrophoretically determined values of Z^i/Z^{ref} agree well with those calculated from model compound pK_A values (Fig. S4). Finally, to account also for the partial histidine protonation observed under intracellular conditions at pH 6.5 to 6.7 (*SI Controls*), we added the charge^{density} offsets 0.13, 0.25, and 0.32 for TTHA, HAH1, and SOD1^{barrel}, respectively (*SI Materials and Methods*). Although this minor correction falls largely within the data scatter, it is added for completeness of analysis.

In-Cell Mobility Shows Common Dependence on Net Charge Density. Analysis of a comprehensive set of mutations, including alterations of both surface charge and hydrophobicity (Table S1), shows that the bacterial protein TTHA undergoes a progressive increase of in-cell mobility upon increasing the net-negative charge (Fig. 4). Despite considerable scatter in the mobility$^{in\ cell}$ vs. charge^{density} plot (Eqs. **3** and **5**), the result is statistically significant with $R = 0.73$. A similarly scattered dependence on charge^{density} is found for the surface mutations of the human homolog HAH1 ($R = 0.84$), but, in this case, the plot is overall offset to lower mobility$^{in\ cell}$ values (Table S1 and Fig. 4). The trend is completed by the surface mutations of SOD1^{barrel}, adding a third—again seemingly parallel—data set below that of HAH1 ($R = 0.61$) (Table S1 and Fig. 4). On the whole, the results are consistent with net-negative repulsion being a fundamental factor in modulating in-cell solubility and protein–protein interactions (20–22). However, the discrete offsets in the mobility$^{in\ cell}$ vs. charge^{density} plots in Fig. 4 indicate that there are more factors than surface net charge at play in controlling the in-cell mobility. Contributions from such additional factors are also emphasized by the conspicuous data scatter, which is clearly

outside the errors of the individual measurements (Fig. 4). The data spread for the whole data set as measured by linear-fit deviation per point, i.e., the residual sum squared (RSS), is RSS = 0.015 (*SI Controls*, Fig. S3, and Table S5).

Notes on Variability, Protein Leakage, and Self-Interaction. As a measure of experimental variability, the data set in Fig. 4 contains—without any exclusions—all measurements performed, including the subset where the supernatant controls revealed some extent of leakage. The latter constitute 38 of a total of 203 experiments, and, on average, the indicated leakage was 16% (Table S1). To single out these experiments, the statistics of the analysis are given both for the total data set (RSS^{all}) and for the data set where the leakage set is excluded (RSS^{sub}), where the unit is in "per point," i.e., RSS total/number of data points. As seen below, however, the difference between RSS^{all} and RSS^{sub} is very small (Table S5). The explanation seems to be that the detected leakage is generally not present during the determination of mobility$^{in\ cell}$ in sample$^{in\ cell}$, but occurs during the handling of fatigued cells in the supernatant preparation. In support of this possibility, (*i*) the supernatant sample in some cases show distinct narrow cross-peaks that are missing in sample$^{in\ cell}$, (*ii*) repeated measurements with varying degrees of leakage yield still small variations in mobility$^{in\ cell}$, and (*iii*) the sample$^{in\ cell}$ signal is, in all cases, stable over the acquisition time and not gradually increasing as would be expected from on-going leakage (Fig. S1). Second, to examine whether the results show any dependence on expression level, we plotted mobility$^{in\ cell}$ and deviation from fit vs. intracellular concentration of overexpressed protein (Fig. S1). The controls reveal no self-interaction between the overexpressed proteins, i.e., the mobility$^{in\ cell}$ appears to stem from interaction with endogenous cellular components.

Surface Hydrophobicity: A Matter of Depth. As second modulator of in-cell mobility, we examined hydrophobicity. The trend is evident already at the level of the TTHA^{PWT}, HAH1^{PWT}, and SOD1^{barrel} structures, where the hydrophobic surface area as measured simply by exposure of V, L, and I methyl groups sums up to 8%, 12%, and 17%, respectively (Fig. 4). Moreover, the plot of mobility$^{in\ cell}$ vs. this fractional exposure seems to account for the protein-specific offsets in the mobility$^{in\ cell}$ vs. charge^{density} data (Fig. 4). To quantify the effect more strictly, we estimated, for each mutant, the change in hydrophobic exposure from energy minimizations of the TTHA^{PWT}, HAH1^{PWT}, and SOD1^{barrel} structures (*SI Materials and Methods*) according to

$$\text{hydrophobicity}^{SASA} = \text{SASA}_{hp}^{local}/\text{SASA}^{global}, \quad [6]$$

where SASA$_{hp}^{local}$ is the hydrophobic SASA and SASA^{global} is the total SASA (*SI Materials and Methods*). The resulting plot of mobility$^{in\ cell}$ vs. [charge^{density}; hydrophobicity^{SASA}] reveals a plane that captures the in-cell motion for all three proteins (Fig. 4). In essence, the offsets in the mobility$^{in\ cell}$ vs. charge^{density} plots (Fig. 4) are orderly displaced in the added hydrophobicity^{SASA} dimension, with an accompanying reduction of the scalar-fit deviations from RSS^{all} = 0.015 and RSS^{sub} = 0.014 to RSS^{all} = 0.012 and RSS^{sub} = 0.011 (Table S5). When it comes to the detailed effects of point mutations, however, the hydrophobicity^{SASA} measure appears relatively insensitive. Although we at mutant level generally discern decreased in-cell mobility upon increasing side-chain hydrophobicity, the resulting change in hydrophobicity^{SASA} can go either way because of local repacking (Table S1). Analogous surface adjustment is often seen upon binding of designed ligands (41, 42) and is here illustrated computationally by docking isobutane (valine mimic) to the SOD1^{barrel}: The ligand becomes almost entirely engrossed in the surface, driven by contacts that are buried in the crystal structure (Fig. S4). In an attempt to
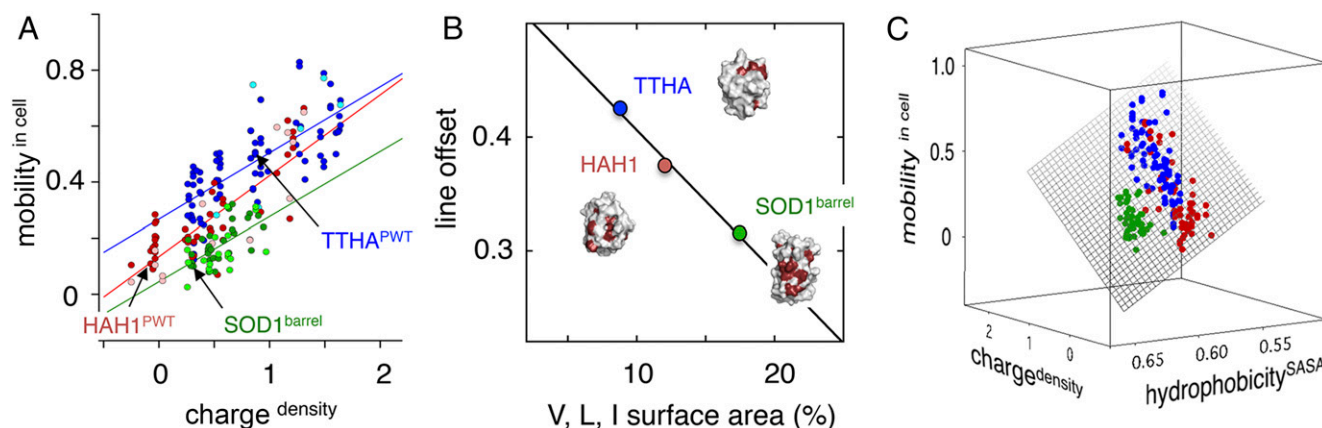
Fig. 4. First step toward quantification of mutational effects on in-cell motion. (*A*) The observed in-cell motions (mobility$^{\text{in cell}}$; Eq. **3**) display an overall scattered increase upon mutational increase of the negative charge density (charge$^{\text{density}}$; Eq. **5**). Even so, the three proteins show differences by being offset in the plot. Given the relatively small measurement errors, this finding indicates that other factors than surface charge density are at play in controlling the in-cell motion. Notably, the experiments with indicated leakage (lightly colored markers) do not stand out in the correlations, suggesting that this leakage occurs during cell handling after in-cell detection (*SI Controls*). (*B*) Plot of in-cell motion vs. exposure of the hydrophobic side chains V, L, and I. The simplistic estimate of surface hydrophobicity accounts for the fitted line offsets in *A* at charge$^{\text{density}}$ = 1. (*C*) Plot with refined estimate of hydrophobic surface area–based energy minimization of the mutant structures (hydrophobicity$^{\text{SASA}}$; Eq. **6**). The plot includes the entire data set in Table S1 and is extended to three dimensions, i.e., mobility$^{\text{in cell}}$ vs. [charge$^{\text{density}}$; hydrophobicity$^{\text{SASA}}$]. Upon accounting for the surface hydrophobicity variation, the data in *A* convert to a well-defined plane. Although the fit is overall improved, the parameter hydrophobicity$^{\text{SASA}}$ captures poorly the details of individual point mutations because of considerable local surface rearrangements in the energy minimizations.

quantify this dynamically accessible surface area (DASA), we assumed crudely that all residues that are not buried in the protein's hydrophobic core are free to interact with the cellular environment upon dynamic rearrangement (Fig. 5). Second, we assigned a hydrophilicity to each of the residues in the DASA layer according to their Guy solvation energies (43), $\Delta F_i'$ (Table S1). In essence, $\Delta F_i'$ is a statistical term describing the depth distribution of the different residues in protein crystal structures, which also correlates with the residue water-to-octanol partition free energies (43). Normalization to surface area was finally obtained by dividing the sum of the Guy solvation energies with the number of residues in the DASA layer (n^{DASA}) construct according to

$$\text{hydrophilicity}^{\text{Guy}} = \frac{1}{n^{\text{DASA}}} \sum_{n^{\text{DASA}}} \Delta F_i'. \qquad [7]$$

Interestingly, the plot mobility$^{\text{in cell}}$ vs. [charge$^{\text{density}}$; hydrophilicity$^{\text{Guy}}$] yields an alternative plane with RSS$^{\text{all}}$ = 0.011 and

RSS$^{\text{sub}}$ = 0.009 (Fig. 6 and Table S5), even though the values of hydrophilicity$^{\text{Guy}}$ and hydrophobicity$^{\text{SASA}}$ are themselves poorly correlated ($R = -0.32$; Table S1). Consistently, the RSS deviations decrease yet another step to RSS$^{\text{all}}$ = RSS$^{\text{sub}}$ = 0.009 upon including both parameters in the fit. Although hydrophobicity$^{\text{SASA}}$ and hydrophilicity$^{\text{Guy}}$ have similar weight in improving the overall correlation, the latter captures better the effects of point mutation. In the hydrophilicity$^{\text{Guy}}$ plot, the individual mutants of TTHA and HAH1 spread out and mix across the plane (Fig. 6 and Movie S1). Whether this indicates that the protein surfaces are indeed plastic by exposing buried material upon in-cell encounters, or simply reflects that the Guy potential—by including also the solvation energies of polar and charged residues—captures properties complementary to hydrophobicity$^{\text{SASA}}$ remains to be found out. Even so, the very convergence of experimental data from three unrelated proteins into a single plane shows that the quinary in-cell interactions—despite their detailed complexity—follow common and simplistic rules, governed by the proteins' macroscopic surface properties (Fig. 6). Moreover, the
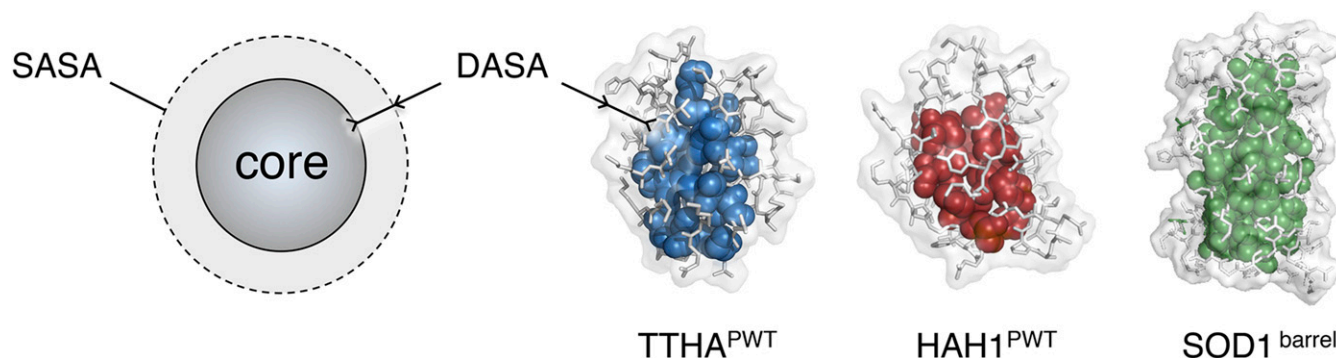
Fig. 5. Alternative method for determining the exposure of hydrophobic surface area. To account for the notion that protein surfaces are not perfectly "fixed" but, to some extent, adjust upon contact with external ligands (41, 42), we assumed that all residues outside the hydrophobic cores (transparent structural regions) are free to interact with the environment via dynamic rearrangement. The atoms in this dynamic layer define the DASA, which is considerably more extensive than the crystallographic SASA. For quantification and comparison with in-cell mobility data, we finally derived the Guy solvation energies (43) of each residue in the DASA layers to obtain the macroscopic parameter hydrophilicity$^{\text{Guy}}$ (Eq. **7**). The final relation between in-cell motion and hydrophilicity$^{\text{Guy}}$ is shown in Fig. 6 and Movie S1.
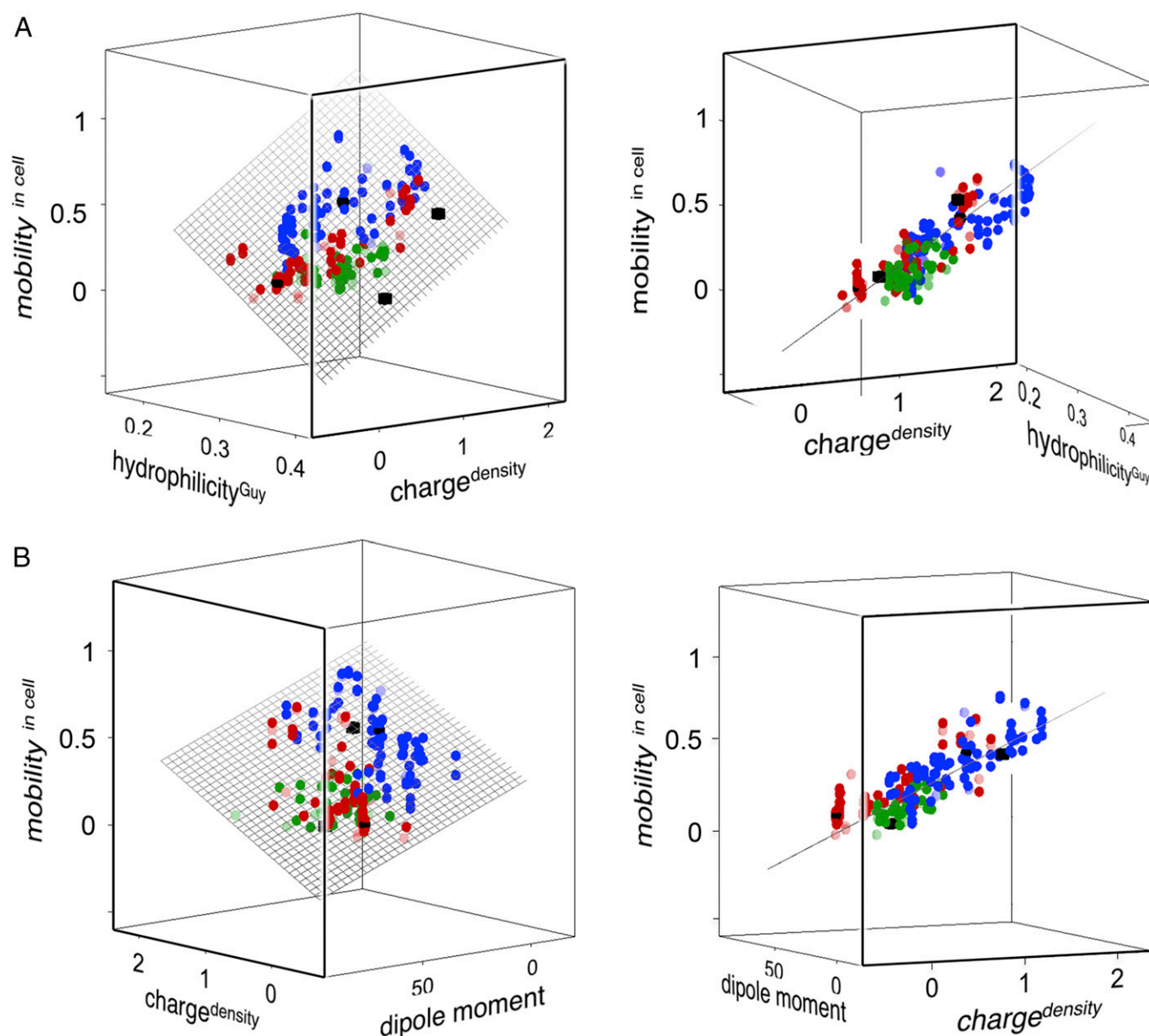
Fig. 6. The high-dimensional relation between *in-cell* motion and the proteins' macroscopic surface properties. Wild-type proteins are denoted by ■ (black solid squares), pseudo-wild-type variants are denoted by ● (black solid circles), and mutants are denoted by ○ (colored solid circles), color-coded as in previous figures. TTHA variants are in blue, HAH1 variants are in red, and SOD1[barrel] variants are in green. Data are from Table S1. Rotations of plots are shown in Movies S1 and S2. (*A*) A 3D plot of the experimentally measured mobility[in cell] (Eq. **3**) vs. the surface net charge density (charge[density]; Eq. **5**) and the surface hydrophilicity (hydrophilicity[Guy]; Eq. **7**). (*B*) Corresponding plot of mobility[in cell] (Eq. **3**) vs. the surface net charge density (charge[density]; Eq. **5**) and the electric dipole moment (||**p**||; Eq. **8**). The correlations in *A* and *B* call attention to the strikingly systematic response of the quinary in-cell interactions to point mutation. Also, the correlations appear generic because they capture all protein constructs, including those of different evolutionary origin.

