Protein stability and mobility in live cells

Revelation of the intracellular diffusive interaction organization mechanisms

Xin Mu
Protein stability and mobility in live cells
Revelation of the intracellular diffusive interaction organization mechanisms

Xin Mu

Academic dissertation for the in Biochemistry at Stockholm University to be publicly defended on Thursday 19 December 2019 at 10.00 in Nordenskiöldsalen, Geovetenskapens hus, Svante Arrhenius väg 12.

Abstract
Biochemical processes inside living cells take place in a confined and highly crowded environment. As such, macromolecular crowding, one of the most important physicochemical properties of cytoplasm, is an essential element of cell physiology. It not only gives rise to steric repulsion, but also promotes non-specific, transient, interactions (referred to as diffusive interactions) between molecules. Since diffusive interactions are a key way to achieving a highly organized intracellular environment, without such interactions, the cell is just “a bag of molecules”. Therefore, understanding how diffusive interactions modulate protein behavior in live cells is of fundamental importance for revealing the mechanisms of molecular recognition, as well as for understanding the cause of protein misfolding diseases.

This thesis focuses on how macromolecular crowding influences the stability and diffusive motions of proteins within living cells by modulating their diffusive interactions. First, we investigated the thermal stability of superoxide dismutase 1 (SOD1), a protein involved in the development of familial amyotrophic lateral sclerosis (ALS), in mammalian and E. coli cells. Intriguingly, the major thermodynamic consequence of macromolecular crowding is due not only to conventional steric repulsions, but primarily to the detailed chemical nature of the diffusive protein interactions in live cells. Secondly, we presented a mutational study of how these diffusive interactions influence the rotation of proteins in the mammalian and bacterial cytosol. The result is a quantitative description of the physicochemical code for the intracellular protein motion, showing that it depends critically on the surface details of protein and the type of the host cell as well. Thirdly, we characterized the impact of intracellular protein concentration by altering the volume of E. coli cells by osmotic shock. The results obtained show that the intracellular diffusion of proteins is not determined by the chemical properties of the protein surface alone, but also by the frequency of concentration-dependent encounters. Moreover, it appears that eukaryotes and bacteria have achieved fidelity of biological processes through different evolutionary strategies. Overall, these observations have numerous implications for both functional protein design and deciphering the evolution of the surface characteristics of proteins.

Subsequently, we attempted to shed new light on the Hofmeister series, using protein-folding kinetics as observable. The results indicate that the Hofmeister series cannot be explained entirely by the traditional Kosmotropes/Chaotropes classification. Strong hetero-ion pairing cannot be ignored when trying to understand the effects of salts on protein salting-in and salting-out behaviors.

Keywords: diffusive interactions, macromolecular crowding, protein thermodynamic stability, protein mobility, in-cell NMR, Hofmeister series.

Stockholm 2019
http://urn.kb.se/resolve?urn=urn:nbn:se:diva-175632

ISBN 978-91-7797-932-6

Department of Biochemistry and Biophysics
Stockholm University, 106 91 Stockholm
Protein stability and mobility in live cells
Revelation of the intracellular diffusive interaction organization mechanisms

Xin Mu
To: Huaguo Mu
   Xiaojun Mu
   Biao Zhang
   Bin Yang
   Boyan Yang
List of Publications

Primary publications

I Thermodynamics of protein destabilization in live cells

II Physicochemical code for quinary protein interactions in Escherichia coli
Mu X, Seongil C, Lang L, Mowray D, Dokholyan N, Danielsson J*, Oliveberg M*.

III Diffusive protein interactions in human versus bacterial cells
Leeb S, Grönlund T, Yang F, Mu X, Oliveberg M*, Danielsson J*.
Submitted.

IV Quantification of the crowding effect on diffusive protein interactions in Escherichia coli by cell-volume modulation
Mu X, Danielsson J, Oliveberg M*.
Manuscript.

V Salt effects on protein folding behavior: revealing the molecular origins and limitation of Hofmeister theory
Mu X, Danielsson J, Oliveberg M*.
Manuscript.

* X.M. and J.D. contributed equally to this work.
Additional publications

VI The cost of long catalytic loops in folding and stability of the ALS-associated protein SOD1

VII Fast association and slow transitions in the interaction between two intrinsically disordered protein domains

VIII The transition state structure for coupled binding and folding of disordered protein domains

IX A frustrated binding interface for intrinsically disordered proteins
Contents

List of Publications ............................................................................................................. viii

Contents .......................................................................................................................... x

Abbreviations .................................................................................................................... xii

Preface ............................................................................................................................. 1

Part I: Protein stability in live cells .............................................................................. 2
  Introduction ................................................................................................................... 2
    Protein folding ............................................................................................................. 3
    Protein thermodynamic stability ............................................................................. 7
    Nuclear magnetic resonance (NMR) ......................................................................... 9
    Our model protein ................................................................................................... 11
  Results and Discussion .............................................................................................. 13
    Global unfolding of SOD1^{135A} within cells ..................................................... 13
    The effect of crowding on thermodynamic stability ........................................... 13
    The behavior of SOD1^{135A} in different cellular environments ....................... 16
  Summary ..................................................................................................................... 18

Part II: Organization of diffusive protein interactions in live cells ......................... 19
  Introduction ................................................................................................................ 19
    The rotational correlation time ............................................................................. 20
    The molecular situation in E. coli cytoplasm ....................................................... 21
    Osmoregulation mechanisms in E. coli .................................................................... 22
    Our model proteins ............................................................................................... 23
  Results and Discussion .............................................................................................. 24
    Physicochemical code for diffusive protein interactions in E. coli ..................... 24
    Protein mobility in different cellular environments ............................................ 27
    Effects of protein size and surface charge on mobility^{in cell} ................................ 30
    Effects of intracellular protein concentration on mobility^{in cell} ....................... 31
  Summary ..................................................................................................................... 34
  Outlook ....................................................................................................................... 35

Part III: Effects of salt on protein folding ................................................................. 36
  Introduction ................................................................................................................ 36
The Hofmeister series.........................................................................................37
Protein solubility ................................................................................................37
Screening electrostatic interactions...................................................................38
Formation of water surface tension at the molecular level .........................38
The effects of ion-water interactions on water surface tension .....................39
Our model proteins ..........................................................................................41
Experimental parameters and nomenclature ..................................................42
Results and Discussion ....................................................................................44
  The Debye-screening effect at low [salt] ......................................................44
  The relevance of $M_p(k_l)$ to protein solubility ..........................................45
  Ionic hydration cannot rationalize the Hofmeister effect fully ......................45
  Strong ion pairing by Gdm$^+$ and SO$_4^{2-}$ at high concentrations ..............46
Conclusions .......................................................................................................50
Future perspectives ..........................................................................................50
Populärvetenskaplig Sammanfattning .............................................................51
Acknowledgements .........................................................................................54
References ........................................................................................................56
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALS</td>
<td>amyotrophic lateral sclerosis</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>SOD1</td>
<td>superoxide dismutase 1</td>
</tr>
<tr>
<td>N</td>
<td>the native ensemble</td>
</tr>
<tr>
<td>D</td>
<td>the denatured ensemble</td>
</tr>
<tr>
<td>GdmCl</td>
<td>guanidine hydrochloride</td>
</tr>
<tr>
<td>$k_f$</td>
<td>refolding rate constant</td>
</tr>
<tr>
<td>$k_u$</td>
<td>unfolding rate constant</td>
</tr>
<tr>
<td>$m_f$</td>
<td>refolding m-value</td>
</tr>
<tr>
<td>$m_u$</td>
<td>unfolding m-value</td>
</tr>
<tr>
<td>$G$</td>
<td>Gibbs free energy</td>
</tr>
<tr>
<td>$H$</td>
<td>enthalpy</td>
</tr>
<tr>
<td>$S$</td>
<td>entropy</td>
</tr>
<tr>
<td>Trp</td>
<td>tryptophan</td>
</tr>
<tr>
<td>Tyr</td>
<td>tyrosine</td>
</tr>
<tr>
<td>SASA</td>
<td>solvent accessible surface area</td>
</tr>
<tr>
<td>TS</td>
<td>transition state</td>
</tr>
<tr>
<td>$C_p$</td>
<td>molar heat capacity</td>
</tr>
<tr>
<td>$T_c$</td>
<td>cold unfolding temperature</td>
</tr>
<tr>
<td>$T_m$</td>
<td>melting temperature</td>
</tr>
<tr>
<td>R</td>
<td>gas constant/linear correlation coefficient</td>
</tr>
<tr>
<td>E. coli</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>CPP</td>
<td>cell-penetrating peptide</td>
</tr>
<tr>
<td>$\tau_r$</td>
<td>rotational correlation time</td>
</tr>
<tr>
<td>$\eta$</td>
<td>solvent viscosity</td>
</tr>
<tr>
<td>$R_H$</td>
<td>effective hydrodynamic radius</td>
</tr>
<tr>
<td>$k_B$</td>
<td>Boltzmann constant</td>
</tr>
<tr>
<td>FT</td>
<td>Fourier transform</td>
</tr>
<tr>
<td>MW</td>
<td>molecular weight</td>
</tr>
<tr>
<td>DASA</td>
<td>dynamically accessible surface area</td>
</tr>
<tr>
<td>3D</td>
<td>three-dimensional</td>
</tr>
<tr>
<td>$R_2$</td>
<td>transverse relaxation rate</td>
</tr>
<tr>
<td>$\eta_{app}$</td>
<td>apparent viscosity</td>
</tr>
</tbody>
</table>
\( \eta_{\text{int}} \) intrinsic microscopic solvent viscosity
GFP green fluorescent protein
\( s \) protein solubility
\( K_s \) salting-out constant
\( \varepsilon_r \) dielectric constant
\( \varepsilon_0 \) permittivity of vacuum
\( l_D \) Debye screening length
\( F \) inward pulling force
\( \gamma \) water surface tension
Preface

Many metabolic pathways are located in the cell cytosol, an extraordinarily complex, heterogeneous, and dense aqueous solution of, mostly, ions, metabolites, and a large variety of macromolecules (1), being even more crowded in prokaryotes than eukaryotes (2). This cytosol is not a simple solution, but contains multiple levels of self-organization that ensure the fidelity and efficiency of the biochemical processes that occur there (3). Although the structures and functions of proteins have been explored extensively in vitro, it remains unclear how these are organized in the crowded interior of living cells. To what extent are data obtained under simplified in vitro conditions relevant to the environment in live cells?