data in Fig. 6 indicate that the cellular components responsible for the in-cell retardation carry repulsive negative charge and attractive surface hydrophobicity. Species matching these characteristics are, e.g., other proteins, membranes, and nucleic acid material.

Enhancing the Precision One Step Further: The Protein Dipole Moment. In some analogy with frustrated spin-glass systems (44), charged proteins possess electric dipole moments that are expected to influence their motions under crowded in-cell conditions. Upon mutation of surface charges, these dipole moments undergo distinct changes with possible impact on mobility[in cell] (Eq. **3**). The magnitude of the dipole moment (||**p**||) for each protein construct in Table S1 was calculated from the charge positions in

the PDB structures of TTHA[PWT], HAH1[PWT], and SOD1[barrel] according to (45)

$$\|\boldsymbol{p}\| = \left\| \sum_{i=1}^{N} q_i(\mathbf{r_i} - \mathbf{r}) \right\|, \qquad [8]$$

where q_i is the elemental charge at position $\mathbf{r_i}$, and $\mathbf{r}$ is the protein center of mass, treating the R/K and E/D side chains as positive and negative point charges, respectively. The histidines were omitted from the calculation because their pK_A values are not yet known. Conveniently, because ||**p**|| is a general vector in space, it does not need normalization to protein surface area to scale with in-cell interaction, as required for the protein net

charge Z (Eq. **5**) and the Guy solvation energies $\Delta F_i'$ (Eq. **7**). The impact of $\|\boldsymbol{p}\|$ on in-cell motions emerges clearly in a plot of mobility$^{\text{in cell}}$ vs. [charge$^{\text{density}}$; $\|\boldsymbol{p}\|$], which defines a plane (RSS$^{\text{all}}$ = 0.010 and RSS$^{\text{sub}}$ = 0.009) similar to that of mobility$^{\text{in cell}}$ vs. [charge$^{\text{density}}$; hydrophilicity$^{\text{Guy}}$] (Fig. 6 and Movie S2). In other words, the dipolar interactions that influence in-cell protein motion are on par with hydrophilicity$^{\text{Guy}}$. Upon final extension of the correlation to four dimensions, i.e., mobility$^{\text{in cell}}$ vs. [charge$^{\text{density}}$; hydrophilicity$^{\text{Guy}}$; $\|\boldsymbol{p}\|$], the scalar fit deviations decrease just slightly to RSS$^{\text{all}}$ = 0.009 and RSS$^{\text{sub}}$ = 0.008, indicating that we have reached the limit of the data scatter (Fig. S3 and Table S5).

Discussion

The Nature of In-Cell Encounters. Although the question of how a given protein senses and moves in the intracellular compartment of *E. coli* may at first seem intricate, our results indicate that this process conforms to simplistic physical–chemical rules. Measured by the rotational freedom (Eq. **3**), the intracellular motions (mobility$^{\text{in cell}}$) of 127 mutant constructs fall coarsely on a common plane in the protein property space (Fig. 6). The parameters of this property space are, moreover, easy to deduce for any protein with known structure, i.e., the surface net charge density (charge$^{\text{density}}$; Eq. **5**), the surface hydrophilicity (hydrophilicity$^{\text{Guy}}$; Eq. **7**), and the charge dipole moment ($\|\boldsymbol{p}\|$; Eq. **8**). Out of these properties, the net charge density is the most decisive (*SI Controls*), consistent with findings in earlier in-cell NMR studies (38, 46) and the notion that net charge repulsion generically assures dispersion of the cellular components (21, 47, 48). The lack of correlation between the intracellular concentration of overexpressed protein (1.6 ± 0.8 mM; *SI Materials and Methods*) and mobility$^{\text{in cell}}$ (Fig. S1) indicates further that the dynamic changes are not from self-interactions but stem from quinary interactions with other cytoplasmic components. Because the in-cell measurements only report on the specific isotope-labeled construct being overexpressed, however, it is not yet possible to determine which components of the *E. coli* cytoplasm cause the in-cell retardation; it could be neighboring soluble proteins, cytoskeleton structures, proteomembrane surfaces, nucleotide material (49), or any species presenting repulsive negative charge and attractive surface hydrophobicity. In any case, the overall morphology of the bacteria is similar when overexpressing the different proteins (*SI Controls* and Fig. S3), suggesting that the intracellular environment remains unperturbed. However, it is apparent, from the NMR spectral features, that the underlying quinary encounters are short-lived and diffusive, in agreement with previous observations (5, 37, 38, 50). As the correlations in Fig. 6 concur with the average properties of the mutated proteins, we further conclude that the motions are mainly controlled nonspecifically by diffusive encounters with the "bulk" cellular background, i.e., the quinary crosstalk (15). By no means does this exclude that the data in Fig. 6 also involve in-cell interactions that rely on the site-specific details of the protein surfaces, but, as yet, these seem to be within the data scatter.

Implications for Functional Optimization and Protein Design. In view of the rapid divergence of surface composition in protein evolution (51), it is interesting that the human HAH1$^{\text{PWT}}$ and SOD1$^{\text{barrel}}$, which are seen to tumble freely in mammalian cells (31, 52), tend to get stuck in the *E. coli* cytoplasm (Fig. 2). The bacterial protein TTHA$^{\text{PWT}}$, on the other hand, yields high-resolution HMQC spectra in *E. coli* from the very start (Fig. 2). Whether these differences in cytoplasmic mobility indeed stem from evolutionary divergence, i.e., mammalian proteins are not fit for *E. coli*, or simply reflect the natural span of intracellular mobilities is not yet clear. Even so, it is evident from the ease with which our model proteins can be moved across the

mobility span (Fig. 6) that the quinary crosstalk is readily open to optimization. One conceivable gain of such optimization is in molecular search (12). At one extreme, very long-lived encounter complexes will increase the chance of two functional partners finding their right orientations by Brownian surface diffusion (3), but also slow down the shift to alternative partners when the match is wrong. Encounter complexes that are very swift will conversely allow probing of more putative partners, but with the risk that correct fits are missed by premature dissociation. Regardless of what the answer may be, the quinary interactions stand out as a significant part of the intracellular crosstalk and provide also the ever-present background for in vivo function (10). It is further evident from the results in Figs. 2–5 and previous studies (9, 46) that the quinary interactions, despite their complexity, at some level follow simplistic macroscopic rules. As such, the correlations in Fig. 6 not only add physical–chemical detail to the effects of intracellular crowding (4, 12, 27, 53) but also provide a tool for rational protein surface design. Applications can include optimization of target proteins for in-cell NMR detection (54), surface optimization of protein therapeutics (55), and mutational examination of the yet poorly understood relation between protein motion, spatial localization, and function (7, 56, 57). The message stands clear and simple to test: Can any protein be tuned to desired rotational motion in *E. coli.* by mutational tweaking of surface net charge density (Eq. **5**), surface hydrophilicity (Eqs. **6** and **7**), and the electric dipole moment (Eq. **8**)?

Materials and Methods

Protein Engineering. Mutagenesis, expression, and purification of HAH1, TTHA, SOD1$^{\text{barrel}}$, and their variants were as in refs. 5 and 32, where the encoding genes were subcloned into the vector pET3a (GenScript) (*SI Materials and Methods*).

Overexpression and In-Cell NMR Sample Preparation. The proteins (Table S1 and Fig. S1) were produced and measured in *E. coli* cells. Before induction, BL21(DE3)pLysS cells (Thermo Fisher Scientific) were grown overnight at 37 °C in 200 mL of LB. The cells were then harvested and resuspended in M9 medium with labeled ^{15}NH$_4$Cl as sole nitrogen source, and overexpression was induced for 4 h (*SI Materials and Methods* and Fig. S1). The culture was then nondisruptively centrifuged at 800 × *g* for 8 min to obtain a pellet of intact, viable cells. One gram of this cell pellet was dissolved in 1 mL of M9 buffer, transferred to an NMR tube, and subjected to in-cell spectral analysis (*SI Materials and Methods*). All experiments were performed at 37 °C unless otherwise stated.

NMR Spectroscopy. NMR experiments were performed on a Bruker Avance 500- or 700-MHz spectrometer equipped with a triple-resonance cryogenically cooled probe head. In-cell and in vitro spectra were obtained by 1D and 2D ^{1}H-[^{15}N]-band-selective optimized flip-angle short-transient heteronuclear multiple quantum coherence (^{1}H-[^{15}N]-SOFAST HMQC) pulse schemes (58, 59) with 32 scans, 64 increments for the 2D experiment and 1,024 scans for 1D experiment. The relaxation delay was set to 0.2 s, and the acquisition time was to 40 ms. Hydrodynamic radii were determined by pulsed field gradient (PFG) NMR diffusion experiments, where the *z*-gradient strength was calibrated by the known diffusion coefficient of α-cyclodextrin and HDO in 99.6% D$_2$O (60) (*SI Materials and Methods*).

Mobility Determination on Native Gel. To quantify net charge density, 2 μL to 5 μL of lysate samples of the various protein variants (Table S1) were loaded on precasted gels (Bio-Rad), with TTHA$^{\text{PWT}}$ as references (R_f^{ref}) and HAH1$^{\text{PWT}}$ as reference control. For the SOD1$^{\text{barrel}}$ variants, we used, additionally, SOD1$^{\text{barrel}}$ for benchmarking/verification of R_f^{ref} normalization. Running time was 90 min to 210 min, at 125 V and 4 °C, and the running buffer was 25 mM Tris plus 192 mM Glycine at pH 8.3 (*SI Materials and Methods*).

1. Moreira IS, Fernandes PA, Ramos MJ (2007) Hot spots—A review of the protein-protein interface determinant amino-acid residues. *Proteins* 68:803–812.

2. Schreiber G, Keating AE (2011) Protein binding specificity versus promiscuity. *Curr Opin Struct Biol* 21:50–61.

3. Schreiber G, Fersht AR (1996) Rapid, electrostatically assisted association of proteins. *Nat Struct Biol* 3:427–431.

4. Ross JL (2016) The dark matter of biology. *Biophys J* 111:909–916.

5. Danielsson J, et al. (2015) Thermodynamics of protein destabilization in live cells. *Proc Natl Acad Sci USA* 112:12402–12407.