The present thesis focuses primarily on how intracellular crowding modulates protein stability (Part I and Paper I), as well as diffusive motions (Part II and Papers II, III, IV) as assessed by in-cell NMR. In addition, we have explored the effects of salts on protein folding (Part III and Paper V) to discuss the molecular origins of protein salting-out and salting-in behaviors.
Part I: Protein stability in live cells

(Paper I: Thermodynamics of protein destabilization in live cells)

Introduction

The thermodynamic stability of a protein is essential to its structure and function. Although protein stability has been probed extensively *in vitro*, it remains unclear how these basic events are regulated in the crowded matrix inside living cells. As mentioned above, the interior of cells is highly crowded, with a high concentration of macromolecules (4), which gives rise to “excluded volume effect”\(^1\). Such effect is present in all cells and essential for metabolic efficiency (5–7). For example, it can promote protein folding (*i.e.*, an increase in protein stability), because a protein occupies a smaller volume when folded than an unfolded (8). However, in some cases, cells appear to destabilize the native structure of a protein (9–15), in contrast to the protein stabilization (16–20), resulting from the intracellular excluded volume effects (5–7). This is because, in such an environment, various kinds of ‘non-native’ interactions (referred to as diffusive interactions) inevitably take place between macromolecules. Diffusive interactions\(^2\), although typically weak and short-lived, are nonetheless expected to influence cellular functions simply due to their large numbers, altering both the stability (21) and mobility (22) of proteins. Clearly, it is of considerable importance to understand equilibria between the folded and unfolded states of proteins under physiological conditions. Accordingly, we have characterized the thermodynamic behavior of a marginally stable variant of SOD1 in cell, utilizing *in-cell* NMR.

---

1 Excluded volume effect stems from the steric repulsion, which limits the available volume of macromolecules, because two molecules cannot present simultaneously at the same place (2, 5–7).

2 Diffusive interactions include repulsive and attractive nonspecific interactions, with the former enhancing the excluded volume effect and the latter opposing (15).
Protein folding

Although the primary structure of a protein determines its tertiary structure and thereby its biological function (23, 24), it is difficult to predict this tertiary structure from the amino acid sequence (25). Currently, X-ray crystallography and nuclear magnetic resonance (NMR) are employed to examine the tertiary structure of a protein (26–28). However, to predict how a protein folds into its final sophisticated structure solely on the basis of its amino acid sequence, we need to understand protein folding at the atomic level.

Two-state folding

Small globular proteins transition from an unfolded to a folded state without the formation of any stable intermediate state, as shown in the simple Scheme 1 (28, 29).

\[ \text{D} \rightleftharpoons \text{N} \]  
(Scheme 1)

where D represents an ensemble of unfolded protein conformations and N is an ensemble of folded state conformations. \( K_{D-N} \), defined as the equilibrium constant for this process, is the ratio of the concentration of the folded protein ([N]) and the unfolded protein concentration ([D]) at the chemical equilibrium (Equation 1).

\[ K_{D-N} = \frac{[N]}{[D]} \]  
(Equation 1)

The energy difference between these two states is referred to as \( \Delta G_{D-N} \) and provides an indicator of the stability of a protein:

\[ \Delta G_{D-N} = G_{N} - G_{D} = -RT\ln K_{D-N} = -2.3RT\log K_{D-N} \]  
(Equation 2)

where \( G_{N} \) and \( G_{D} \) are the free energies of the folded and unfolded states, respectively; R is the gas constant; T is the temperature in degree Kelvin. When \( K_{D-N} \) is 1, \( \Delta G_{D-N} \) is 0, and the numbers of folded and unfolded protein molecules are identical.

Chemical denaturation

Proteins can be unfolded by a variety of conditions, including heat, high pressure, altering the acidity/alkalinity or addition of a chemical denaturant.
Typically, small proteins can “recover” from such denaturation, *i.e.*, the unfolding/folding process is usually reversible (Scheme 1) (29).

**Tryptophan fluorescence**

The intrinsic fluorescence of tryptophan (Trp) and tyrosine (Tyr), both of which can be excited at 280 nm and are sensitive to changes in their microenvironment, are widely used to monitor protein folding and unfolding. Both the intensity of their fluorescence and wavelength of maximal emission change during folding/unfolding (Figure 1). Plotting these parameters against the concentration of the denaturant ([denaturant]) provides sigmoidal equilibrium curves (Figure 1) (30).

**Protein folding equilibrium**

Protein folding equilibrium, which can give key information about conformational stability (29), is studied using a wide range of [denaturant]s from high (favoring D, so that \( K_{D-N} < 1 \)) to low [denaturant] (\( K_{D-N} > 1 \)) (Figure 1).

![Figure 1](image)

Figure 1. A. Fluorescence profiles of a folded (N) and unfolded (D) proteins. In this case, the folded protein exhibits greater fluorescent intensity. B. The equilibrium unfolding curve corresponding to A. \( I_{300-400} \) is the sum of the fluorescent intensity from 300 nm to 400 nm. At low [denaturant], most of the protein molecules are folded, but as the [denaturant] increase, more begin to unfold. The midpoint of the transition (\( M_p \)) is the [denaturant] at which \( K_{D-N} = 1 \).

The \( \log K_{D-N} \) of a two-state protein demonstrates linear relationship to the [denaturant]:

\[
\log K_{D-N} = \log K_{D-N}^{H_2O} + m_{D-N} [\text{denaturant}]
\]

(Equation 3)
where, \( \log K_{D-N}^{H_2O} \) is the log \( K_{D-N} \) in water, and \( m_{D-N} \) is the slope of the log \( K_{D-N} \) versus [denaturant] plot. This slope is referred to as the equilibrium m-value and is related to changes in the solvent-accessible surface area \( \Delta \text{SASA} \) of the protein during the process of unfolding (31).

### Protein folding kinetics

The detailed kinetics of protein folding provide key information concerning the pathway involved. The relaxation rate constant (\( k_{\text{obs}} \), Equation 4) for the change in fluorescence intensity during the refolding/unfolding process can be obtained by fitting an exponential function to the corresponding plot (Figure 2) (32).

\[
k_{\text{obs}} = k_f + k_u \tag{Equation 4}
\]

where \( k_f \) and \( k_u \) are the rate constants of refolding and unfolding, respectively. In the case of a two-state system, the dependence of \( \log k_{\text{obs}} \) on [denaturant] is V-shaped (a so-called chevron plot), with both unfolding (N \( \rightarrow \) D) and refolding (D \( \rightarrow \) N) limbs (Figure 2).

The logarithms of \( k_f \) and \( k_u \) are linearly correlated with [denaturant], and extrapolation of the refolding and unfolding limbs of the chevron plot to the y-axis can give \( \log k_f^{H_2O} \) and \( \log k_u^{H_2O} \), both in water (Figure 2). The corresponding slopes represent the m-values for refolding \( (m_f) \) and unfolding \( (m_u) \), which are proportional to the \( \Delta \text{SASA} \) from D to the transition state (TS) and from the TS to N, respectively (Equations 5 and 6). The equilibrium m-value \( (m_{D-N}) \) is equal to the sum of the absolute values of \( m_f \) and \( m_u \) (28).

The unfolding and refolding limbs intersect at the transition midpoint \( (M_p) \) (Figure 2) At \( M_p \), \( k_f = k_u \); Below \( M_p \), \( k_u << k_f \) and \( k_f \) is the predominant determinant of \( k_{\text{obs}} \); Above \( M_p \), \( k_u >> k_f \) and \( k_u \) is the predominant determinant of \( k_{\text{obs}} \) (28).

\[
\log k_f = \log k_f^{H_2O} + m_f \text{[denaturant]} \tag{Equation 5}
\]

\[
\log k_u = \log k_u^{H_2O} + m_u \text{[denaturant]} \tag{Equation 6}
\]

\[
m_{D-N} = |m_f| + |m_u| \tag{Equation 7}
\]
\[
\log K_{D-N}^{H_2O} = \log k_f^{H_2O} - \log k_u^{H_2O}
\]  
(Equation 8)

In practice, \( \log K_{D-N}^{H_2O} \) indicates the stability of the protein in water. The mathematical expression of \( \log k_{obs} \) (Equation 9) can be deduced from the combination of Equations 5, 6 and 7. Parameters that determine the folding pathway can be extracted from the chevron plot by applying Equation 9 and utilizing these parameters, the folding free-energy can be represented as shown in Figure 3 (28, 33).

\[
\log k_{obs} = \log \left( 10^{\log k_f^{H_2O} + m_f[\text{denaturant}]} + 10^{\log k_u^{H_2O} + m_u[\text{denaturant}]} \right)
\]  
(Equation 9)

![Figure 2](image_url)

**Figure 2.** A. A curve reflecting the kinetics of protein unfolding/refolding. The fluorescence intensity increases with time until the steady-state is reached. \( k_{obs} \) can be extracted from the single exponential of this curve. B. The chevron plot for a two-state protein (28).