6. Zarrinpar A, Park SH, Lim WA (2003) Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* 426:676–680.

7. French JB, et al. (2016) Spatial colocalization and functional link of purinosomes with mitochondria. *Science* 351:733–737.

8. McGuffee SR, Elcock AH (2010) Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLOS Comput Biol* 6:e1000694.

9. Smith AE, Zhou LZ, Gorensek AH, Senske M, Pielak GJ (2016) In-cell thermodynamics and a new role for protein surfaces. *Proc Natl Acad Sci USA* 113:1725–1730.

10. Bhattacharyya S, et al. (2016) Transient protein-protein interactions perturb E. coli metabolome and cause gene dosage toxicity. *eLife* 5:e20309.

11. McConkey EH (1982) Molecular evolution, intracellular organization, and the quinary structure of proteins. *Proc Natl Acad Sci USA* 79:3236–3240.

12. Mika JT, Poolman B (2011) Macromolecule diffusion and confinement in prokaryotic cells. *Curr Opin Biotechnol* 22:117–126.

13. Dhar A, et al. (2011) Protein stability and folding kinetics in the nucleus and endoplasmic reticulum of eucaryotic cells. *Biophys J* 101:421–430.

14. Guo M, Xu Y, Gruebele M (2012) Temperature dependence of protein folding kinetics in living cells. *Proc Natl Acad Sci USA* 109:17863–17867.

15. Monteith WB, Cohen RD, Smith AE, Guzman-Cisneros E, Pielak GJ (2015) Quinary structure modulates protein stability in cells. *Proc Natl Acad Sci USA* 112:1739–1742.

16. Ignatova Z, et al. (2007) From the test tube to the cell: Exploring the folding and aggregation of a beta-clam protein. *Biopolymers* 88:157–163.

17. Inomata K, et al. (2009) High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* 458:106–109.

18. Sarkar M, Smith AE, Pielak GJ (2013) Impact of reconstituted cytosol on protein stability. *Proc Natl Acad Sci USA* 110:19342–19347.

19. Guzman I, Gelman H, Tai J, Gruebele M (2014) The extracellular protein VlsE is destabilized inside cells. *J Mol Biol* 426:11–20.

20. Gitlin I, Carbeck JD, Whitesides GM (2006) Why are proteins charged? Networks of charge-charge interactions in proteins measured by charge ladders and capillary electrophoresis. *Angew Chem Int Ed Engl* 45:3022–3060.

21. Kurnik M, Hedberg L, Danielsson J, Oliveberg M (2012) Folding without charges. *Proc Natl Acad Sci USA* 109:5705–5710.

22. Andersson O (2013) Cellular electrolyte metabolism. *Encyclopedia of Metalloproteins* (Springer, New York), pp 580–587.

23. Ghaemmaghami S, Oas TG (2001) Quantitative protein stability measurement in vivo. *Nat Struct Biol* 8:879–882.

24. Monteith WB, Pielak GJ (2014) Residue level quantification of protein stability in living cells. *Proc Natl Acad Sci USA* 111:11335–11340.

25. Dhar A, Ebbinghaus S, Shen Z, Mishra T, Gruebele M (2010) The diffusion coefficient for PGK folding in eukaryotic cells. *Biophys J* 99:L69–L71.

26. Smith AE, Zhang Z, Pielak GJ, Li C (2015) NMR studies of protein folding and binding in cells and cell-like environments. *Curr Opin Struct Biol* 30:7–16.

27. Rivas G, Minton AP (2016) Macromolecular crowding in vitro, in vivo, and in between. *Trends Biochem Sci* 41:970–981.

28. Sakakibara D, et al. (2009) Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* 458:102–105.

29. Anastassopoulou I, et al. (2004) Solution structure of the apo and copper(I)-loaded human metallochaperone HAH1. *Biochemistry* 43:13046–13053.

30. Danielsson J, et al. (2013) Global structural motions from the strain of a single hydrogen bond. *Proc Natl Acad Sci USA* 110:3829–3834.

31. Danielsson J, et al. (2013) Pruning the ALS-associated protein SOD1 for in-cell NMR. *J Am Chem Soc* 135:10266–10269.

32. Danielsson J, Kurnik M, Lang L, Oliveberg M (2011) Cutting off functional loops from homodimeric enzyme superoxide dismutase 1 (SOD1) leaves monomeric β-barrels. *J Biol Chem* 286:33070–33083.

33. Halle B, Wennerström H (1981) Interpretation of magnetic resonance data from water nuclei in heterogeneous systems. *J Chem Phys* 75:1928–1943.

34. Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 2. Analysis of experimental results. *J Am Chem Soc* 104:4559–4570.

35. Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* 104:4546–4559.

36. Lee AL, Wand AJ (1999) Assessing potential bias in the determination of rotational correlation times of proteins by NMR relaxation. *J Biomol NMR* 13:101–112.

37. Wang Q, Zhuravleva A, Gierasch LM (2011) Exploring weak, transient protein–protein interactions in crowded in vivo environments by in-cell nuclear magnetic resonance spectroscopy. *Biochemistry* 50:9225–9236.

38. Barbieri L, Luchinat E, Banci L (2015) Protein interaction patterns in different cellular environments are revealed by in-cell NMR. *Sci Rep* 5:14456.

39. Oliveberg M, Arcus VL, Fersht AR (1995) pKA values of carboxyl groups in the native and denatured states of barnase: The pKA values of the denatured state are on average 0.4 units lower than those of model compounds. *Biochemistry* 34:9424–9433.

40. Cantor CR, Schimmel PR (1980) *Biophysical Chemistry, Part 2: Techniques for the Study of Biological Structure and Function* (W. H. Freeman, New York).

41. Teague SJ (2003) Implications of protein flexibility for drug discovery. *Nat Rev Drug Discov* 2:527–541.

42. Tsao D, Diatchenko L, Dokholyan NV (2011) Structural mechanism of S-adenosyl methionine binding to catechol O-methyltransferase. *PLoS One* 6:e24287.

43. Guy HR (1985) Amino acid side-chain partition energies and distribution of residues in soluble proteins. *Biophys J* 47:61–70.

44. Ferreiro DU, Komives EA, Wolynes PG (2014) Frustration in biomolecules. *Q Rev Biophys* 47:285–363.

45. Antosiewicz J (1995) Computation of the dipole moments of proteins. *Biophys J* 69:1344–1354.

46. Cohen RD, Pielak GJ (2016) Electrostatic contributions to protein quinary structure. *J Am Chem Soc* 138:13139–13142.

47. Lang L, Kurnik M, Danielsson J, Oliveberg M (2012) Fibrillation precursor of superoxide dismutase 1 revealed by gradual tuning of the protein-folding equilibrium. *Proc Natl Acad Sci USA* 109:17868–17873.

48. Vendruscolo M, Dobson CM (2007) Chemical biology: More charges against aggregation. *Nature* 449:555.

49. Frembgen-Kesner T, Elcock AH (2013) Computer simulations of the bacterial cytoplasm. *Biophys Rev* 5:109–119.

50. Luh LM, et al. (2013) Molecular crowding drives active Pin1 into nonspecific complexes with endogenous proteins prior to substrate recognition. *J Am Chem Soc* 135:13796–13803.

51. Sasidharan R, Chothia C (2007) The selection of acceptable protein mutations. *Proc Natl Acad Sci USA* 104:10080–10085.

52. Banci L, et al. (2013) Atomic-resolution monitoring of protein maturation in live human cells by NMR. *Nat Chem Biol* 9:297–299.

53. Gnutt D, Ebbinghaus S (2016) The macromolecular crowding effect—From in vitro into the cell. *Biol Chem* 397:37–44.

54. Freedberg DI, Selenko P (2014) Live cell NMR. *Annu Rev Biophys* 43:171–192.

55. Leader B, Baca QJ, Golan DE (2008) Protein therapeutics: A summary and pharmacological classification. *Nat Rev Drug Discov* 7:21–39.

56. Wirth AJ, Gruebele M (2013) Quinary protein structure and the consequences of crowding in living cells: Leaving the test-tube behind. *BioEssays* 35:984–993.

57. Srere PA (2000) Macromolecular interactions: Tracing the roots. *Trends Biochem Sci* 25:150–153.

58. Schanda P, Brutscher B (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J Am Chem Soc* 127:8014–8015.

59. Schanda P, Kupce E, 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.

60. Danielsson J, Jarvet J, Damberg P, Gräslund A (2004) Two-site binding of beta-cyclodextrin to the Alzheimer Abeta(1-40) peptide measured with combined PFG-NMR diffusion and induced chemical shifts. *Biochemistry* 43:6261–6269.

61. Sezonov G, Joseleau-Petit R, D'Ari R (2007) *Escherichia coli* physiology in Luria-Bertani broth. *J Bacteriol* 189:8746–8749.

62. Kubitschek HE, Friske JA (1986) Determination of bacterial cell volume with the Coulter Counter. *J Bacteriol* 168:1466–1467.

63. Lindberg MJ, Normark J, Holmgren A, Oliveberg M (2004) Folding of human superoxide dismutase: Disulfide reduction prevents dimerization and produces marginally stable monomers. *Proc Natl Acad Sci USA* 101:15893–15898.

64. Nordlund A, Oliveberg M (2006) Folding of Cu/Zn superoxide dismutase suggests structural hotspots for gain of neurotoxic function in ALS: Parallels to precursors in amyloid disease. *Proc Natl Acad Sci USA* 103:10218–10223.

65. Fersht AR (1995) Optimization of rates of protein folding: The nucleation–condensation mechanism and its implications. *Proc Natl Acad Sci USA* 92:10869–10873.

66. Damberg P, Jarvet J, Gräslund A (2001) Accurate measurement of translational diffusion coefficients: A practical method to account for nonlinear gradients. *J Magn Reson* 148:343–348.

67. Cavallo L, Kleinjung J, Fraternali F (2003) POPS: A fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucleic Acids Res* 31:3364–3366.

68. Yin S, Ding F, Dokholyan NV (2007) Eris: An automated estimator of protein stability. *Nat Methods* 4:466–467.

69. Yin S, Ding F, Dokholyan NV (2007) Modeling backbone flexibility improves protein stability estimation. *Structure* 15:1567–1576.

70. Ding F, Dokholyan NV (2006) Emergence of protein fold families through rational design. *PLOS Comput Biol* 2:e85.

71. Ding F, Yin S, Dokholyan NV (2010) Rapid flexible docking using a stochastic rotamer library of ligands. *J Chem Inf Model* 50:1623–1632.

72. Abolliel AA, Zedan H (2015) Synthesis, cloning and expression of a novel preminiproinsulin analogue gene in *Escherichia coli*. *J Adv Res* 6:663–671.

73. Tai J, Dave K, Hahn V, Guzman I, Gruebele M (2016) Subcellular modulation of protein VlsE stability and folding kinetics. *FEBS Lett* 590:1409–1416.

74. Matsuo H, et al. (1999) Identification by NMR spectroscopy of residues at contact surfaces in large, slowly exchanging macromolecular complexes. *J Am Chem Soc* 121:9903–9904.

75. Cornett JB, Shockman GD (1978) Cellular lysis of *Streptococcus faecalis* induced with triton X-100. *J Bacteriol* 135:153–160.

76. Smith PK, et al. (1985) Measurement of protein using bicinchoninic acid. *Anal Biochem* 150:76–85.

77. Zimmerman SB, Trach SO (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J Mol Biol* 222:599–620.

78. Bring J (1994) How to standardize regression-coefficients. *Am Stat* 48:209–213.

Supporting Information

Mu et al. 10.1073/pnas.1621227114

SI Materials and Methods

Mutagenesis, Expression, and Purification. Mutagenesis, expression, and purification of HAH1, TTHA, SOD1$^{\text{barrel}}$, and their variants were as in refs. 5 and 32, where the encoding genes were subcloned into the vector pET3a (GenScript). To remove functional binding of metals and prevent disulphide linking, the involved cysteine ligands were replaced by serine. For TTHA, the replacements were C11S/C14S, and, for HAH1, they were C12S/C15S/C41S. The resulting pseudo-wild-type constructs are denoted HAH1$^{\text{PWT}}$ and TTHA$^{\text{PWT}}$.

NMR Sample Preparations and Experimental Conditions. Before induction, the *E. coli* BL21 (DE3) pLysS cells (Thermo Fisher Scientific) were grown overnight at 37 °C in 200 mL of LB medium in the presence of 100 μg/mL of carbenicillin and 34 μg/mL of chloramphenicol. The cells were then harvested by centrifugation (8 min at $800 \times g$) and resuspended in M9 medium to $OD_{600} \approx 1.5$ with labeled $^{15}NH_4Cl$ as sole nitrogen source, and protein expression was induced by 0.5 mM IPTG. All experiments were performed in M9 buffer at 37 °C unless otherwise stated.

Intracellular Concentration of Overexpressed Proteins. The cytoplasmic concentration of ^{15}N-labeled protein ([protein]$^{\text{in cell}}$) following 4 h of overexpression was estimated as follows. First, the protein concentration in the NMR detection volume ([protein]$^{\text{lysate}}$) was determined from the peak intensities of decrowded cell lysate (lysate$^{\text{supernatant}}$), using reference spectra obtained at fixed concentrations of [protein] = 100 μM. Subsequently, the intracellular protein concentration was calculated from [protein]$^{\text{in cell}}$ = [protein]$^{\text{lysate}}/\varphi$, where $\varphi \approx 0.29$ is the *E. coli* volume fraction of sample$^{\text{in cell}}$. The volume fraction $\varphi \approx 0.29$ was calculated from the concentration of cells in the sample$^{\text{in cell}}$, derived from the number of cells in 1 g of cell pellet using optical density at 600 nm (OD_{600}) $\approx 10^9$ cells per milliliter per OD (58), and the volume of the individual *E. coli* cells, i.e., 1 μm^3 per cell (59). Based on the [protein]$^{\text{lysate}}$ values in Table S2, the [protein]$^{\text{in cell}}$ were estimated to 1.6 ± 0.8 mM (Fig. S1*E*) for the various overexpressed constructs in this study. There is no correlation between [protein]$^{\text{in cell}}$ and mobility$^{\text{in cell}}$ (Fig. S1 *F* and *G*).

Mobility Determination on Native Gel. To quantify net charge density, 2 μL to 5 μL of lysate samples of the various protein variants (Table S1) were loaded on precasted gels (Bio-Rad), with TTHA$^{\text{PWT}}$ as references (R_f^{ref}) and HAH1$^{\text{PWT}}$ as reference control. For the SOD1 variants, we used, additionally, SOD1$^{\text{barrel}}$ for benchmarking/verification of R_f^{ref} normalization. Running time was 90 min to 210 min, at 125 V and 4 °C, and the running buffer was 25 mM Tris plus 192 mM Glycine at pH 8.3. After staining and destaining, the mobility in millimeters of each protein variant (R_f) was determined and normalized according to Fig. S4 *A* and *B* and Eq. **4**. The vertical electrophoresis cell was from Bio-Rad, and the power supply was from Thermo Fisher Scientific.

Correction for Partial Histidine Protonation. Because the electrophoretic mobility at pH 8.3 (R_f^i/R_f^{ref}; Fig. S4*A*) does not report on the differences in histidine charge between TTHA$^{\text{PWT}}$, HAH1$^{\text{PWT}}$, and SOD1$^{\text{barrel}}$ under intracellular conditions at pH 6.5 to 6.7 (*SI Controls*), these were accounted for separately as follows. From the partial histidine protonation observed by *E. coli* in-cell NMR (*SI Controls*) and the structures in Fig. S1*A*, we estimate the total histidine charge (His$^{\text{tot}}$) to be +0.5 for TTHA (H13), +1 for

HAH1 (H4, H46), and +2 for SOD1$^{\text{barrel}}$ (H46, H48, H80, H90). These His$^{\text{tot}}$ values can, in turn, be translated to Δcharge$^{\text{density}}$ for each protein according to

$$\Delta\text{charge}^{\text{density}} = \left(\text{His}^{\text{tot}} \times \delta(Z^i/Z^{\text{ref}})/\delta Z^{\text{calc}}\right)/r_{\text{H}}^2, \quad \text{[S1]}$$

where $\delta(Z^i/Z^{\text{ref}})/\delta Z^{\text{calc}} = -0.26$ is the slope of the linear fit in Fig. S4*F*, and r_{H}^2 is the surface normalization by protein's hydrodynamic radius (Fig. S4 *C–E* and Table S4). The values of Δcharge$^{\text{density}}$ are −0.13, −0.25, and −0.32 for TTHA, HAH1, and SOD1, respectively, and have been used as offsets for the charge$^{\text{density}}$ (Eq. **5**) throughout the manuscript.

Folding Kinetics. Measurements of folding kinetics were as described in references (60, 61). In short, guanidinium chloride (AppliChem) and urea (MP Biomedicals) of ultrapure grade was used for protein denaturation. All kinetic measurements were performed in 20 mM pH 6.3 Mes buffer. The final concentration of SOD1$^{\text{barrel}}$ was 4 μM, whereas, for TTHA$^{\text{PWT}}$, TTHA$^{\text{E57K}}$, HAH1$^{\text{PWT}}$, and HAH1$^{\text{K57E}}$, it was increased to 20 μM to allow for tyrosine fluorescence detection. The observed relaxation rates, k_{obs}, were fitted by standard procedures according to (62)

$$\log k_{\text{obs}} = \log(k_{\text{f}} + k_{\text{u}}) = \log\left(10^{\log k_{\text{f}}^{\text{H}_2\text{O}} + m_{\text{f}}[\text{denaturant}]}\right.$$
$$\left. + 10^{\log k_{\text{u}}^{\text{H}_2\text{O}} + m_{\text{u}}[\text{denaturant}]}\right), \quad \text{[S2]}$$

where, k_{f} and k_{u} are the refolding and unfolding rate constants, respectively, $k_{\text{f}}^{\text{H}_2\text{O}}$ and $k_{\text{u}}^{\text{H}_2\text{O}}$ are their values at [denaturant] = 0 M, and m_{f} and m_{u} are the slopes used for extrapolation. Data analysis was with KaleidaGraph (Synergy Software).

NMR Spectroscopy. NMR experiments were performed on a Bruker Avance 500- or 700-MHz spectrometer equipped with a triple-resonance cryogenically cooled probe head. The in-cell, as well as the in vitro, experiments were obtained using 1D and 2D HMQC pulse schemes (63, 64) with 32 scans, 64 increments for the 2D experiment and 1,024 scans for the 1D experiment. The relaxation delay was set to 0.2 s, and the acquisition time was set to 40 ms. Line broadening was measured from the sum of peak heights for resonances between 8.6 ppm and 9.4 ppm, and a normalized value of the in-cell retardation was obtained from the ratio of the peak heights in the lysate supernatant and the in-cell samples according to Eq. **3**.

The hydrodynamic radii of TTHA$^{\text{PWT}}$ and HAH1$^{\text{PWT}}$ were determined by PFG NMR diffusion experiments, where the *z*-gradient strength was calibrated by the known diffusion coefficient of α-cyclodextrin and HDO in 99.6% D_2O (65). The diffusion experiments were conducted by altering the effective gradient strength from 0.5 G/cm to 25.5 G/cm in 16 steps, and the attenuation signal was fitted to a modified Stejskal−Tanner equation (66) to account for nonlinearity of the gradient field.

Transmission Electron Microscopy. The bacterial cells were fixed in 2.5% glutaraldehyde and 1% paraformaldehyde in 0.1 M phosphate buffer at pH 7.4, and incubated overnight at 4 °C. After centrifugation, the pellet was rinsed in 0.1 M phosphate buffer at pH 7.4, followed by postfixation in 2% osmium tetroxide in 0.1 M phosphate buffer at pH 7.4 for 2 h at 4 °C. The sample was then dehydrated in ethanol, followed by acetone, and finally embedded in LX-112 (Ladd). Ultrathin sections (∼50 nm to 60 nm) were

cut by a Leica ultracut UCT (Leica). Sections were contrasted with uranyl acetate followed by lead citrate and examined in a Hitachi HT 7700 at 80 kV. Digital images were taken with a Veleta camera (Olympus Soft Imaging Solutions, GmbH). The transmission electron microscopy was performed at the Electron Microscopy Unit at Karolinska University Hospital, in Sweden.

Calculation of Hydrophobicity^{SASA}. Hydrophobic SASA were calculated for each protein construct using the Parameter Optimised Surfaces (POPS) method (67). The values presented are the mean values calculated over 20 mutant structures. To obtain the mutant structures, we used the program Eris (68, 69). In each round of the Eris calculation, 20 structures of the native and mutant proteins were produced. Each structure was the result of 20 steps of Monte Carlo simulation repacking the side chains within 10 Å of the mutation site. From the 20 mutant and 20 native structures, the program also computed the mean change in free energy between the mutant and native structures using the Medusa force field (70). For each mutation, we performed 20 independent rounds of Eris calculations. We selected the round with the lowest mean change in free energy between mutant and native proteins. The 20 mutant structures within this round of the Eris calculation were used to calculate the solvent accessible surface areas, $SASA_{hp}^{local}$ and $SASA^{global}$ in Eq. **6**.

Docking of Isobutane to the SOD1 Surface. The docking analysis was performed using MedusaDock, a docking program that simultaneously models the flexibility of the receptor and ligand (71). We performed 200 independent docking calculations and selected, finally, the structure with the lowest energy as evaluated by the Medusa force field (70) (Fig. S4 G and H).

SI Controls

Expression Levels, Cytoplasmic Protein Concentration, and Signal Stability over Time. *E. coli* BL21 (DE3) pLysS cells were used as overexpression system. This strain has minimal background leakage expression (72), enabling us to keep the levels of nonlabeled protein low. After 4 h of expression, the cytoplasmic concentration of ^{15}N-labeled protein was estimated to be 1.6 ± 0.8 mM (Fig. S1E) from the number of *E. coli* cells [OD$_{600}$ ≈ 10^9 cells per milliliter per OD (58)], the *E. coli* volume [~1 μm^3 per cell (59)], and the peak intensity of the lysate NMR spectra, using reference spectra with known concentrations. We find no correlation ($R = 0.22$) between expression level and observed mobility$^{in\ cell}$ (Fig. S1F), or between data deviation from the fitted plane and expression level (Fig. S2G and Table S2). The stability of the in-cell signal was verified by comparing the 1D HMQC spectrum recorded upon initiation of the experiments with that recorded at the end. In all cases, identical spectra suggested stable conditions (Fig. S1 H–J).

Cytoplasmic Environment Controls. Some of the macroscopic properties seen to affect quinary interactions, e.g., surface charge density, the electric dipole moment, and hydrophilicity^{Guy}, are intrinsically dependent on pH (46). Following the methodology in refs. 5 and 31, i.e., by comparing the in-cell ^{1}H chemical shifts of TTHA^{PWT} with in vitro NMR spectra at series of known pHs, we determined the pH of *E. coli* cytoplasm to be 6.7 ± 0.1. Moreover, in terms of chemical shifts, the in-cell spectra are superimposable with those of the lysate supernatants at pH = 6.5.

The observed retardation of molecular tumbling in the *E. coli* cytoplasm (mobility$^{in\ cell}$) is assumed to report on the surface properties of the proteins. However, it is conceivable that overexpression-induced differences in the cytoplasmic environment can also contribute to the observed effects. To pinpoint such putative morphological differences between *E. coli* overexpressing TTHA^{PWT}, HAH1^{PWT}, and SOD1^{barrel}, we examined the bacterial interior by transmission electron microscopy. Bacterial

cells overexpressing either TTHA^{PWT}, HAH1^{PWT}, or SOD1^{barrel} were prepared using the same protocol as for the NMR experiments. As a control, we included cells without overexpressing vector. The bacteria were fixated and contrast enhanced as described in *SI Materials and Methods*. All preparations showed nondisrupted cells with little variation in morphologies (Fig. S3 G–I). The cells displayed, generally, signs of chromatin assemblies, but without significant differences between the samples. The control bacteria with no overexpression, however, displayed slightly less condensed interior, whereas the number of ribosomes seemed up-regulated in the overexpressing cells (Fig. S3 G–I). On this basis, we conclude that the overall morphology of the cytoplasmic compartment remains similar across the experiments and is unlikely to explain the observed differences in mobility$^{in\ cell}$.

Protein Leakage Controls. To assure that the NMR signals stem from the overexpressed ^{15}N-labeled proteins in the *E. coli* cytoplasm, rather than from proteins that have leaked out into the intervening medium during sample preparation, handling, or acquisition, leakage controls were conducted. Each in-cell NMR sample was collected and gently centrifuged after the NMR experiment to obtain a cell-free supernatant of the intervening medium (sample^{supernatant}). Representative 1D HMQC leakage spectra of TTHA^{PWT}, HAH1^{PWT}, and SOD1^{barrel} are shown in Fig. S2 A–C. In all cases, the spectra of the sample^{supernatant} indicate small, but detectable, amounts of protein leakage (Fig. S2 A–C and Table S1). Data from samples with excessive leakage (>25%) were omitted from the analysis, and data from samples with protein leakage of >10% are annotated in boldface in Table S1. The latter data points are omitted in the subset denoted "RSS^{sub}," but are kept in the full analysis denoted "RSS^{all}." Interestingly, the data points from samples with indicated leakage do not systematically differ from those with a low degree of leakage (Figs. 4 and 6). Generally, protein leakage is expected to result in systematically too high values of mobility$^{in\ cell}$, because proteins that are freely moving outside the cells contribute to increase the NMR peak heights. Also, the level of leaked protein is expected to increase during the cause of the acquisition. As neither of these effects can be distinguished in the data sets, we conclude that the detected leakage occurs after the in-cell acquisition, i.e., during the preparation of the supernatant sample as a result of disturbing the analysis-stressed cells by pipetting, transfer, and centrifugation.

Protein Stability Controls. The assumption that changes in line width reflect retardation of rotational motion is only valid for folded proteins where the transversal relaxation is dominated by global Brownian motions, rather than local side-chain dynamics (33–35). This assumption holds only for rigid, folded proteins (36). To assure that the proteins are indeed fully folded during analysis, the thermodynamic stabilities of all studied variants were validated in both the cytosol and the cell lysates (5, 13, 73). First, the in vitro stabilities of the pseudo-wild-type proteins were determined by standard chevron data analysis, verifying that all three proteins are fully folded and show archetypical two-state behavior (Table S3 and Fig. S2 H and I). Chevron analysis was also used to verify folded structures of the TTHA and HAH1 mutations with the largest changes in mobility$^{in\ cell}$ (Fig. S2 H and I). For the rest of the 135 protein variants, the thermodynamic stability was verified directly from their respective in-cell NMR spectra (95% of which were 2D), where even small populations of unfolded material would be identified (Fig. S2 E–G). For eight variants of the SOD1^{barrel}, however, the stability was affected to such an extent that the unfolded state was significantly populated, and these were omitted from the data set. The omitted variants were SOD1$^{Q15E/I17E/N19E}$, SOD1$^{K9E/R85E/K92E}$, SOD1$^{R85E/K92E/K61E}$, SOD1$^{K9E/R85E/K92E/K61E/R100E}$, SOD1$^{K9E/K61E/R85E/K92E/K36E/R100E}$, SOD1$^{K9E/K23E/K36E/K61E/R85E/K92E/}$

^{R100E}, SOD1$^{K9E/K23E/K30E/K36E/K61E/R85E/K92E/R100E}$, and SOD1$^{K3E/K9E/K23E/K30E/K36E/K61E/R85E/K92E/R100E}$.

Line-Broadening Controls. Line broadening in NMR is not always due to retardation in global rotational dynamics; it can also result from chemical exchange on intermediate time scales (74). If such chemical exchange were to be the main cause of the in-cell line broadening, it would be biased to surface-exposed residues, as all of the variants in this study are fully folded in the *E. coli* cytoplasm (see above). The observed line broadening, however, is relatively uniformly distributed along the protein sequence, involving both the core and surface-exposed residues (Fig. S2*J*); this indicates that chemical exchange is not a major contributor to the in-cell line broadening.

Lysis Efficiency Control. To further validate that the NMR signal stems from cytoplasmic protein localization, we measured the NMR signal recovery following cell lysis by sonication (10×19 s, 1-s pulses, 40% amplitude). After cell lysis, we removed large portions of the cellular debris by centrifugation, keeping the total soluble protein concentration constant. This sample is denoted lysate^{supernatant} and was analyzed with the same experimental setup as for the sample$^{in\ cell}$. As the lysate^{supernatant} sample is used for quantification of the line broadening, it is crucial that the lysing protocol result in near-complete release of all cytosolic overexpressed protein. To test the lysate preparation protocol, we used Triton X-100 to further enhance cell disintegration. The sample$^{in\ cell}$ was divided into two equal parts, where one was treated with 1% (wt/wt) Triton X-100 to chemically disrupt the cell membrane (75), and the other was handled according to the original protocol. No difference in NMR signal intensity or spectral appearance of the lysate^{supernatant} from the two preparations was observed (Fig. S2*D*), suggesting that the sonication protocol results in near-full recovery of the cytosolic soluble protein material.