![Figure 3](image_url)

**Figure 3.** A free energy diagram of two-state folding. The protein proceeds from D to N through an energy barrier, called the transition state (‡), which is unstable and short-lived. The height of the free-energy barrier indicates the amount of energy required to transition from N to ‡ (\( \Delta G_u \)) or from D to ‡ (\( \Delta G_f \)). The difference in energy between N and D is \( \Delta G_{D-N} \). The reaction coordinates during the folding process can be presented as \( \Delta \text{SASA} \) with the reaction rate from D to the TS being the rate of folding \( k_f \) and , from N to the TS being the rate of unfolding \( k_u \) (33).
Protein thermodynamic stability

The stability of the various states involved in protein-folding can be described in terms of $\Delta G$, which incorporates changes in both the entropy ($\Delta S$) and enthalpy ($\Delta H$) (Equation 10), where $S$ is the measure of states available in the system. The number of states is much higher for D, so that entropy is reduced upon folding. $H$ is the heat gained or lost and protein folding is exothermic ($\Delta H < 0$) with the formation of numerous intramolecular interactions (34).

$$\Delta G = \Delta H - T\Delta S$$

(Equation 10)

The heat capacity associated with thermal unfolding of proteins

The heat capacity of a protein ($C$) is the heat added or removed in order to increase or decrease the temperature by 1 K. Accordingly, the molar heat capacity $C_p$ (where $p$ means under constant pressure) is the amount of heat required to alter the temperature of one mole of protein by 1 K (28, 34). Unfolding a protein elevates its heat capacity, an important determinant of the temperature dependence of this process and, thus, of protein stability (35–37). The overall enhancement of heat capacity during protein unfolding is due primarily to hydration of non-polar groups, as more and more of these groups normally located inside the hydrophobic core are exposed to the solvent (38). Therefore, heat capacity is an indicator of the $\Delta$SASA of the protein. This explains why the heat capacity of the unfolded protein is higher than that in the folded state (37–41), more heat being required to “melt” the larger hydration shell of the unfolded polymer.

$\Delta C_p$ and m-value relevant to $\Delta$SASA upon protein unfolding

For a two-state protein, the $\Delta$SASA upon unfolding is a key determinant of the m-value and $\Delta C_p$ (31). At the same time, the thermodynamic parameters ($\Delta S$ and $\Delta H$) connected with folding are distinct and temperature-dependent (42, 43). According to classical thermodynamics, the relationship of $\Delta C_p$ to $\Delta H$ and $\Delta S$ can be expressed as in Equations 11 and 12 (31, 42–44).

$$\Delta C_p = \frac{\partial \Delta H}{\partial T}$$  \hspace{1cm}  (Equation 11)

$$\frac{\Delta C_p}{T} = \frac{\partial \Delta S}{\partial T}$$  \hspace{1cm}  (Equation 12)
For the type of two-state system illustrated in Scheme 1, $\Delta H_{D-N}(T)$ between the folded and unfolded states at any given temperature ($T$) can be calculated with Equation 13 (31, 42–44),

$$\Delta H_{D-N}(T) = H_N(T) - H_D(T) = \Delta H_{D-N}(T_m) - (T_m - T)\Delta C_p^{D-N}$$  \hspace{1cm} (Equation 13)

where $H_N(T)$ and $H_D(T)$ are the overall enthalpies of the folded and unfolded protein at that temperature, respectively; $\Delta C_p^{D-N}$ is the difference between N and D with respect to $C_p$; and $T_m$ is the transition temperature with $\Delta G_{D-N}(T_m) = 0$. Equation 10 gives rise to Equation 14 (31, 42–44):

$$\Delta G_{D-N}(T_m) = \Delta H_{D-N}(T_m) - T_m\Delta S_{D-N}(T_m) = 0$$  \hspace{1cm} (Equation 14)

$\Delta S_{D-N}(T_m)$ is the difference between N and D at $T_m$ with respect to S and $\Delta S_{D-N}(T)$ is (31, 42–44):

$$\Delta S_{D-N}(T) = S_N(T) - S_D(T) = \frac{\Delta H_{D-N}(T_m)}{T_m} - \Delta C_p^{D-N}\ln(T_m/T)$$  \hspace{1cm} (Equation 15)

Equations 10, 13 and 15 can be combined to derive the following equation (31, 42–44):

$$\Delta G_{D-N}(T) = \Delta H_{D-N}(T) - T\Delta S_{D-N}(T)$$

$$= \Delta H_{D-N}(T_m) \frac{T_m - T}{T_m} - (T_m - T)\Delta C_p^{D-N} + T\Delta C_p^{D-N}\ln(T_m/T)$$

$$= (T_m - T)\Delta S_{D-N}(T_m) - (T_m - T)\Delta C_p^{D-N} + T\Delta C_p^{D-N}\ln(T_m/T)$$  \hspace{1cm} (Equation 16)

**Stability curves for proteins**

The stability curve of a protein plots the free energy of unfolding against temperature, resulting in a parabola (Figure 4) (44), with the protein being most stable when the free energy is lowest ($\Delta G_{D-N}_{min}$. $\Delta G = 0$ (i.e., there are equal amounts of folded and unfolded protein) at two temperatures, namely, the melting temperature ($T_m$) and cold unfolding temperature ($T_c$). Typically, $T_m$ is above room temperature, while $T_c$ is below freezing. All temperatures between $T_m$ and $T_c$ are associated with a negative $\Delta G$, i.e., more of the population is folded than unfolded. Conversely, when $\Delta G > 0$, there is more unfolded protein (Figure 4) (31, 42–44).
Fitting the stability curve with Equation 16 allows the thermodynamic parameters associated with thermal unfolding to be extracted.

In connection with this stability curve, $\Delta C_p$, representing the extent of curvature, is obtained as shown below (44):

$$\frac{\partial^2 \Delta G}{\partial T^2} = -\frac{\Delta C_p}{T}$$

(Equation 17)

**Nuclear magnetic resonance (NMR)**

Multidimensional correlation experiments have been used extensively to detect pairs of coupled nuclei, such as $^1$H-$^{15}$N and $^1$H-$^{13}$C in proteins labeled appropriately. Each such pair gives rise to a resonance cross-peak at a distinct frequency in the NMR spectrum (45, 46), a frequency determined by the chemical environment of the protein. Thus, NMR spectra can reflect the overall conformational state of a protein, as well as local environmental changes due to structural dynamics or interactions with other molecules. For a folded protein, each cross-peak is influenced by very different local environments, resulting in chemical shifts for all residues and cross-peaks distributed across a wide range of frequencies in the $^1$H dimension. In the case of an unfolded protein, on the other hand, the rapid reconfiguration and overall dynamics of the disordered polypeptide chain result in an average environment for each residue that is similar, so that the chemical shifts will also be similar and clustered within a narrow range, typically between 7.5-8.5 ppm in the proton dimension. This difference allows definitive discrimination of folded from unfolded proteins (21), both in live cells and in vitro.
Consequently, information related to protein stability can be obtained by quantifying the \([N]/[D]\) equilibrium (Equation 1) from the integral of identified cross-peaks for N and D states in an NMR spectrum (21).

**Application of NMR to living cells**

Biophysical approaches, such as X-ray crystallography and cryo-electron microscopy, can be utilized to characterize the structural features of biological macromolecules *in vitro*, but have severely limited capacity to deal with extremely heterogeneous *in-cell* samples. The increasing focus on the functional characteristics of biological macromolecules in their native environment demands high-resolution techniques for exploring their behavior in living cells at atomic levels. *In-cell* NMR provides one such possibility (11, 47, 48).

A prerequisite for *in-cell* NMR is the introduction of the protein of interest labeled with \(^{15}\)N or \(^{13}\)C (NMR-active nuclei) into living cells, which can be accomplished simply in prokaryotes through overexpression of the protein (Figure 5). Here, the protein of interest has been overexpressed in *Escherichia coli* (*E. coli*), which with its viral promoter and polymerase accumulates exogenous proteins rapidly after induction. Application of *in-cell* NMR spectroscopy to yeast, insect cells and mammalian cells has been successfully achieved (47, 49, 50). However, with eukaryotic cells, it’s more difficult to attain sufficient overexpression of the labeled protein to allow acquisition of high-quality *in-cell* NMR spectra. In addition, the cost of isotope-labeled medium is higher in the case of eukaryotic cells than for *E. coli*. Moreover, recombinant proteins are expressed by eukaryotic cells for extended periods (usually several days), which can increase background noise due to accumulation of isotope-labeled endogenous cellular material.

Alternatively, isolated isotope-labeled proteins can be transferred into eukaryotic cells by several delivery procedures. For example, oocytes of *Xenopus laevis* have been transfected with concentrated protein by microinjection (51, 52). Furthermore, a cell-penetrating peptide (CPP), covalently linked to the protein via a disulfide bond has been utilized to transfer model proteins across the plasma membrane of HeLa cells, with subsequent release of the CPP by disulfide bond cleavage in the reducing environment of the cytoplasm (11). Moreover, labeled proteins can enter the cytoplasm through pores formed in the cell membrane by *e.g.*, the toxin Streptolysin O. (which forms pores that easily can be blocked by excess Ca\(^{2+}\)) (53). Two other methods transiently change the permeability of the plasma membrane by electroporation (54) or hypotonic swelling (55). Proteins can translocate into the cytoplasm through the transiently permeable membranes. In our investigations, we employed electroporation to deliver \(^{15}\)N-labeled proteins into the
cytoplasm of eukaryotic cells, a procedure established by Selenko and coworkers for *in-cell* NMR (Figure 5) (21, 54). Although these approaches eliminate background noise arising from endogenous labeling, they are challenging in practice, demanding a high concentration of the isotope-labeled protein and involving a complicated experimental procedure, as well as threats to cell viability.