"Crowding Level" of the Lysate Supernatant. As cell lysis releases not only the target proteins but the entire *E. coli* content, it needs to be established that the line width of the lysate^{supernatant} sample is not skewed by quinary interactions with coreleased proteins. Accordingly, the concentration of endogenous *E. coli* proteins in the lysate^{supernatant} sample was determined to be 10.3 ± 1 mg/mL (76), using the Bicinchoninic acid assay kit (Sigma-Aldrich). The lysate supernatant is thus much more diluted than the >300 mg/mL *E. coli* cytoplasm (77) and should have negligible effect on the NMR line widths. To more directly test possible lysate effects on the NMR line broadening, we also compared the lysate^{supernatant}

spectrum with a spectrum recorded in dilute buffer (Fig. S1 *B–D*). No significant difference in line widths or spectral appearance was observed, indicating that neither interactions with endogenous protein nor self-interaction with the overexpressed protein affect the lysate supernatant sample line width.

Native Gel Mobility Control. To validate that the native gel mobility is not skewed by cellular contaminants, but mainly reflects the properties of the studied proteins, we compared native gel mobility of TTHA^{PWT}, HAH1^{PWT}, and SOD1^{barrel} in *E. coli* lysate and in aqueous solution. The experiment reports on whether the intrinsic properties of the analyzed proteins, i.e., hydrodynamic radius and surface charge density, are the main determinant of gel mobility, or if interactions with other, comigrating lysate components interfere with the mobility. The results show that proteins have indistinguishable mobility in aqueous solution and in *E. coli* lysate (Table S1 and Fig. S4*B*), indicating that the lysate supernatant sample yields unbiased results.

Standardized Regression Coefficients. Although the protein properties charge^{density}, hydrophilicity^{Guy}, hydrophobicity^{SASA}, and $||\boldsymbol{p}||$ correlate orderly with the mobility$^{in\ cell}$, their relative importance/weight is difficult to quantify in a nonbiased way. To get an indication of the response on mobility$^{in\ cell}$ by a given perturbation, we normalized the parameters as (78)

$$\chi_i^{norm} = \frac{\chi_i - \langle \chi_i \rangle}{\sigma_i}, \quad \text{[S3]}$$

where χ_i is any parameter i, $\langle \chi_i \rangle$ is the average value, and σ_i is the SD of the parameter. Then, by fitting a hyperdimensional plane to the observed mobility$^{in\ cell}$ using the normalized parameters as

$$\text{mobility}^{in\ cell} = \alpha_0 + \sum_{i=1}^{N} \alpha_i \chi_i^{norm}, \quad \text{[S4]}$$

we obtain $\alpha_i = \delta\text{mobility}^{in\ cell}/\delta\chi_i^{norm}$, which gives an indication of the response on a small perturbation in the dimension i. Using this approach, we determined α_i for all four parameters. The results show that charge^{density} has the strongest impact, with $\alpha_{charge} = 0.11 \pm 0.008$, and the other independent parameters are $\alpha_{Guy} = 0.03 \pm 0.008$, $\alpha_{SASA} = -0.06 \pm 0.008$, and $\alpha_{dipole} = -0.05 \pm 0.007$. We thus conclude that the charge^{density} is the main effector on the cytoplasmic mobility.
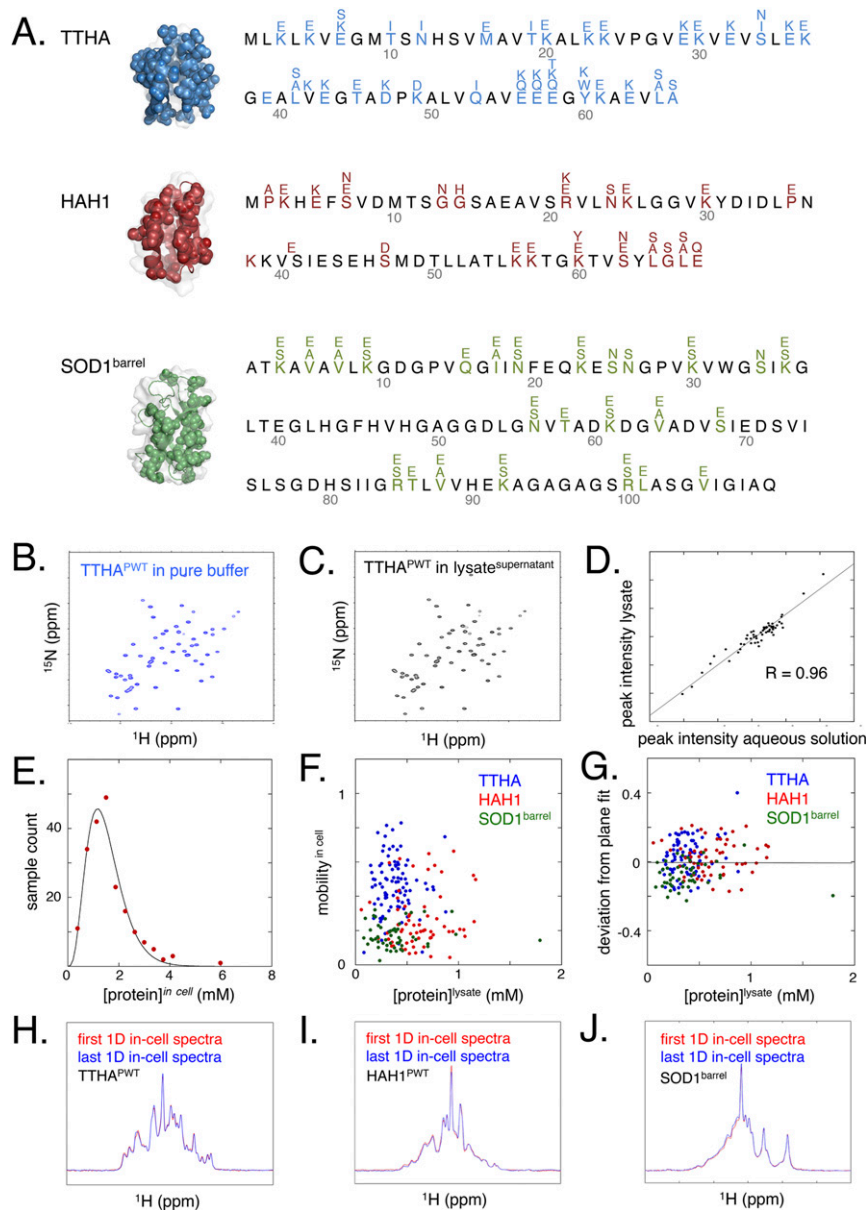
Fig. S1. (*A*) Structures, sequences, and mutations of the analyzed proteins TTHA [PDB 2ROE (28)], HAH1 [PDB 1TL4 (29)], and SOD1^barrel [PDB 4BCZ (30)]. (*B* and *C*) The HMQC spectrum of TTHA^PWT in (*B*) pure buffer is very similar to that in (*C*) lysate^supernatant. (*D*) The peak heights in dilute buffer and the lysate supernatant at the same protein concentration are very similar and correlate with $R = 0.96$, underlining that the lysate^supernatant conditions correspond to the "free" rotational tumbling observed in pure buffer. (*E*) Intracellular concentrations of overexpressed protein show a skewed normal distribution, with average 1.6 mM and SD ± 0.8 mM. (*F*) Although the intracellular concentration of overexpressed protein is relatively high at 1.6 ± 0.8 mM, there is no apparent correlation between expression levels and mobility^in cell. (*G*) Likewise, there is no correlation between expression levels and deviation from the fitted plane in Fig. 6. (*H–J*) Temporal stability of the 1D ^15N-filtered NMR spectra during the acquisition of mobility^in cell. The overlap of spectra obtained at the start and end of acquisition shows that the levels and properties of overexpressed protein remain the same during the course of the experiment.

Fig. S2. (*A–C*) Representative NMR spectra of (*A*) TTHA[PWT], (*B*) HAH1[PWT], and (*C*) SOD1[barrel] used for determining mobility[in cell] and leakage. The reduction in tumbling freedom was quantified by comparing the signal intensity from the sample[in cell] and the lysate[supernatant] sample, and the leakage was quantified by the intensity of the sample[supernatant]. (*D*) Overlaid lysate spectra obtained by sonication alone (red) and by chemical treatment with Triton X in combination with sonication (blue). The almost identical spectra show that sonication alone results in near-complete release of cytosolic soluble proteins. (*E–G*) Representative HMQC spectra for (*E*) TTHA[PWT], (*F*) HAH1[R21E], and (*G*) SOD1[R100E]. The spectra show single folded states in both the sample[in cell] (colored spectra) and the lysate[supernatant] (black). (*H* and *I*) Protein stability was further confirmed by chevron analysis for the pseudo-wild-type variants and selected surface mutants. (*J*) The line broadening, quantified by mobility[in cell], shows no systematic variation along the sequence or projected on the structure (not shown). The homogenous structural distribution suggests that the line broadening mainly stems from reduction in rotational correlation time. The dotted line corresponds to the determined mobility[in cell] value (Table S1).

Fig. S3. (*A*−*D*) Results and residuals from fits of data to linear or planar models. (*A* and *B*) Linear fit to the complete set of mobility$^{in\ cell}$ vs. charge^{density} data shows a strong systematic deviation in the residuals. In particular, the systematic deviation is between the three proteins, and only HAH1 data fall onto the line. (*C*) Increasing the dimensionality to three dimensions by adding hydrophilicity^{Guy} as a parameter improves the fit significantly, whereas (*D*) adding $||p||$ as an independent parameter in 4D only marginally improves the fit. (*E*) The corresponding distributions of residuals shown together with the best Gaussian fit, to highlight the narrowing in distribution widths by extending the fit to higher dimensions. (*F*−*I*) Morphological effects of *E. coli* induced by of overexpression. Transmission electron micrographs of *E. coli* cells overexpressing (*F*) TTHA^{PWT}, (*G*) HAH1^{PWT}, and (*H*) SOD1^{barrel}, compared with (*I*) nonoverexpressing cells. The general morphology is very similar for all overexpressing bacteria, and only very small differences in morphology are seen between the overexpressing and nonoverexpressing cells.
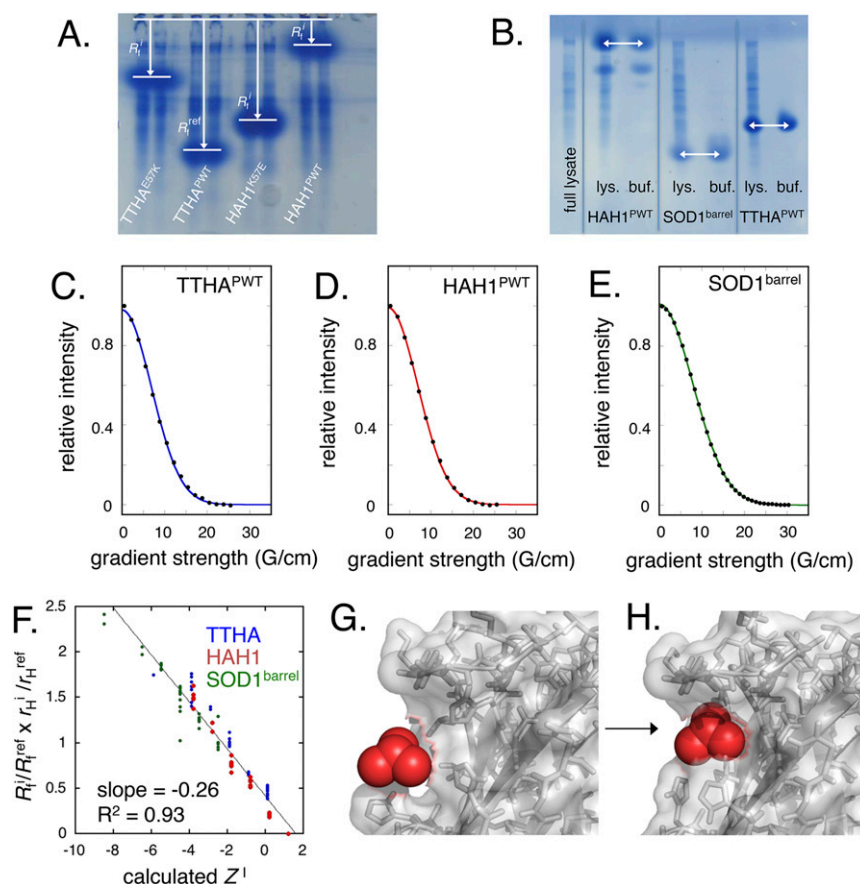
Fig. S4. (*A*) Native gel migration profiles, showing the variation in migration length due to single surface mutations. Surface charge density (Eq. **5**) is determined by dividing the migration length of the mutant (R_f^i) to the reference molecule TTHA^{PWT} (R_f^{ref}, lane 2), according to Eq. **4**. (*B*) Native gel migration was measured for each protein variant in lysate^{supernatant}, and comparison of migration in lysate and pure buffer shows no significant difference. The results show that, even in a more complex solution, the native gel mainly reports on surface charge density and hydrodynamic radius of the target proteins. (*C–E*) Signal attenuation profiles from PFG NMR diffusion determination of translational diffusion coefficient for (*C*) TTHA^{PWT}, (*D*) HAH1^{PWT}, and (*E*) SOD1^{barrel}. Data for SOD1^{barrel} are from ref. 32. All experiments were performed at 298 K in 100% D$_2$O. The fitted lines are least square fits for the Stejskal−Tanner equation with correction for nonlinear gradient field (66). (*F*) Plot of size-corrected electrophoretic mobility ($R_f^i/R_f^{ref} \times r_H^i/r_H^{ref} = Z^i/Z^{ref}$) vs. net charge calculated from model compound pK_A values (Z^i). (*G* and *H*) Docking of isobutene to the surface of SOD1^{barrel}. The simulation shows a substantial reorganization of the protein surface layer upon binding the isobutene moiety.

Table S1. Parameters of all mutational variants of TTHA, HAH1, and SOD1^{barrel}

Protein	mobility$^{in\ cell}$*	Leakage[†]	R_f^i/R_f^{ref}[‡]	charge^{density}[§]	hydrophobicity^{SASA}[¶]	hydrophilicity^{Guy}[#]	$\lVert p \rVert$[||]
TTHA^{PWT}	0.54 ± 0.02	0.02	1 ± 0	0.87	0.58	0.32	18.83
TTHA^{PWT}	0.5 ± 0.02	0.03	1 ± 0	0.87	0.58	0.32	18.83
TTHA$^{L41A/L65A}$	0.43 ± 0.01	0.01	1 ± 0	0.87	0.57	0.38	18.83
TTHA$^{L41A/L65A}$	0.51 ± 0.01	0.02	1 ± 0	0.87	0.57	0.38	18.83
TTHA^{E57Q}	0.41 ± 0.01	0.01	0.67 ± 0.01	0.54	0.58	0.33	18.81
TTHA^{E57Q}	0.5 ± 0	0.03	0.67 ± 0.02	0.54	0.58	0.33	18.81
TTHA^{L41S}	0.5 ± 0.01	0.05	1 ± 0	0.87	0.57	0.36	18.83
TTHA^{Y60W}	0.49 ± 0.02	0.02	0.96 ± 0.03	0.83	0.59	0.32	18.83
TTHA^{Y60W}	0.58 ± 0.01	0.04	0.96 ± 0.03	0.83	0.59	0.32	18.83
TTHA^{Y60W}	0.58 ± 0	0.01	0.96 ± 0.03	0.83	0.59	0.32	18.83
TTHA^{Y60K}	0.39 ± 0.03	0.04	0.69 ± 0.01	0.56	0.58	0.36	11.88
TTHA^{Y60K}	0.47 ± 0.01	0.02	0.69 ± 0.01	0.56	0.58	0.36	11.88
TTHA$^{L65S/A66S}$	0.43 ± 0.01	0	1.05 ± 0.05	0.92	0.58	0.37	18.83
TTHA$^{L65S/A66S}$	0.56 ± 0.01	0.04	1.05 ± 0.05	0.92	0.58	0.37	18.83
TTHA^{L41A}	0.54 ± 0.01	0.03	1.05 ± 0.05	0.92	0.58	0.35	18.83
TTHA^{L41A}	0.7 ± 0.01	0.02	1.05 ± 0.05	0.92	0.58	0.35	18.83
TTHA^{L65A}	0.69 ± 0.01	0.02	1.05 ± 0.05	0.92	0.57	0.35	18.83
TTHA^{L65A}	0.45 ± 0.01	0.01	1.05 ± 0.05	0.92	0.57	0.35	18.83
TTHA^{K37E}	0.64 ± 0.01	0.02	1.76 ± 0.19	1.63	0.56	0.31	38.55
TTHA^{K37E}	0.59 ± 0.01	0.08	1.76 ± 0.19	1.63	0.56	0.31	38.55
TTHA^{K24E}	0.69 ± 0.01	0.03	1.77 ± 0.13	1.64	0.57	0.31	14.39
TTHA^{K24E}	0.60 ± 0.01	0.02	1.77 ± 0.13	1.64	0.57	0.31	14.39
TTHA^{K24E}	0.68 ± 0.01	**0.12**	1.77 ± 0.13	1.64	0.57	0.31	14.39
TTHA^{K3E}	0.56 ± 0.01	0.06	1.59 ± 0.03	1.46	0.57	0.31	25.34
TTHA^{K3E}	0.49 ± 0.01	0.01	1.59 ± 0.03	1.46	0.57	0.31	25.34
TTHA^{K3E}	0.58 ± 0.01	0.04	1.59 ± 0.03	1.46	0.57	0.31	25.34
TTHA^{K20E}	0.79 ± 0.01	0	1.62 ± 0.02	1.49	0.57	0.31	17.23
TTHA^{K20E}	0.77 ± 0.01	**0.12**	1.62 ± 0.02	1.49	0.57	0.31	17.23
TTHA^{K23E}	0.75 ± 0.01	0.01	1.68 ± 0.06	1.55	0.57	0.31	8.44
TTHA^{K23E}	0.67 ± 0.01	0.08	1.68 ± 0.06	1.55	0.57	0.31	8.44
TTHA^{K23E}	0.67 ± 0.01	0.05	1.68 ± 0.06	1.55	0.57	0.31	8.4
TTHA^{M16E}	0.83 ± 0.01	0.01	1.4 ± 0.15	1.27	0.56	0.37	20.98
TTHA^{M16E}	0.81 ± 0.01	0.07	1.4 ± 0.15	1.27	0.56	0.37	20.98
TTHA^{T45E}	0.61 ± 0.01	0.01	1.37 ± 0.07	1.27	0.57	0.34	10.83
TTHA^{T45E}	0.53 ± 0.01	0.04	1.37 ± 0.07	1.24	0.57	0.34	10.83
TTHA^{E58K}	0.18 ± 0.01	0.05	0.45 ± 0.01	0.32	0.58	0.34	21.62
TTHA^{E58K}	0.14 ± 0.03	0	0.45 ± 0.01	0.32	0.58	0.34	21.62
TTHA^{E58K}	0.12 ± 0.04	0.04	0.45 ± 0.01	0.32	0.58	0.34	21.62
TTHA^{E7S}	0.48 ± 0.02	0.02	0.68 ± 0.06	0.55	0.58	0.32	8.66
TTHA^{E7S}	0.36 ± 0.01	0.04	0.68 ± 0.06	0.55	0.58	0.32	8.66
TTHA$^{Y60W/E57K}$	0.28 ± 0.02	0.01	0.42 ± 0.03	0.29	0.59	0.33	24.27
TTHA$^{Y60W/E57K}$	0.27 ± 0.02	0.03	0.42 ± 0.03	0.29	0.59	0.33	24.27
TTHA^{E32K}	0.19 ± 0.01	0.02	0.54 ± 0.09	0.41	0.58	0.34	29.7
TTHA^{E32K}	0.25 ± 0.01	0.02	0.54 ± 0.09	0.41	0.58	0.34	29.7
TTHA^{D47K}	0.34 ± 0.01	0.01	0.4 ± 0.03	0.27	0.58	0.34	36.72
TTHA^{D47K}	0.35 ± 0.01	0.03	0.4 ± 0.03	0.27	0.58	0.34	36.72
TTHA^{D47K}	0.34 ± 0.01	0.04	0.4 ± 0.03	0.27	0.58	0.34	36.72
TTHA^{E29K}	0.25 ± 0.01	0.02	0.51 ± 0.03	0.38	0.58	0.34	43.72
TTHA^{E29K}	0.36 ± 0.01	0.03	0.51 ± 0.03	0.38	0.58	0.34	43.72
TTHA^{K61E}	0.45 ± 0.01	0.01	1.62 ± 0.12	1.49	0.57	0.31	37.9
TTHA^{K61E}	0.47 ± 0.01	0.03	1.62 ± 0.12	1.49	0.57	0.31	37.9
TTHA^{K49D}	0.5 ± 0.01	0.03	1.73 ± 0.16	1.60	0.58	0.31	27.13
TTHA^{K49D}	0.58 ± 0.02	0.05	1.73 ± 0.16	1.60	0.58	0.31	27.13
TTHA^{E43K}	0.27 ± 0.01	0.06	0.5 ± 0.08	0.37	0.58	0.34	39.12
TTHA^{E43K}	0.36 ± 0.03	0.02	0.5 ± 0.08	0.37	0.58	0.34	39.12
TTHA^{E56K}	0.46 ± 0.01	0.05	0.44 ± 0.02	0.31	0.58	0.34	17.8
TTHA^{E56K}	0.43 ± 0.01	0.03	0.44 ± 0.02	0.31	0.58	0.34	17.8
TTHA$^{E36k.}$	0.32 ± 0.03	0	0.43 ± 0.03	0.30	0.57	0.34	22.13
TTHA^{E36K}	0.27 ± 0.01	0.03	0.43 ± 0.03	0.30	0.57	0.34	22.13
TTHA^{E7K}	0.28 ± 0.02	0.03	0.39 ± 0.03	0.26	0.58	0.34	2.37
TTHA^{E7K}	0.39 ± 0.01	0.01	0.39 ± 0.03	0.26	0.58	0.34	2.37
TTHA^{E39K}	0.39 ± 0.01	0.02	0.44 ± 0.03	0.31	0.58	0.34	17.01
TTHA^{E39K}	0.42 ± 0.01	0.03	0.44 ± 0.03	0.31	0.58	0.34	17.01

Protein	mobility$^{\text{in cell}}$*	Leakage†	R_f^i/R_f^{ref}‡	charge$^{\text{density}}$§	hydrophobicity$^{\text{SASA}}$¶	hydrophilicity$^{\text{Guy}}$#	$\lVert p \rVert$$^{\lVert}$
TTHA$^{\text{E63K}}$	0.41 ± 0.01	0.02	0.5 ± 0.05	0.37	0.59	0.34	15.95
TTHA$^{\text{E63K}}$	0.44 ± 0.01	0.02	0.5 ± 0.05	0.37	0.59	0.34	15.95
TTHA$^{\text{K30E/K61E}}$	0.58 ± 0.01	0.01	1.75 ± 0.05	1.62	0.57	0.3	17.27
TTHA$^{\text{K30E/K61E}}$	0.58 ± 0.02	0.02	1.75 ± 0.05	1.62	0.57	0.3	17.27
TTHA$^{\text{K5E}}$	0.48 ± 0.01	0.01	1.45 ± 0.03	1.32	0.57	0.31	32.28
TTHA$^{\text{K5E}}$	0.41 ± 0.01	0.01	1.45 ± 0.03	1.32	0.57	0.31	32.28
TTHA$^{\text{K30E}}$	0.6 ± 0.01	0.04	1.41 ± 0.12	1.28	0.57	0.31	8.98
TTHA$^{\text{K30E}}$	0.59 ± 0.01	**0.13**	1.41 ± 0.12	1.28	0.57	0.31	8.98
TTHA$^{\text{Y60W/E57Q}}$	0.48 ± 0.01	0.06	0.65 ± 0.02	0.52	0.58	0.32	18.81
TTHA$^{\text{Y60W/E57Q}}$	0.41 ± 0.01	0.02	0.65 ± 0.02	0.52	0.58	0.32	18.81
TTHA$^{\text{Y60W/E57Q}}$	0.50 ± 0.01	0.06	0.65 ± 0.02	0.52	0.58	0.32	18.81
TTHA$^{\text{Y60W/E58K}}$	0.28 ± 0.04	**0.16**	0.65 ± 0.02	0.52	0.58	0.32	17.49
TTHA$^{\text{Y60W/E56Q}}$	0.34 ± 0.01	0.03	0.65 ± 0.02	0.52	0.58	0.32	15.84
TTHA$^{\text{E56Q/E58Q}}$	0.22 ± 0.01	0.04	0.44 ± 0.02	0.31	0.58	0.33	19.06
TTHA$^{\text{S34I}}$	0.75 ± 0	**0.12**	0.98 ± 0.01	0.85	0.57	0.29	18.83
TTHA$^{\text{N12I}}$	0.31 ± 0.01	0.1	1.03 ± 0.03	0.90	0.58	0.29	18.83
TTHA$^{\text{T19I}}$	0.37 ± 0.01	0.03	0.98 ± 0	0.85	0.57	0.3	18.83
TTHA$^{\text{Q53I}}$	0.38 ± 0.01	0.02	0.98 ± 0.01	0.85	0.59	0.28	18.83
TTHA$^{\text{T10I}}$	0.33 ± 0.01	0.05	1.02 ± 0.02	0.89	0.58	0.3	18.83
TTHA$^{\text{T45I}}$	0.44 ± 0.01	0.02	1.12 ± 0.09	0.99	0.59	0.3	18.83
TTHA$^{\text{E57K}}$	0.44 ± 0.01	0.01	0.47 ± 0.02	0.34	0.58	0.34	24.27
TTHA$^{\text{E58Q}}$	0.45 ± 0.01	0.03	0.68 ± 0.01	0.55	0.58	0.33	17.49
TTHA$^{\text{S34N}}$	0.51 ± 0.01	0.06	1.12 ± 0.09	0.99	0.56	0.32	18.83
TTHA$^{\text{E56Q}}$	0.51 ± 0.01	0.07	0.68 ± 0.01	0.55	0.58	0.33	15.85
TTHA$^{\text{L41S/L65S/A66S}}$	0.42 ± 0.01	0.02	1 ± 0	0.87	NA	0.4	18.83
TTHA$^{\text{L41S/L65S/A66S}}$	0.49 ± 0.01	0.01	1 ± 0	0.87	NA	0.4	18.83
TTHA$^{\text{E56K/E57K/E58T}}$	0.08 ± 0.07	**0.13**	0	−0.13	NA	0.33	44.62
TTHA$^{\text{E56K/E57K/E58T}}$	0.07 ± 0.02	**0.16**	0	−0.13	NA	0.33	44.62
HAH1$^{\text{PWT}}$	0.11 ± 0.03	0.08	0.18 ± 0.01	−0.07	0.58	0.31	54.26
HAH1$^{\text{PWT}}$	0.16 ± 0.05	0.02	0.18 ± 0.01	−0.07	0.58	0.31	54.26
HAH1$^{\text{P36E/K38E}}$	0.2 ± 0.01	0.09	1.1 ± 0.1	0.82	0.58	0.3	69.09
HAH1$^{\text{P36E/K38E}}$	0.2 ± 0.01	**0.16**	1.1 ± 0.1	0.82	0.58	0.3	69.09
HAH1$^{\text{K60E/S63E}}$	0.29 ± 0.01	0.03	1.2 ± 0	0.92	0.58	0.3	42.09
HAH1$^{\text{R21E/K25E}}$	0.66 ± 0.01	0	1.6 ± 0.1	1.31	0.57	0.27	43.45
HAH1$^{\text{R21E/K25E}}$	0.65 ± 0	**0.18**	1.6 ± 0.1	1.31	0.57	0.27	43.45
HAH1$^{\text{K60E/K56E}}$	0.27 ± 0.02	0.05	1.47 ± 0.04	1.18	0.56	0.28	31.42
HAH1$^{\text{K60E/K56E}}$	0.34 ± 0.01	**0.15**	1.47 ± 0.04	1.18	0.56	0.28	31.42
HAH1$^{\text{K60E/K57E}}$	0.58 ± 0	0.08	1.5 ± 0.1	1.21	0.56	0.28	24.07
HAH1$^{\text{K60E/K57E}}$	0.54 ± 0	**0.19**	1.5 ± 0.1	1.21	0.56	0.28	24.07
HAH1$^{\text{N24S}}$	0.16 ± 0.01	**0.14**	0.22 ± 0	−0.04	0.58	0.31	54.26
HAH1$^{\text{N24S}}$	0.11 ± 0.01	**0.12**	0.22 ± 0	−0.04	0.58	0.31	54.26
HAH1$^{\text{R21K}}$	0.18 ± 0.02	0.04	0.22 ± 0.02	−0.04	0.58	0.3	54.26
HAH1$^{\text{R21K}}$	0.15 ± 0.01	0.07	0.22 ± 0.02	−0.04	0.58	0.3	54.26
HAH1$^{\text{G13N}}$	0.09 ± 0.03	0.05	0.2 ± 0.02	−0.06	0.58	0.31	54.26
HAH1$^{\text{G13N}}$	0.09 ± 0.03	0	0.2 ± 0.02	−0.06	0.58	0.31	54.26
HAH1$^{\text{G14H}}$	0.05 ± 0.04	**0.22**	0.29 ± 0.01	0.03	0.58	0.29	54.26
HAH1$^{\text{G14H}}$	0.07 ± 0.02	**0.22**	0.29 ± 0.01	0.03	0.58	0.29	54.26
HAH1$^{\text{P2A}}$	0.26 ± 0.01	0.05	0.23 ± 0.01	−0.03	0.56	0.29	54.26
HAH1$^{\text{P2A}}$	0.20 ± 0.01	0.04	0.23 ± 0.01	−0.03	0.56	0.29	54.26
HAH1$^{\text{L65A/L67A}}$	0.28 ± 0	0.08	0.22 ± 0.02	−0.04	0.58	0.35	54.26
HAH1$^{\text{L65A/L67A}}$	0.25 ± 0.01	0.04	0.22 ± 0.02	−0.04	0.58	0.35	54.26
HAH1$^{\text{L65S/L67S}}$	0.21 ± 0.02	0.02	0.22 ± 0.02	−0.04	0.57	0.37	54.26
HAH1$^{\text{L65S/L67S}}$	0.2 ± 0.01	0.07	0.22 ± 0.02	−0.04	0.57	0.37	54.26
HAH1$^{\text{R21E/K30E}}$	0.58 ± 0.01	**0.12**	1.45 ± 0.05	1.16	0.57	0.27	59.3
HAH1$^{\text{R21E/K30E}}$	0.62 ± 0	0.07	1.45 ± 0.05	1.16	0.57	0.27	59.3
HAH1$^{\text{R21E/K30E}}$	0.5 ± 0.01	0.01	1.45 ± 0.05	1.16	0.57	0.27	59.3
HAH1$^{\text{R21E/K38E}}$	0.56 ± 0	0.07	1.5 ± 0	1.21	0.59	0.27	48.48
HAH1$^{\text{R21E/K38E}}$	0.52 ± 0.01	0.02	1.5 ± 0	1.21	0.59	0.27	48.48
HAH1$^{\text{S63N}}$	0.13 ± 0.03	0.08	0.22 ± 0.02	−0.04	0.58	0.3	54.26
HAH1$^{\text{K3E}}$	0.32 ± 0.05	**0.11**	0.66 ± 0.01	0.39	0.58	0.29	76.39
HAH1$^{\text{K60Y}}$	0.21 ± 0.01	0.05	0.61 ± 0.01	0.34	0.57	0.27	46.24
HAH1$^{\text{P36E}}$	0.19 ± 0.01	0.04	0.58 ± 0	0.32	0.56	0.31	58.48