![Figure 5](image)

**Figure 5.** The procedures employed here to introduce proteins into living cells. A. $^{15}$N-labeled protein molecules (green dots) are over-expressed in bacterial (22). B: $^{15}$N-labeled proteins are delivered into mammalian cells by electroporation. First, a highly concentrated solution of the protein is mixed with the cells in an electroporation cuvette, following which a strong pulsed electric field is applied. This permeabilizes the cell membrane, allowing $^{15}$N-labeled protein to enter the cytoplasm. After 4 hours of recovery, the cells are ready for *in-cell* NMR measurements (21).

Our model protein

The folding/unfolding of human SOD1, a ubiquitous intracellular antioxidant enzyme (56), is of considerable importance. Not only can absence of SOD1 result in permanent cellular damage, but aggregation of this protein can lead to amyotrophic lateral sclerosis (ALS), a neurodegenerative disease characterized by a loss of motor neurons that causes muscle atrophy.

Human SOD1 is a homo-dimer, in which each monomer consists of an 8-stranded $\beta$-barrel structure, a metal-binding loop IV and a so-called electrostatic loop VII (Figure 6). Since our aim was to investigate the intracellular stability rather than the activity of this protein (57), loops IV and VII involved in the dimerization and the catalytically active Cu$^{2+}$/Zn$^{2+}$ site were removed, producing a non-functional, but well-behaved monomeric $\beta$-barrel scaffold (SOD1$^{\text{barrel}}$) (Figure 6) (57). This SOD1$^{\text{barrel}}$ offers several advantages in connection with *in-cell* NMR studies of protein folding, display-
ing not only two-state folding (57) but also a fully resolved *in-cell* NMR spectrum (58).

To capture a mixture of both the N and D states\(^3\) with unchanged structure and surface properties at physiological temperatures, a destabilizing hydrophobic core mutation I35A (SOD1\(^{135A}\)) was introduced into SOD1\(_{\text{barrel}}\) (Figure 6) (21). This mutant protein is thermodynamically and kinetically similar to monomeric reduced full-length SOD1 (apoSOD1\(^{\text{pwt}}\)) (21, 59), which has been used extensively as a model for the aggregate precursors involved in ALS (60). Moreover, the marginally stable SOD1\(^{135A}\) is ideal for studying thermal unfolding, because its thermodynamic stability is sensitive to environmental perturbation by *e.g.*, temperature or ionic strength (21).

![Figure 6](image-url)

Figure 6. A. The native SOD1 dimer (PDB code 1HL5). In this holo state, both of the metal-binding loop IV and the electrostatic loop VII surround the active site in a compact and orderly manner, with loop IV also being involved in formation of the dimer interface (blue). The residues involved in binding Cu\(^{+2/0}\) in the active site and forming the C57–C146 disulfide bond are highlighted. B. The monomeric SOD1\(_{\text{barrel}}\) (PDB code 4BCZ) after truncating the loops (IV in blue and VII in cyan) by protein engineering (21, 57). C. The \(^1\)H–\(^15\)N 2D spectrum of SOD1\(_{\text{barrel}}\) at 37 °C, displaying uniformly folded protein. D. The corresponding \(^1\)H–\(^15\)N 2D spectrum of SOD1\(^{135A}\) (PDB code 4XCR) at 37 °C, showing mixed population of N and D states. The \([N]/[D]\) equilibrium (Equation 1) is quantified from the cross-peak volumes of the C-terminal resonance Q153. E. The stability curves for SOD1\(_{\text{barrel}}\) and SOD1\(^{135A}\) obtained with NMR thermal scans. The stability curve for SOD1\(^{135A}\) is shifted upwards relative to that for SOD1\(_{\text{barrel}}\), with \(T_m\) declining from 60 to 35.4 °C, and \(T_c\) rising from ~ 32.5 to ~ 2.5 °C, indicating destabilization by the mutation introduced (I35A).

\(^3\) It’s essential to capture a mixture of both the N and D states for the quantification of \(K_{D,N}\).
Results and Discussion

Global unfolding of SOD1\textsuperscript{I35A} within cells

SOD1\textsuperscript{I35A} is significantly less stable in A2780 cells than \textit{in vitro} (Figure 7) (21), with the thermal unfolding equilibrium at 37 °C being shifted four-fold in favor of D. This finding is contrary to the expectation that the excluded volume effect would increase protein stability (5–7), emphasizing the importance of diffusive interactions to modulate protein stability intracellularly (15, 21, 61). Notably, thermal unfolding in A2780 cells is accompanied by an increase in \( \Delta C_p \) by approximately 37 %\textsuperscript{4} (Figures 7 and 8) of the corresponding value \textit{in vitro}. As mentioned above, \( \Delta C_p \) is an indicator of the SASA of protein thermal unfolding. Since \textit{in-cell} NMR spectra show that N is relatively unaffected\textsuperscript{5}, it is reasonable to conclude that this increase in \( \Delta C_p \) is due primarily to modulation of D. The D state becomes more extensively hydrated (31), perhaps because its dispersed sites of diffusive interaction render its flexible conformation even more extended (7, 62).

What causes this destabilization of SOD1\textsuperscript{I35A} in live cells? One possibility is that with respect to mass action, the D and N states interact preferentially with the surrounding environment. If the number, strength and duration of diffusive interactions have a greater effect on N (\textit{i.e.}, the conformational equilibrium favors N), stabilization will occur; whereas if these are more pronounced for D, destabilization will be observed. Because of the relatively larger SASA associated with the D state, facilitating productive contacts, it is reasonable to speculate that D engages in a larger number of interactions intracellularly than N (Figure 8) (7, 62, 63).

The effect of crowding on thermodynamic stability

To determine the cause of this intracellular destabilization of SOD1\textsuperscript{I35A}, experiments with chemically distinct crowding agents were carried out \textit{in vitro}. Notably, ficoll\textsuperscript{70} and PEG\textsuperscript{400}, which are preferentially excluded from the protein surface, favor the folded state, resulting overall in stabilization of SOD1\textsuperscript{I35A} (Figure 9) (6). However, on a molecular level, the origin of this stabilization is more complicated than simply volume exclusion due to steric

\textsuperscript{4} Since \( \Delta C_p \) represents the extent of curvature in stability curve, \textit{i.e.}, the deeper, the greater the value, these curves in Figure 7 indicate the destabilization of SOD1\textsuperscript{I35A} in A2780 cells is accompanied by an increase in \( \Delta C_p \).

\textsuperscript{5} The NMR chemical shifts and line broadening of N are relatively unaffected, suggesting the structure of N is unchanged.
repulsions (64). Such synthetic crowders are generally considered to be hard-sphere/rod, accounting for their excluded volume effects at high concentrations (65–67), but ignoring potential attractive interactions with other molecules. Indeed, synthetic crowders do have attractive interactions with proteins (5) and should not be represented by hard-spheres. In addition, crowding agents are also osmolytes that influence protein hydration (68). PEG$^{400}$ has been reported to be intermediate between a stabilizing osmolyte and a chemical denaturant (64). Unlike ficoll$^{70}$, PEG$^{400}$ stabilizes SOD1$^{135A}$ with a significant increase in $\Delta C_p$, indicating expansion of the denatured state of SOD1$^{135A}$ in line with both a previous study (64) and our observations on human cells (21). The stabilizing effects of PEG$^{400}$ and ficoll$^{70}$ on SOD1$^{135A}$ are in contrast to the findings in living cells, suggesting that stabilization due to excluded volume and osmotic pressure are outweighed by the opposite diffusive interactions.

Crowding by proteins is more physiologically relevant and, as expected, when globular proteins were used as crowders, the results depended on their surface properties. More specifically, positively charged lysozyme (+ 8.5 e at pH 6.5) can potentially form strong electrostatic interactions with the negatively charged SOD1$^{135A}$ (− 0.5 e at pH 6.5), and, indeed, lysozyme destabilizes SOD1$^{135A}$ significantly in a manner similar to the intracellular destabilization (Figure 9). At the same time, BSA (− 8.5 e at pH 6.5), the bacterial putative heavy metal binding protein TTHA$^{PWT}$ (− 1.5 e at pH 6.5) and the cysteine-depleted holoSOD1$^{dimer}$ (− 5 e at pH 6.5) have only weak effects on SOD1$^{135A}$ stability (Figure 9). Taken together, these findings indicate that the impact of diffusive interactions on protein stability is sequence-specific (69), depending not only on the detailed surface features of the protein crowder, but also on the identity of the given protein itself (SOD1$^{135A}$) (21).

\[\text{Thermal unfolding of SOD1}^{135A} \text{in the presence of 100 mg/ml PEG}^{400} \text{is accompanied by an increase in } \Delta C_p \text{ by approximately 18 } \% \text{ of the corresponding value in pure buffer, whereas, in the case of 100 mg/ml ficoll}^{70}, \text{protein is stabilized with a decrease in } \Delta C_p \text{ by approximately 22 } \%.\]
Figure 7. Quantification of intracellular protein stability. The entire stability curves in A2780 and E. coli cells are shifted overall to a lower level, shifting \( T_c \) to a physiological range. \( T_m \) is reduced less extensively in E. coli than in A2780 cells. At the same time, the temperature at which SOD1\(^{I35A}\) is most stable is 14 °C in mammalian cells versus 20 °C in E. coli (21).