Protein	mobility$^{\text{in cell}}$*	Leakage†	$R_\text{f}^i/R_\text{f}^\text{ref}$‡	charge$^{\text{density}}$§	hydrophobicity$^{\text{SASA}}$¶	hydrophilicity$^{\text{Guy}}$#	$\lVert p \rVert$‖
HAH1^{P36E}	0.14 ± 0.01	0.04	0.58 ± 0	0.32	0.56	0.31	58.48
HAH1^{S63E}	0.23 ± 0.01	0.04	0.5 ± 0.03	0.24	0.58	0.31	53.99
HAH1^{S63E}	0.2 ± 0.01	0.04	0.5 ± 0.03	0.24	0.58	0.31	54
HAH1^{S7E}	0.21 ± 0.01	0.01	0.53 ± 0.04	0.26	0.58	0.31	56.51
HAH1^{S7E}	0.19 ± 0.02	0.05	0.53 ± 0.04	0.26	0.58	0.31	56.51
HAH1^{S47D}	0.13 ± 0.02	0.06	0.57 ± 0.03	0.30	0.57	0.31	59.84
HAH1^{S47D}	0.11 ± 0.03	0.06	0.57 ± 0.03	0.30	0.57	0.31	59.84
HAH1^{K57E}	0.18 ± 0.02	**0.12**	0.75 ± 0.01	0.48	0.56	0.29	35.5
HAH1^{K57E}	0.37 ± 0.01	0.06	0.75 ± 0.01	0.48	0.56	0.29	35.5
HAH1^{K57E}	0.39 ± 0.01	0.06	0.75 ± 0.01	0.48	0.56	0.29	35.5
HAH1^{K56E}	0.23 ± 0.01	0.06	0.73 ± 0.02	0.46	0.56	0.29	42.5
HAH1^{K56E}	0.23 ± 0.02	0.04	0.73 ± 0.02	0.46	0.56	0.29	42.5
HAH1^{K56E}	0.29 ± 0.01	0.03	0.73 ± 0.02	0.46	0.56	0.29	42.5
HAH1$^{K57E/K56E}$	0.6 ± 0.01	**0.12**	1.35 ± 0.05	1.07	0.56	0.28	30.33
HAH1$^{K57E/K56E}$	0.43 ± 0.01	0.04	1.35 ± 0.05	1.07	0.56	0.28	30.33
HAH1^{K60E}	0.24 ± 0.02	0.08	0.85 ± 0.05	0.58	0.57	0.29	39.51
HAH1^{K60E}	0.19 ± 0.02	0.07	0.85 ± 0.05	0.58	0.57	0.29	39.51
HAH1^{K38E}	0.22 ± 0.02	0.03	0.74 ± 0.02	0.47	0.59	0.29	60.59
HAH1^{R21E}	0.4 ± 0.01	0.05	0.77 ± 0.05	0.50	0.58	0.28	43.66
HAH1^{K30E}	0.16 ± 0.01	**0.19**	0.73 ± 0.07	0.46	0.56	0.29	66.16
HAH1^{K30E}	0.16 ± 0.02	0.04	0.73 ± 0.07	0.46	0.56	0.29	66.16
HAH1^{K25E}	0.32 ± 0.01	0.04	0.85 ± 0.01	0.58	0.57	0.29	46.17
HAH1^{K25E}	0.34 ± 0.02	0.05	0.85 ± 0.01	0.58	0.57	0.29	46.17
HAH1^{S7N}	0.17 ± 0.01	0.08	0.2 ± 0.03	−0.06	0.58	0.30	54.26
HAH1^{E68Q}	0.11 ± 0.02	0.08	0**	−0.25	0.57	0.31	40.73
HAH1^{E5K}	0.04 ± 0.02	**0.12**	0**	−0.25	0.59	0.31	41.72
HAH1$^{L65S/G66S/L67S}$	0.2 ± 0.01	0.09	0.21 ± 0.04	−0.05	NA	0.37	54.26
SOD1^{barrel}	0.14 ± 0.01	0.01	0.79 ± 0.01	0.30	0.60	0.29	45.66
SOD1^{barrel}	0.11 ± 0.02	0.05	0.79 ± 0.01	0.30	0.60	0.29	45.66
SOD1^{R100E}	0.31 ± 0.01	0.05	1.22 ± 0.01	0.63	0.60	0.28	41.03
SOD1^{K3E}	0.15 ± 0.01	0.09	1.22 ± 0.02	0.63	0.61	0.29	50.82
SOD1$^{K92E/R85E}$	0.3 ± 0.01	**0.16**	1.53 ± 0.03	0.88	0.6	0.27	40.03
SOD1$^{K92E/R85E}$	0.31 ± 0.01	**0.17**	1.53 ± 0.03	0.88	0.6	0.27	40.03
SOD1^{R85E}	0.21 ± 0.01	**0.13**	1.21 ± 0	0.63	0.61	0.28	36.47
SOD1^{K92E}	0.14 ± 0.02	**0.12**	1.21 ± 0.03	0.63	0.60	0.29	53.13
SOD1^{K92E}	0.15 ± 0.02	0.2	1.21 ± 0.03	0.63	0.60	0.29	53.13
SOD1^{K92S}	0.15 ± 0.02	0.05	0.98 ± 0.01	0.45	0.6	0.28	47.08
SOD1^{K23S}	0.12 ± 0.01	0.03	1 ± 0.01	0.46	0.61	0.28	54.34
SOD1^{K23S}	0.08 ± 0.01	**0.19**	1 ± 0.01	0.46	0.61	0.28	54.34
SOD1^{K30S}	0.08 ± 0.02	0.07	0.99 ± 0.01	0.45	0.61	0.28	53.02
SOD1^{K30S}	0.09 ± 0.02	**0.17**	0.99 ± 0.01	0.45	0.61	0.28	53.02
SOD1^{K36E}	0.13 ± 0.02	**0.12**	1.24 ± 0.06	0.65	0.6	0.29	48.31
SOD1^{R100S}	0.2 ± 0.01	0.1	0.91 ± 0.06	0.39	0.6	0.28	40.7
SOD1$^{K9E/K36E}$	0.24 ± 0.01	0.1	1.62 ± 0.07	0.98	0.6	0.28	43.12
SOD1^{R85S}	0.16 ± 0.01	0.1	0.98 ± 0	0.45	0.6	0.28	39.93
SOD1^{R85S}	0.15 ± 0.02	0.07	0.98 ± 0	0.45	0.61	0.28	39.93
SOD1$^{K92E/R100E}$	0.28 ± 0.02	0.06	1.56 ± 0.06	0.90	0.6	0.27	63.91
SOD1^{K3S}	0.17 ± 0.02	0.08	0.99 ± 0.01	0.45	0.61	0.28	46.31
SOD1^{K61E}	0.09 ± 0.02	**0.14**	1.23 ± 0.05	0.64	0.61	0.29	68.31
SOD1$^{K92S/R100S}$	0.27 ± 0.01	0.08	1.26 ± 0.03	0.67	0.6	0.26	46.97
SOD1^{K9S}	0.21 ± 0.01	0.1	1 ± 0.09	0.46	0.6	0.28	36.52
SOD1^{K9S}	0.11 ± 0.01	**0.24**	1 ± 0.09	0.46	0.6	0.28	36.52
SOD1^{K36S}	0.13 ± 0.02	0.07	1.04 ± 0.05	0.49	0.60	0.28	45.32
SOD1$^{V5E/V7E}$	0.32 ± 0.02	0.02	1.26 ± 0.03	0.67	0.59	0.35	35.42
SOD1$^{Q15E/N19E}$	0.22 ± 0.02	0.02	1.28 ± 0.03	0.68	0.6	0.3	40.59
SOD1$^{N56E/T58E}$	0.23 ± 0.02	0.03	1.26 ± 0.02	0.67	0.60	0.31	61.71
SOD1$^{L101E/V105E}$	0.1 ± 0.01	0.06	0.81 ± 0.01	0.31	0.59	0.35	31.13
SOD1^{S34N}	0.16 ± 0.02	0.08	1.02 ± 0.02	0.47	0.60	0.29	45.66
SOD1$^{K36E/K61E}$	0.15 ± 0.02	**0.12**	1.65 ± 0.02	0.97	0.6	0.28	81.93
SOD1$^{K92E/R100S}$	0.28 ± 0.02	0.03	1.46 ± 0.01	0.82	0.60	0.27	56.76
SOD1$^{K92E/R85S}$	0.27 ± 0.04	0.06	1.44 ± 0.02	0.81	0.6	0.27	45.87
SOD1$^{K9S/K36E}$	0.14 ± 0.04	0.06	1.48 ± 0.04	0.84	0.6	0.27	44.33

Table S1. Cont.

Protein	mobility$^{\text{in cell}}$*	Leakage†	$R_f^i/R_f^{ref‡}$	charge$^{\text{density§}}$	hydrophobicity$^{\text{SASA¶}}$	hydrophilicity$^{\text{Guy\#}}$	$\|\boldsymbol{p}\|^{\|}$
SOD1^{N19S}	0.03 ± 0.04	0.2	0.74 ± 0.05	0.26	0.61	0.29	45.66
SOD1^{K23E}	0.13 ± 0.01	0.07	1.1 ± 0	0.54	0.61	0.29	67.21
SOD1^{K30E}	0.08 ± 0.03	0.1	1.06 ± 0	0.51	0.60	0.29	61.6
SOD1^{V64A}	0.11 ± 0.02	**0.2**	0.75 ± 0.07	0.27	0.6	0.31	45.66
SOD1^{K61S}	0.12 ± 0.02	**0.2**	0.93 ± 0.04	0.41	0.6	0.28	54.5
SOD1^{K9E}	0.19 ± 0.02	0.1	1.16 ± 0.08	0.59	0.6	0.29	29.15
SOD1^{I17A}	0.16 ± 0.02	**0.19**	0.74 ± 0.03	0.26	0.6	0.31	45.66
SOD1^{V5AV7A}	0.10 ± 0.03	0.09	0.78 ± 0.05	0.29	0.62	0.33	45.66
SOD1^{N26S}	0.2 ± 0.01	0.06	0.75 ± 0.05	0.27	0.61	0.29	45.66
SOD1$^{K92S/R85E}$	0.31 ± 0	**0.15**	1.2 ± 0	0.62	0.6	0.27	35.09
SOD1$^{V64E/S68E}$	0.14 ± 0	**0.22**	1.1 ± 0	0.54	0.6	0.33	58.69
SOD1$^{T86E/V88E}$	0.20 ± 0.03	**0.15**	1.06 ± 0.03	0.51	0.6	0.33	36.18
SOD1^{N56S}	0.13 ± 0.02	0.1	0.78 ± 0.05	0.29	NA	0.29	45.66
SOD1^{S25N}	0.2 ± 0.01	0.06	0.73 ± 0.04	0.25	NA	0.29	45.66
SOD1^{V88A}	0.16 ± 0.02	**0.2**	0.69 ± 0.04	0.22	NA	0.31	45.66
SOD1$^{Q15E/I17E/N19E}$	0.18 ± 0.01	0.08	1.42 ± 0.01	0.79	NA	0.32	40.26

*Calculated as the ratio of in-cell peak heights and lysate peak heights. The error is estimated from the noise-to-signal ratio of the NMR spectra.