Figure 8. A. The \( \Delta C_p \) in live cells is approximately 37 % higher than the corresponding value \textit{in vitro}. B. The unfolding equilibrium shifts towards the D state due to diffusive interactions with the interior of the cell. Both D and N interact with other cellular molecules (the light blue cycles), but these interactions are more pronounced for the structurally expanded and flexible D. The elevated \( \Delta C_p \) observed in cells is attributed to the increased SASA (dotted boundary) of D, promoted by transient association with neighboring macromolecules (21). We estimate that at all times each molecule of SOD1\(^{I35A}\) would encounter approximately five potential interaction partners in its immediate surroundings (61).

Figure 9. Comparison of findings in live cells and \textit{in vitro} reveals that synthetic crowders (ficolI\(^{70}\) and PEG\(^{400}\)) stabilize SOD1\(^{I35A}\), in contrast to the intracellular destabilization, whereas protein crowders cause a wide variety of effects, underlining the dependence of protein-protein interactions on primary structure (21).
The shift in $T_m$ is generally accompanied by a variation in $\Delta H_{D-N}$. Based on the research of Ebbinghaus and coworkers (64, 70), the crowding effects observed can be classified on the basis of differences in the excess enthalpy ($\Delta \Delta H_{D-N}$) in relationship to the changes in $T_m$ ($\Delta T_m$). Figure 10 illustrates how PEG$^{400}$ and ficoll$^{70}$ cause enthalpic stabilization of SOD1$^{I35A}$ by preferential hydration with increasing values for both $\Delta H_{D-N}$ and $T_m$ ($\Delta \Delta H_{D-N} > 0$; $\Delta T_m > 0$). With lysozyme, the trend was opposite to that with PEG$^{400}$, i.e., lysozyme caused enthalpic destabilization ($\Delta \Delta H_{D-N} < 0$) by preferential binding, accompanied by a lowering of $T_m$. The same phenomenon was observed with BSA and TTHA$^{PWT}$, whereas holoSOD1$^{dimer}$ had almost no effect on $T_m$ (21). Moreover, the destabilization of SOD1$^{I35A}$ in both E. coli and A2780 cells could originate from preferential binding (diffusive interactions) of surrounding cellular components to the unfolded protein (Figure 10) (21). Altogether, the classification of in cell and in vitro crowders illustrated in Figure 10 indicates that a natural crowder (protein) with a heterogeneous surface provides a better model of the intracellular environment with respect to diffusive interactions than synthetic crowders, revealing the impact of sequence-specific diffusive interactions on protein stability (69).

The behavior of SOD1$^{I35A}$ in different cellular environments

We found that SOD1$^{I35A}$ is also destabilized in E. coli, but in a slightly different manner than in A2780 cells (Figures 7 and 10): SOD1$^{I35A}$ is more stable in E. coli above room temperature, but more stable in mammalian cells at lower temperatures. Moreover, even in one and the same cell line, protein stability is dependent on the phase of the cell cycle, as well as intracellular (organelle) localization (14, 19). Such distinct differences in protein stability depending on localization focus attention on the divergence of organisms, since the protein surface evolves much more rapidly than its structural scaffolds (71), resulting in different surface properties in bacteria and human cells. To obtain further insight into such differences, we transferred SOD1$^{I35A}$ from A2780 cells into the cytosol of E. coli and observed stabilization, based on an increase in $T_m$ ($\Delta T_m > 0$) together with a negative $\Delta \Delta H_{D-N}$ (Figure 10). Thus, it seems reasonable to conclude that the thermodynamic features of this behavior reflect more pronounced excluded volume effects in the even more crowded E. coli cytosol, providing a larger number of partners for diffusive interaction. The impact of this evolutionary diver-

---

7 The excluded volume effect generally stabilizes protein with $\Delta \Delta H_{D-N} < 0$, since the crowders also form certain non-specific attractive interactions (5).
gence, with an even denser and more challenging macromolecular environment in the bacterial cytosol on protein mobility in live cells is explored in detail in section II below (Paper II, III and IV).

Figure 10. Thermodynamic classification of the effects of crowding agents in live cells and in vitro. Based on the research of Ebbinghaus and coworkers, the effects of different crowding agents can be classified, utilizing the relationship between $\Delta \Delta H_{D-N}$ and $\Delta T_m$ (64). In vitro, the synthetic crowding agents PEG$_{400}$ and ficoll$^{70}$ stabilize SOD$_{1}^{135A}$, with both positive $\Delta T_m$ and $\Delta \Delta H_{D-N}$, and can be classified as PH (i.e., causing preferential hydration like an osmolyte). Lysozyme exerts the opposite effect on SOD$_{1}^{135A}$ stability, with both negative $\Delta T_m$ and $\Delta \Delta H_{D-N}$, as a result of diffusive interaction (PB) with the unfolded SOD$_{1}^{135A}$. The same behavior is observed with the other natural crowders (BSA and TTHA$^{PW7}$), but with less extensive destabilization, except for holoSOD$_{1}^{dimer}$ which influences SOD$_{1}^{135A}$ stability slightly ($\Delta T_m = 0$). Crowding destabilizes SOD$_{1}^{135A}$ in the cytosol of both A2780 (blue cycle) and E. coli (green triangle) cells in the same manner as lysozyme. However, transfer from mammalian cells to bacteria results in a negative $\Delta \Delta H_{D-N}$ accompanied by a positive $\Delta T_m$, an excluded volume effect involving enhanced transient binding.
Summary

The present findings answers how to translate protein stability from *in vitro* to *in live cells* conditions (7, 61, 72). In our specific case, SOD1<sup>135A</sup> is destabilized in living cells. Although the excluded volume and osmotic effects must, in general, be taken into consideration, the rules of the game are determined by the molecular details: intracellular protein stability is determined by all interaction partners, depending not only on the sequence of the protein itself, but also its interactions with its own intracellular environment. Consequently, different proteins are expected to exhibit different behavior in live cells, due to sequence-specific diffusive interactions (9, 10, 13, 15–20). This explains why it is difficult to mimic physiological environments with synthetic crowders such as ficoll<sup>70</sup> and PEG<sup>400</sup>. Furthermore, our study provides an example of how classical investigation of protein properties *in vitro* can easily neglect certain key physiological features, which may introduce bias. Although A2780 cells are not neuronal tissue, *in-cell* destabilization might indicate that the aggregate precursors (apoSOD1<sup>wt</sup>) involved in ALS are mostly unfolded in neurons (73, 74). Moreover, the soluble aggregated species of apoSOD1<sup>wt</sup> in the spinal cord of a murine model of ALS is recognized fully by antibodies targeting the disordered peptide epitope (75).
Part II: Organization of diffusive protein interactions in live cells

(Paper II: Physicochemical code for quinary protein interactions in Escherichia coli; Paper III: Diffusive protein interactions in human versus bacterial cells; Paper IV: Quantification of the crowding effect on diffusive protein interactions in Escherichia coli by cell-volume modulation.)

Introduction

Soluble cytosolic proteins are constantly moving and undergoing countless non-specific collisions with surrounding macromolecules (61). These diffusive interactions help the proteins to ‘sense’ their environment and find functional binding partners, a process critical to life (76–78). In comparison to specific interactions, the affinity of which is typically high, diffusive interactions are relatively weak and short-lived, making them difficult to quantify (79). For example, the sheer number of diffusive encounters in the crowded cellular interior gives rise to competition with specific binding (80), and control diffusion (81), as well as influencing the structural stability of proteins (21).

Importantly, diffusive interactions reduce the diffusion of proteins in the cytosol and too much reduction would make it difficult for proteins to encounter their functional interaction partners within a reasonable period of time (22, 82). The extent of crosstalk between a protein and its cellular environment can be quantified on the basis of retardation of protein rotational mobility by diffusive interactions, allowing us to examine the extent to which these ubiquitous background interactions are regulated and optimized (21, 61, 80, 83–85). To shed light on these issues, we employed in-cell NMR to monitor the response of diffusive protein interactions to specific point mutations, as reflected in the change in rotational tumbling (Paper II).

The strength and duration of diffusive interactions between a given protein and its intracellular environment depend not only on the surface characteris-
tics of the protein but also on the nature of surrounding molecules, i.e., the type of host cell (58, 86). With the exceptions of specific binding sites, the surface features of proteins are, in general, not well conserved evolutionarily and thus vary widely between organisms. The best example of this is the remarkably unique high-charge density of the proteome in the cytosol of halophilic bacteria (87, 88). Thus, an interesting question becomes how rotational motions of a protein are affected by transferring the given proteins from a prokaryotic to eukaryotic cell. In Papers II and III, we attempt to answer this question by exploring the retardation of protein rotation in both E. coli and human A2780 cells.

Intracellular crowding is particularly important for prokaryotes, which lack membrane bound-organelles (2). To determine how E. coli modulates diffusive interactions through macromolecular crowding, we explored the tuning of such interactions by crowdedness in the cytosol of E. coli in detail (Paper IV).

The rotational correlation time

Rotational correlation time ($\tau_r$), the average time required for a molecule to lose its rotational orientation correlation, is related to temperature, solvent viscosity, and molecular size (89). The Stokes–Einstein–Debye relationship (90) expression is as follows:

$$\tau_r = \frac{4\pi R_H^2 \eta}{3 k_B T} \quad \text{(Equation 18)}$$

where $\eta$ is the solvent viscosity, $R_H$ is the effective hydrodynamic radius of the protein, $k_B$ is the Boltzmann constant and $T$ is the temperature. $\tau_r$ is the characteristic time constant for the stochastic Brownian rotational diffusion of soluble molecules in aqueous solution.

For performance of liquid-state NMR, the protein of interest must rotate freely in its environment. The freedom of rotation influences the relaxation of each spin, thus affecting the peak intensities and line-widths in the NMR spectrum. In general, small molecules rotate more rapidly, thereby slowing down the overall relaxation and, eventually, providing an NMR signal of relatively narrow line-width (Figure 11). Conversely, with larger molecules or increasing solvent viscosity (Equation 18 and Figure 11), the molecular rotation slows down and relaxation becomes faster and the line-width of the NMR signal wider. When an isotope-labeled protein is placed in a crowded environment, such as the cytoplasm of E. coli, its rotational freedom could be limited by constant diffusive collisions with other macromolecules, there-
by broadening the line-widths of NMR signals (Figure 11). The larger the number of diffusive interactions, the wider the NMR signal, sometimes becoming too broad to be captured (22, 47, 91). Moreover, the line-width is dependent not only on $\tau$, but also on local protein dynamics and conformational exchange (91, 92). Accordingly, we optimized our model proteins by reducing the molecular size, simplifying the conformational states and altering surface properties for a high quality in-cell NMR spectrum (22, 47).

**Figure 11.** The relationship between the rate of molecular rotation and line-width of the corresponding NMR signal (inspired by Selenko’s illustration (47)). A. Rapid rotation, i.e., slow relaxation. The time domain is converted into a frequency domain with a Fourier transform (FT), resulting in a narrow, sharp NMR signal. B. If relaxation becomes rapid, the NMR signal will be broad. C. Small molecules rotate more rapidly in solution, giving a narrow, sharp NMR signal (A). Conversely, macromolecules rotate more slowly and relaxation becomes faster, and the signal broader (B). If the solvent viscosity is increased, the molecular rotation will slow down and the line-width becomes broad (B). In a complex and crowded environment (such as inside the cell), the target molecule would interact with other macromolecules, forming transient complexes that slow down its rotation and broaden line-width (B).

The molecular situation in *E. coli* cytoplasm

*E. coli* has been studied for decades, and the wealth of information accumulated makes this a desirable model for studying protein dynamics. Of the macromolecules in this bacterium, proteins occupy the most space, account-
ing for approximately 55 % of its dry weight (93). The cytoplasm is crowded with macromolecular assemblies, including ribosomes and RNA polymerase, with an overall macromolecular content of 300-400 mg/ml (Figure 12) (3, 94, 95). The cytoplasm of E. coli is unstructured, with no membrane-bound organelles (96) and the DNA is folded into compact cyclic structures located in the center of this cytoplasm (97). In addition, a large number of small molecules and ions present in the cytoplasm of E. coli to cope with changes in the external osmotic pressure (3, 98). Although the cytoplasm of E. coli is highly crowded and lacks clear functional divisions, its processes are organized in an impressive manner (3, 99–101). Macromolecular assemblies (e.g., RNA polymerase and ribosomes) are present in relatively high numbers and demonstrate impressive functional fidelity and efficiency (3). But how is this achieved? Characterization of protein behavior in E. coli could provide useful insights in the context?

Figure 12. A model of the E. coli cytoplasm. This “snapshot” shows that the cytoplasm is quite crowded with molecules constantly colliding with macromolecules in their surrounding (inspired by the model constructed by Elcock and colleagues (61).

Osmoregulation mechanisms in E. coli

E. coli possess a number of interconnected systems that allow them to cope with external stress, and grow under a wide range of osmotic pressures (102), to which they respond both passively and actively (102, 103). E. coli whose growth is inhibited, responds passively to hyperosmotic/hypooosmotic shock⁸ by losing/obtaining water. In the case of growing cells, the initial response to sudden hyperosmotic/hypooosmotic shock is also passive, accompanied by changes in intracellular osmolality, that trigger active osmoregula-

⁸ When the external osmotic pressure increases, hyperosmotic shock, water effluxes from the cell’s cytoplasm.

⁹ When the intracellular osmotic pressure is higher than that of the external environment, this so called hypooosmotic shock causes rapid influx of water into the cell.
tion, including elevated synthesis of osmolytes and their transporters (104, 105). Such active osmoregulation allows cell volume and growth to gradually return to normal. In addition, plasmolysis, separation of the cell wall from the inner membrane, can occur in response to hyperosmotic shock (106).

Our model proteins

Our two heavy metal transporters, one from human (human HAH1) and the other from bacteria (*Thermus thermophilus*, bacterial TTHA) are homologous, each consisting of 4 β-strands and two α-helices (Figure 13). Previous reports on *in-cell* NMR studies on these two proteins aided in the design of our approach (86, 107). The functional regions of the wild-type proteins were removed, resulting in scaffolds (TTHA PWT and HAH1 PWT) that undergo two-state folding. Moreover, TTHA PWT, HAH1 PWT, and SOD1 barrel (as well as their mutant forms) are fully folded within cells, ensuring that the broadening of *in-cell* NMR signals reflects primarily the protein’s τ, rather than local chemical exchanges and conformational dynamics (Papers II and III).

Figure 13. Bacterial TTHA and human HAH1 are homologous, each being composed of 4 β-strands and two α-helices (22).
Results and Discussion

To quantify the intracellular diffusive interactions, the rotational movements of three phylogenetically and/or structurally divergent proteins, *i.e.*, bacterial TTHA, human HAH1, and human SOD1\textsuperscript{barrel} inside cells were probed (Figure 14).

Figure 14. The cytoplasmic movements of three evolutionary divergent proteins (bacterial TTHA, human HAH1, and human SOD1\textsuperscript{barrel}) were assessed by *in-cell* NMR (22).

Physicochemical code for diffusive protein interactions in *E. coli*

Bacterial TTHA\textsuperscript{PWT} moves in a relatively unrestricted manner (*i.e.*, its NMR cross-peaks well-resolved with narrow linewidth) in the cytosol of *E. coli*, whereas the motion of human HAH1\textsuperscript{PWT} and SOD1\textsuperscript{barrel} is significantly retarded, resulting in broader NMR peaks (Figure 15). However, this behavior can be altered by changing the surface charges on the proteins: reduction of negative charges restricts the intracellular motion of TTHA\textsuperscript{E58K}, while enhancing their negative surface charges allows HAH1\textsuperscript{K58E} and SOD1\textsuperscript{R100E} to move more freely (Figure 15). These encouraging qualitative results motivated our subsequent investigation of the quantitative relationship between protein mobility inside cells (mobility\textsuperscript{in cell})\textsuperscript{10} and protein surface charge density (charge\textsuperscript{density})\textsuperscript{11}, employing a comprehensive set of surface mutations in live cells.

\textsuperscript{10} The parameter of mobility\textsuperscript{in cell} characterizes the freedom of protein rotation in cells (Papers II and IV). Its value = 1 means that protein rotate freely in *E. coli* cytosol, just like in pure water; when its value = 0 indicates the cross-peaks have broadened out beyond detection (Papers II and IV).

\textsuperscript{11} The parameter of charge\textsuperscript{density}, which is convected from protein electrophoretic mobility on a native gel, quantifies the net charge of a protein (Paper II).
Figure 15. In-cell NMR spectra reveal that bacterial TTHA<sup>PWT</sup> moves relatively freely in the E. coli cytoplasm (well-resolved signals in A), whereas the movement of human HAH1<sup>PWT</sup> and SOD1<sup>barrel</sup> is retarded significantly (the broadened spectra in A). However, this behavior can be reversed readily by altering surface charge (B). The signals in the middle of spectra for HAH1<sup>PWT</sup>, SOD1<sup>barrel</sup> and TTHA<sup>E58K</sup> arise mainly from small metabolites.

Despite considerable scatter in the correlation between mobility<sub>in cell</sub> and charge<sub>density</sub>, these plots demonstrate a linear positive dependence of mobility<sub>in cell</sub> on charge<sub>density</sub> (Figure 16). Interestingly, the mobility<sub>in cell</sub> vs. charge<sub>density</sub> plots for the three proteins differ<sup>12</sup>, indicating that other factors are involved in controlling intracellular mobility. Therefore, we examined the influence of surface hydrophobicity (22).

To quantify the influence of surface hydrophobic on mobility<sub>in cell</sub> more precisely, the area of exposed hydrophobic surface on each protein had to be determined. We hypothesized that all residues other than those in the hydro-

---

<sup>12</sup> The plots for three proteins have different offsets (Figure 16).
phobic core are not firmly “fixed” (108, 109), but can interact freely with the environment upon dynamic rearrangement. Thus, the average of the Guy solvation energies (110) for all residues in the dynamically accessible surface area (DASA) was defined as the hydrophilicity \( G_{\text{uy}} \). Interestingly, the three-dimensional (3D) plot of mobility \( m_{\text{cell}} \) versus charge \( q_{\text{density}} \) and hydrophilicity \( G_{\text{uy}} \) yields a plane that captures the \textit{in-cell} motion of all three proteins (Figure 17). The offsets in the mobility \( m_{\text{cell}} \) versus charge \( q_{\text{density}} \) plots are displaced in an orderly fashion when hydrophilicity \( G_{\text{uy}} \) dimension is added (Figure 17). Furthermore, charged proteins have electric dipole moments \( ||\mathbf{p}|| \) that are expected to affect their motions under crowded intracellular conditions (22, 111). The 3D plot of mobility \( m_{\text{cell}} \) versus charge \( q_{\text{density}} \) and the magnitude of the dipole moment \( ||\mathbf{p}|| \) in the various mutants demonstrates that the impact of \( ||\mathbf{p}|| \) on \textit{in-cell} mobility is similar to that of hydrophilicity \( G_{\text{uy}} \) (Figure 17) (22).