†Defined by the ratio of supernatant peak intensity to in-cell peak intensity. Data points with protein leakage of >10% are highlighted in boldface.

‡Reference molecule is for all values TTHA^{PWT}.

§From relative mobility scaled by hydrodynamic radii, Eq. **5**.

¶From in silico mutagenesis and hydrophobicity calculations (*SI Materials and Methods*). NA, data not available.

$^\#$Calculated from Eq. **7**.

$^{\|}$Calculated from the PDB files and Eq. **8**.

**HAH1^{E68Q} and HAH1^{E5K} do not migrate at all on the native gel.

Table S2. *E. coli* expression levels

[protein]^{lysate}	Millimolars	[protein]^{lysate}	Millimolars	[protein]^{lysate}	Millimolars	[protein]^{lysate}	Millimolars
TTHA^{PWT}	0.25	TTHA^{K49D}	0.17	HAH1^{N24S}	0.72	SOD1^{barrel}	0.45
TTHA^{PWT}	0.19	TTHA^{K49D}	0.39	HAH1^{N24S}	0.66	SOD1^{R100E}	0.54
TTHA$^{L41A/L65A}$	0.49	TTHA^{E43K}	0.26	HAH1^{R21K}	0.75	SOD1^{K3E}	0.58
TTHA$^{L41A/L65A}$	0.41	TTHA^{E43K}	0.45	HAH1^{R21K}	0.32	SOD1$^{K92E/R85E}$	0.49
TTHA^{E57Q}	0.29	TTHA^{E56K}	0.75	HAH1^{G13N}	0.45	SOD1$^{K92E/R85E}$	0.67
TTHA^{E57Q}	0.30	TTHA^{E56K}	0.53	HAH1^{G13N}	0.48	SOD1^{R85E}	0.38
TTHA^{L41S}	0.44	TTHA$^{E36k.}$	0.32	HAH1^{G14H}	0.73	SOD1^{K92E}	0.22
TTHA^{Y60W}	0.41	TTHA^{E36K}	0.18	HAH1^{G14H}	0.52	SOD1^{K92E}	0.16
TTHA^{Y60W}	0.39	TTHA^{E7K}	0.50	HAH1^{P2A}	0.52	SOD1^{K92S}	0.15
TTHA^{Y60W}	0.33	TTHA^{E7K}	0.30	HAH1^{P2A}	0.67	SOD1^{K23S}	0.70
TTHA^{Y60K}	0.22	TTHA^{E39K}	0.29	HAH1$^{L65A/L67A}$	0.52	SOD1^{K23S}	0.46
TTHA^{Y60K}	0.19	TTHA^{E39K}	0.29	HAH1$^{L65A/L67A}$	1.15	SOD1^{K30S}	0.34
TTHA$^{L65S/A66S}$	0.40	TTHA^{E63K}	0.37	HAH1$^{L65S/L67S}$	1.01	SOD1^{K30S}	0.44
TTHA$^{L65S/A66S}$	0.36	TTHA^{E63K}	0.43	HAH1$^{L65S/L67S}$	0.44	SOD1^{K36E}	0.34
TTHA^{L41A}	0.16	TTHA$^{K30E/K61E}$	0.15	HAH1$^{R21E/K30E}$	0.56	SOD1^{R100S}	0.34
TTHA^{L41A}	0.27	TTHA$^{K30E/K61E}$	0.23	HAH1$^{R21E/K30E}$	0.43	SOD1$^{K9E/K36E}$	0.43
TTHA^{L65A}	0.44	TTHA^{K5E}	0.49	HAH1$^{R21E/K30E}$	1.17	SOD1^{R85S}	0.56
TTHA^{L65A}	0.16	TTHA^{K5E}	0.30	HAH1$^{R21E/K38E}$	0.70	SOD1^{R85S}	0.29
TTHA^{K37E}	0.32	TTHA^{K30E}	0.41	HAH1$^{R21E/K38E}$	1.14	SOD1$^{K92E/R100E}$	0.19
TTHA^{K37E}	0.42	TTHA^{K30E}	0.25	HAH1^{S63N}	0.50	SOD1^{K3S}	0.34
TTHA^{K24E}	0.36	TTHA$^{Y60W/E57Q}$	0.68	HAH1^{K3E}	0.06	SOD1^{K61E}	0.30
TTHA^{K24E}	0.38	TTHA$^{Y60W/E57Q}$	0.48	HAH1^{K60Y}	0.45	SOD1$^{K92S/R100S}$	0.29
TTHA^{K24E}	0.33	TTHA$^{Y60W/E57Q}$	0.48	HAH1^{P36E}	0.59	SOD1^{K9S}	0.41
TTHA^{K3E}	0.41	TTHA$^{Y60W/E58K}$	0.19	HAH1^{P36E}	0.55	SOD1^{K9S}	0.70
TTHA^{K3E}	0.59	TTHA$^{Y60W/E56Q}$	0.41	HAH1^{S63E}	0.57	SOD1^{K36S}	0.42
TTHA^{K3E}	0.49	TTHA$^{E56Q/E58Q}$	0.62	HAH1^{S63E}	0.94	SOD1$^{V5E/V7E}$	0.19
TTHA^{K20E}	0.36	TTHA^{S34I}	0.87	HAH1^{S7E}	1.05	SOD1$^{Q15E/N19E}$	0.24
TTHA^{K20E}	0.22	TTHA^{N12I}	0.87	HAH1^{S7E}	0.31	SOD1$^{N56E/T58E}$	0.22
TTHA^{K23E}	0.29	TTHA^{T19I}	0.41	HAH1^{S47D}	0.35	SOD1$^{L101E/V105E}$	0.38
TTHA^{K23E}	0.41	TTHA^{Q53I}	0.51	HAH1^{S47D}	0.35	SOD1^{S34N}	0.22
TTHA^{K23E}	0.26	TTHA^{T10I}	0.47	HAH1^{K57E}	0.41	SOD1$^{K36E/K61E}$	0.19
TTHA^{M16E}	0.44	TTHA^{T45I}	0.76	HAH1^{K57E}	0.53	SOD1$^{K92E/R100S}$	0.17
TTHA^{M16E}	0.30	TTHA^{E57K}	0.70	HAH1^{K57E}	0.14	SOD1$^{K92E/R85S}$	0.09
TTHA^{T45E}	0.51	TTHA^{E58Q}	0.46	HAH1^{K56E}	0.41	SOD1$^{K9S/K36E}$	0.08
TTHA^{T45E}	0.46	TTHA^{S34N}	0.37	HAH1^{K56E}	0.62	SOD1^{N19S}	0.23
TTHA^{E58K}	0.29	TTHA^{E56Q}	0.29	HAH1^{K56E}	0.70	SOD1^{K23E}	0.40
TTHA^{E58K}	0.25	TTHA$^{L41S/L65S/A66S}$	0.43	HAH1$^{K57E/K56E}$	0.33	SOD1^{K30E}	0.39
TTHA^{E58K}	0.42	TTHA$^{L41S/L65S/A66S}$	0.35	HAH1$^{K57E/K56E}$	0.42	SOD1^{V64A}	0.30
TTHA^{E7S}	0.39	TTHA$^{E56K/E57K/E58T}$	0.65	HAH1^{K60E}	0.49	SOD1^{K61S}	0.12
TTHA^{E7S}	0.18	TTHA$^{E56K/E57K/E58T}$	0.08	HAH1^{K60E}	0.73	SOD1^{K9E}	0.33
TTHA$^{Y60W/E57K}$	0.26	HAH1^{PWT}	0.37	HAH1^{K38E}	0.86	SOD1^{I17A}	0.42
TTHA$^{Y60W/E57K}$	0.24	HAH1^{PWT}	0.22	HAH1^{R21E}	0.82	SOD1^{V5AV7A}	0.22
TTHA^{E32K}	0.32	HAH1$^{P36E/K38E}$	0.63	HAH1^{K30E}	0.61	SOD1^{N26S}	0.39
TTHA^{E32K}	0.45	HAH1$^{P36E/K38E}$	0.91	HAH1^{K30E}	0.73	SOD1$^{K92S/R85E}$	0.93
TTHA^{D47K}	0.43	HAH1$^{K60E/S63E}$	0.52	HAH1^{K25E}	0.41	SOD1$^{V64E/S68E}$	1.79
TTHA^{D47K}	0.25	HAH1$^{R21E/K25E}$	0.95	HAH1^{K25E}	0.67	SOD1$^{T86E/V88E}$	0.14
TTHA^{D47K}	0.34	HAH1$^{R21E/K25E}$	0.71	HAH1^{S7N}	0.66	SOD1^{N56S}	0.27
TTHA^{E29K}	0.24	HAH1$^{K60E/K56E}$	0.79	HAH1^{E68Q}	1.06	SOD1^{S25N}	0.36
TTHA^{E29K}	0.52	HAH1$^{K60E/K56E}$	0.48	HAH1^{E5K}	0.56	SOD1^{V88A}	0.22
TTHA^{K61E}	0.26	HAH1$^{K60E/K57E}$	0.77	HAH1$^{L65S/G66S/L67S}$	0.91	SOD1$^{Q15E/I17E/N19E}$	0.97
TTHA^{K61E}	0.24	HAH1$^{K60E/K57E}$	0.83	SOD1^{barrel}	0.43		

Protein concentrations ([protein]^{lysate}) were determined from the NMR signal intensity of the decrowded sample lysates (lysate^{supernatant}), using standard NMR signal intensities at [protein] = 100 μM as references. The intracellular protein concentrations were subsequently estimated from [protein]$^{in\ cell}$ = [protein]$^{lysate}/\varphi$, where $\varphi \approx 0.29$ is the *E. coli* volume fraction of the NMR sample (sample$^{in\ cell}$).

Table S3. Kinetic parameters and protein stabilities from in vitro controls

Protein	log $k_f^{H_2O}$*	m_f,* M^{-1}	log $k_u^{H_2O}$*	m_u,* M^{-1}	log $K_{D-N}^{H_2O}$†	$\Delta G_{D-N}^{H_2O}$‡ kcal/mol
TTHA^{PWT}	1.11 ± 0.02	−0.97 ± 0.02	−3.33 ± 0.04	0.61 ± 0.01	4.44 ± 0.06	6.04 ± 0.08
TTHA^{E57K}	1.06 ± 0.03	−0.87 ± 0.01	−1.96 ± 0.02	0.59 ± 0.01	3.02 ± 0.05	4.11 ± 0.07
HAH1^{PWT}	2.68 ± 0.03	−0.49 ± 0.02	0.27 ± 0.05	0.22 ± 0.01	2.41 ± 0.08	3.28 ± 0.11
HAH1^{K57E}	2.56 ± 0.02	−0.54 ± 0.01	0.51 ± 0.03	0.16 ± 0	2.05 ± 0.05	2.79 ± 0.07
SOD1^{barrel}	−0.28 ± 0.03	−0.78 ± 0.01	−4.03 ± 0.04	0.27 ± 0.01	3.75 ± 0.07	5.1 ± 0.1

*Derived from chevron data according to Eq. **S2**.
†Calculated from log $K_{D-N}^{H_2O}$ = log $k_u^{H_2O}$ − log $k_f^{H_2O}$.
‡Calculated from $\Delta G_{D-N}^{H_2O}$ = −2.3RT log $k_{D-N}^{H_2O}$.

Table S4. Hydrodynamic dimensions of TTHA^{PWT}, HAH1^{PWT}, and SOD1^{barrel} determined by PFG NMR diffusion measurements

Protein	D_T,* 10^{-10} m^2/s	r_H,† Å	M_W,‡ Da	Number of residues
TTHA^{PWT}	1.29 ± 0.02	15.5	7,041	66
HAH1^{PWT}	1.27 ± 0.02	15.7	7,402	68
SOD1$^{barrel\ c}$	1.01 ± 0.01	19.7	10,976	110

*Determined at 298 K in 100% D$_2$O.
†Calculated from Stoke−Einstein relationship.
‡Data from ref. 1.

Table S5. Residual squared sum from data fits to line and expanded plane models

Independent variables	RSS^{all}*	RSS^{all}/point	RSS^{sub}†	RSS^{sub}/point
charge^{density}	2.98	0.015	2.14	0.014
[charge^{density}; hydrophobicity^{SASA}]	2.24	0.012	1.75	0.011
[charge^{density}; hydrophilicity^{Guy}]	2.10	0.011	1.45	0.009
[charge^{density}; hydrophilicity^{Guy}; $||p||$]	1.79	0.009	1.25	0.008

*All data points are included, n = 193.
†Corresponds to the subset with very low leakage, n = 155.

Movie S1. Rotation of the plot in Fig. 6*A*, showing the planar relation between the protein construct's in-cell motions (mobility$^{in\ cell}$; Eq. **3**) and their surface properties, i.e., the surface net charge density (charge^{density}; Eq. **5**) and the surface hydrophilicity (hydrophilicity^{Guy}; Eq. **7**). The wild-type proteins are represented by ■ (black solid squares), the pseudo-wild-type variants by ● (black solid circles), and the mutants by ◯ (colored solid circles), color-coded according to Fig. 1.

Movie S1

Movie S2. Rotation of the plot in Fig. 6*B*, showing the planar relation between the protein construct's in-cell motions (mobility$^{\text{in cell}}$; Eq. **3**) and their surface properties, i.e., the surface net charge density (charge$^{\text{density}}$; Eq. **5**) and the electric dipole moment ($\|p\|$; Eq. **8**). The wild-type proteins are represented by ■ (black solid squares) the pseudo-wild-type variants by ● (black solid circles), and the mutants by ◐ (colored solid circles), color-coded according to Fig. 1.

[Movie S2](#)