![Figure 17](image)

The observation that the mobility \( m_{\text{cell}} \) data for all three proteins falls onto a single physicochemical plan indicates that despite the intricate organization of \textit{in-cell} diffusive interactions, these interactions still obey certain common and simple physicochemical rules. More specifically, diffusive interactions with the “bulk” cellular environment are semiordered and depend predomi-
nantly on the net surface charge density of the protein, after which the electric dipole moment and area of surface hydrophobicity exposed come into play. Thus, we speculate that the cellular components responsible for retardation of protein movement may carry a repulsive negative charge or/and attractive surface hydrophobicity. Furthermore, the observations illustrated by Figure 17 could help optimize diffusive interactions in connection with molecular searches (81), as well as provide a tool for rational protein design.

Protein mobility in different cellular environments

As described above, proteins from divergent organisms behave quite differently in the cytosol of *E. coli* with respect to their diffusive interactions: the bacterial protein (TTHA<sup>PWT</sup>) moves relatively freely in the bacterial interior, while the human homolog (HAH1<sup>PWT</sup>) tends to get “stuck”. Interestingly, it has been reported that human HAH1<sup>PWT</sup> and SOD1<sup>barrel</sup> tumble freely in mammalian cells (58, 86).

Whether such variations in cytoplasmic mobility stem from the different proteomes of different species or other intracellular constraints remains unclear. To explore this issue further, the rotational movement of three model proteins in *E. coli* and human A2780 cells were compared using *in-cell* NMR relaxation (Paper III). The signals from the human proteins in *E. coli* are so broadened out that they are undetectable, whereas in human cells, these signals are well-resolved (Figure 18), indicating freer motion of proteins with fewer retarding diffusive interactions in the cytosol of mammalian cells. The ratios of the line-widths in the H-<sup>15</sup>N 2D NMR spectra (Figure 19) in lysates and *in-cell* show that although the three proteins differ in size and surface properties, they all behave in much the same manner in mammalian cytosol (Figure 19). HAH1<sup>PWT</sup> (ratio: 0.41 ± 0.06) is most retarded, whereas retardation of TTHA<sup>PWT</sup> (ratio: 0.55 ± 0.07) and SOD1<sup>barrel</sup> (ratio: 0.56 ± 0.15) is less pronounced (Figure 19). This difference in the motional retardation of the homologous TTHA<sup>PWT</sup> and HAH1<sup>PWT</sup> proteins may reflect the influence of surface properties, as observed in *E. coli* (Paper II). Moreover, in the case of *E. coli*, the line-width ratio for TTHA<sup>PWT</sup> is only 0.37 ± 0.04 (Figure 19),<sup>13</sup> revealing an even higher degree of retardation in this environment.

<sup>13</sup>The signals from HAH1<sup>PWT</sup> and SOD1<sup>barrel</sup> are so broadened out in *E. coli* that they cannot be detected and line-width analysis is not possible.
Figure 18. Comparison of the in-cell NMR spectra in A2780 cells and E. coli. The NMR spectra of TTHAPWT (blue), HAH1PWT (red) and SOD1\textsubscript{barrel} (green) in E. coli (bottom panel) and A2780 cells (upper panel) are shown. In E. coli, the signals from the human proteins HAH1PWT and SOD1\textsubscript{barrel} are severely broadened due to a large number of diffusive interactions with the cellular matrix. Upon transfer from E. coli to A2780 cells, all three proteins give rise to well-resolved spectra, but with varying line-widths (Paper III).

Figure 19. Quantification and comparison of line broaden in E. coli and A2780 cells. The ratios of in vitro to in cell line-widths were obtained from \textsuperscript{1}H-\textsuperscript{15}N 2D NMR spectra.

In Paper III, in order to obtain more reliable quantification of intracellular protein motion, we decided to determine the transverse relaxation rate ($R_2$) (112). $R_2$ determines the line-width of an NMR signal, with faster relaxation resulting in a broader signal. However, $R_2$ cannot be utilized as a direct measurement of diffusive retardation within a cell if the proteins being studied have different shapes and sizes (113, 114). To obtain the retardation due to the intracellular diffusive interactions, the apparent viscosity ($\eta_{\text{app}}$) is taken into consideration (Paper III). $\eta_{\text{app}}$ is defined as the microscopic viscosity in a pure solution (in our case, water/glycerol mixtures (115)), in which the $R_2$/line-broadening is the same as observed in a more complex environment (116, 117). The $\eta_{\text{app}}$ of a protein in the cytosol is the sum of the intrinsic microscopic solvent viscosity ($\eta_{\text{int}}$) and all interactions with surrounding macromolecules, where $\eta_{\text{int}}$ takes into accounts differences in the shapes and

\footnote{Quantifying protein rotation on the basis of $R_2$ is preferable to the quantitative procedure employed in Paper II, because the former is independent of the protein concentration in the cell and lysates.}
sizes of proteins, allowing them to be comparable (Paper III). When the relaxation rates $R_2$ for each protein in 0-50 % glycerol were characterised (Figure 20 A), $R_2$ was found to be linearly related to $\eta$ in all three cases. Moreover, the plot of $\delta R_2/\delta \eta$ versus protein size in Figure 20 B is linear. From the plots in Figure 20 A, in-cell $R_2$ can be converted to $\eta_{\text{app}}$ to quantify retardation due to intracellular diffusive interactions (Paper III).

![Figure 20](image)

**Figure 20.** Reference curves for the determination of $\eta_{\text{app}}$. A. The $R_2$ for the three probe proteins in water/glycerol mixtures increases linearly with the microscopic viscosity. B. The slope of the plots in A is directly related to protein size (Paper III).

On the basis of this analysis of $\eta_{\text{app}}$, we explored the influence of protein net charge on diffusive interactions in mammalian cells in greater detail. We discovered that $\eta_{\text{app}}$ is dependent on the net charge on the protein, i.e., the greater the net negative charge, the fewer the number of diffusive interactions (Figure 21 A). It is noteworthy that the plots of protein net charge versus $\eta_{\text{app}}$ are not linear (Figure 21 A). The value of $\eta_{\text{app}}$ approaches that in the water when no intracellular diffusive interactions occur, while $\eta_{\text{app}}$ will be excessively high when the strength of the diffusive interactions is sufficient to block protein rotation fully (Figure 21 A). To compare relaxation in mammalian cells (Paper III) and mobility $^{\text{in cell}}_{\text{in cell}}$ in E. coli (Paper II), the mobility $^{\text{in cell}}_{\text{in cell}}$ values for all mutants were translated into $\eta_{\text{app}}^{15}$, revealing that the $\eta_{\text{app}}$ in E. coli ($\eta_{\text{E. coli}}^{\text{app}}$) is significantly higher than in A2780 cells ($\eta_{\text{A2780}}^{\text{app}}$). Moreover, although net charge density is still an important factor, protein movement in the mammalian cytosol is approximately 6-fold less sensitive to this density than in the cytosol of E. coli (22, 118, 119), as revealed by the slope ($\delta \eta_{\text{E. coli}}^{\text{app}}/\delta \eta_{\text{A2780}}^{\text{app}}$) of the correlation for the protein $\eta_{\text{app}}$ values in these two types of cells (Figure 21 B). On the other hand, the lack of membrane-bound functional compartments in E. coli imposes more restrictions on protein mobility that could ensure tighter control of key cellular processes than that achievable by altering protein surfaces. In summary, these observations indicate that protein movement in the cytosol of E. coli is retarded to a much higher extent than in the mammalian cytosol.

$^{15}$ The detailed conversion method is in Paper III.
Figure 21. The $\eta_{app}$ of our model proteins inside human A2780 cells and E. coli. A. The $\eta_{app}$ is dependent on the protein’s net charge, decreasing as the net negative charge increases. B. The $\eta_{app}$ observed in these two different types of cell is correlated with the slope $\delta\eta_{app}^{E. coli}/\delta\eta_{app}^{A2780}$, which is 6, suggesting that protein rotation is 6-fold more dependent on net surface charge.

Effects of protein size and surface charge on mobility in cell

As stated above, we conclude that protein mobility within E. coli is more restricted than inside mammalian cells. Although this bacterium lacks membrane-bound organelles, it regulates protein mobility by tuning diffusive interactions as demonstrated previously. The mobility of evolutionarily diverse proteins in the same cellular environment, is governed by their physicochemical surface properties (Paper II). Accordingly, we analyzed the average net surface charge density and molecular weight (MW) of the proteomes of human cells and E. coli statistically (Figure 22). The distribution curves in Figure 22 indicate the human protein (34 kDa) appears to be slightly larger than the E. coli protein (25 kDa); while the E. coli protein is slightly more negatively charged (120). This observation that both E. coli and human proteins tend to be negatively charged is in line with our previous findings (22), i.e., in both of these species, the negatively charged proteins in the cytosol move freely by electrostatic repulsions. Moreover, the net negatively charged TTHA$^{PWT}$ is freer to move in human cells than that in E. coli; the smaller the size of the protein molecules encountered, the less the retardation. Therefore, we conclude that the MW and net negative charge of proteins in human cells and E. coli cannot explain the difference in protein rotation.
Figure 22. Characteristics of proteins in the cytosol of E. coli and human cell. Left panel: The net surface charge density of cytosolic proteins in E. coli and human cells. The net charge density in E. coli is more negative than that in human cells, generating stronger electrostatic repulsion. Right panel: The size distribution of cytosolic proteins in E. coli (blue) and human cells (red). The human proteome shifts the distribution peak to larger protein size, as well as a significantly longer tail composed of larger proteins (Paper III).

Effects of intracellular protein concentration on mobility\textsuperscript{in cell}

The value of $\eta_{app}$ in human cells is much lower than the corresponding value in E. coli (Paper III), suggesting less macromolecular crowding and lower collision frequencies in the mammalian cytosol. Indeed, macromolecular crowding in E. coli is more pronounced than in mammalian cells (113, 121, 122), which might explain the difference in retardation of protein movement. To test this proposal, sucrose osmotic shock (Figure 23) was applied to modulate the cytoplasmic volume, without changing the components of the cytosol of E. coli, thereby altering macromolecular crowding. The TTHA$^{\text{pWT}}$ rotational mobility is proportional to osmotic pressure in a sigmoidal fashion (Figure 24) (Paper IV): mobility is retarded when the cell shrinks, whereas retardation weakens as the cell swells, \textit{i.e.}, there appears to be a predictable dependence on protein concentration (Paper IV). Notably, since the expansion of these cells is limited by the elasticity of their peptidoglycan cell wall (123), the mobility$^{\text{in cell}}$ of TTHA$^{\text{pWT}}$ cannot reach the maximum value (1). On the other hand, mobility$^{\text{in cell}}$ plateaus out at a minimum value when the cytoplasmic volume shrinks to a certain level (Paper IV).
Figure 23. Altering intracellular macromolecular crowding in *E. coli* by osmotic shock. Sucrose can pass freely through the cell wall into the periplasmic space, but cannot penetrate the plasma membrane. With high [sucrose], water flows out of the cytoplasm, causing the cell to shrink and plasmolysis to occur (left panel). Hypoosmotic shock was achieved by reducing the [sucrose] so that water molecules flow into the cytoplasm and the cell swells (right panel).

In Paper IV, both the cytoplasmic and total volumes were quantified. The volumes increased by as much as 14 % as the [sucrose] was reduced (Figure 25). When the osmotic pressure was elevated, both of these volumes declined simultaneously, but to a different extent, both eventually becoming 50 % smaller (124, 125), a limit probably imposed by intracellular macromolecules (2) and water (98). The plot of mobility$_{\text{in \ cell}}$ as a function of cytoplasmic volume revealed a correlation (Figure 25). In this case, the data for HAH$_1^{\text{R21E/K25E}}$ and SOD$_1^{\text{R100E}}$ in the cytosol of *E. coli*, where their movement is less retarded were used since the *in-cell* signals observed for human HAH$_1^{\text{PWT}}$ and SOD$_1^{\text{barrel}}$ were of poor quality (22). When the cytoplasmic volume expanded by 14 %, the mobility$_{\text{in \ cell}}$ values for TTHA$_{\text{PWT}}$ and HAH$_1^{\text{R21E/K25E}}$ rose by 30 %, while that for SOD$_1^{\text{R100E}}$ increased 70 % (Figure 25). When the cytoplasmic volume contracted by 20 %, the mobility$_{\text{in \ cell}}$ values of HAH$_1^{\text{PWT}}$ and SOD$_1^{\text{barrel}}$ become minimal (0); whereas the cytoplasmic volume had to be reduced by 30 %, in order for the three other proteins to reach their minimum mobility$_{\text{in \ cell}}$ values (0). These observations reveal that modu-
lation of protein motion by macromolecular crowding is on the nature of the specific protein involved (22).

Altogether, the distinct differences in motional retardations in *E. coli* and human cells appear to be caused by differences in physiological protein concentrations. The relatively free movement of TTHA<sub>PWT</sub> in both bacterial and mammalian cells suggests that its surface has been optimized to withstand large numbers of collisions, avoiding capture in short-lived non-productive complexes. In contrast, human proteins appear to have evolved to maintain motion in their own natural environment, where macromolecular crowding is less extensive; therefore, their motion becomes more retarded in the cytosol of *E. coli* with higher level of macromolecular crowding. Consequently, the mobility of bacterial proteins is less affected by changes in intracellular crowdedness than that of human proteins.

![Figure 25](image.png)

**Figure 25.** A. Quantification of the volume of *E. coli* under conditions of sucrose osmotic shock. B. The relationship between the volume of the *E. coli* cytoplasm and mobility<sub>n</sub> of our model proteins. Protein mobility<sub>n</sub> rises in a sigmoidal fashion as this volume expands. The mobility<sub>n</sub> decreases as the volume is reduced until a lower limit is reached, reflecting hindrance of rotation by enhanced crosstalk of diffusive interactions.

This difference in the cytosolic concentration of macromolecules in *E. coli* and eukaryotes influences not only protein rotation, but also the transverse diffusion of proteins. More specifically, the diffusion of free green fluorescent protein (GFP) in *E. coli* is three times slower than in human cells (126). Essentially, the crowding effect increases effective biomolecular concentrations by modulating the volume of solvent available, which exerts a significant impact on biochemical equilibria (77, 127–130). With enhanced intracellular crowdedness (*i.e.*, more diffusive interaction partners), each protein molecule will undergo more diffusive collisions, which result in even stronger crosstalk of diffusive interactions within *E. coli*. Therefore, protein movement inside *E. coli* is more challenging than in mammalian cells.
Summary

First, three model proteins and nearly 130 variants with surface mutations were utilized to examine diffusive interactions, and, consequently, protein rotation in the cytosol of *E. coli*. Protein mobility is related to physicochemical surface properties, including (i) net surface charge density; (ii) the density of hydrophobic side-chains at the surface; and (iii) the global dipole moment (Paper II). Net surface charge has the strongest effect: a high net negative surface charge enables relatively free movement, while with a neutral or net positive surface charge, protein movement in the cytosol is restricted (Paper II). Furthermore, rotational mobility also depends critically on the type of organism in question; human proteins that “get stuck” in bacteria appear to move freely in the cytosol of human cells (Papers II and III). To explore the underlying mechanisms, the influence of intracellular crowding on diffusive interactions was tested by modulating the volume of the *E. coli* cytoplasm (Paper IV). Intracellular macromolecular crowding not only alters the strength of the steric repulsive interactions but also tunes the frequency of random diffusive collisions between molecules (*i.e.*, the concentration of diffusive interaction partners) (Papers III and IV).

Intracellular crowding is of considerable physiological importance. Lacking organelles, *E. coli* must ensure that its life processes proceed with fidelity and efficiency in some other way. Probably through diffusive interactions modulated by the surface properties of cellular components and macromolecular crowding. In the case of human cells, with their membrane-bound organelles, less extensive cytosolic crowding may represent an evolutionary advantage, rendering them more tolerant of mutations that alter the surface properties of proteins. These findings together with those presented in Part I, complement our understanding of intracellular crowding at the physicochemical level: intracellular crowding involves both the excluded volume effect and diffusive interactions, which work together to regulate the molecular recognition and the functional stability of proteins.
Outlook

The properties of intracellular environment are usually determined by a variety of physicochemical parameters, including ionic strength, viscosity, osmotic pressure, the activity of water, pH and macromolecular crowding, etc. Among these parameters, the intracellular ionic strength plays an important role in controlling both specific and non-specific interactions between biomacromolecules (130, 131). This intracellular ionic strength can vary widely in response to fluctuations in the external environment (e.g., osmotic pressure), making it challenging to quantify this parameter. In general, simply using 150 mM NaCl in vitro does not accurately mimic the effective ionic strength in live cells, since Na$^+$ and Cl$^-$ move freely in diluted aqueous solutions. whereas in live cells most of the anions (i.e., the negatively charged groups) are bound covalently to the surface of macromolecules, resulting in a relatively low concentration of free Cl$^-$ (< 30 mM) (132).

Diffusive interactions are highly sensitive to environmental fluctuations (as we found in Paper IV) for rapid fine-tuning of cellular processes (133). The work in Paper II demonstrated that the charges on protein surface play a key role in the modulation of diffusive interactions, then a related concern is how the intracellular environment regulates the electrostatic interactions by modulating the concentration of ions. To this end, we will attempt to investigate the influence of ionic strength on diffusive interactions in live cells.
Part III: Effects of salt on protein folding

(Paper V: Salt effects on protein folding behavior: revealing the molecular origins and limitation of Hofmeister theory.)

Introduction

Many salt ions are involved in controlling key biological functions, ranging from osmoregulation and signaling to the specific roles of certain cations in the activities of metalloproteins (73, 134). Consequently, the effects of salts on protein solubility, stability, assembly, self-association, and biological activity have been of considerable interest for decades (135–137). These effects are of salt concentration ([salt]) dependent. At low concentration, a non-specific effect on electrostatic interactions in dielectric medium (138) referred to as the Debye screening effect, is expected to be most significant (139–142). However, at high [salt], the Debye screening effect cannot account for protein salting-in and salting-out, which are instead explained on the basis of the empirical Hofmeister scale (137). Although salts differ in their abilities to influence the thermodynamics and kinetics of protein folding (143), most of their effects are consistent with the Hofmeister series, the molecular nature and origin of which remain controversial (144, 145). In general, there are two points of view concerning the effects of ions on protein salting-in and salting-out behaviors. The traditional view is that the Hofmeister effect is indirect, mediated by generalized changes in the structure of water (146–149), while the other emerging perspective insists that direct interaction of ions with proteins is the dominant factor (150–153).

The complexity of the physicochemical properties of both salt ions and proteins, make it challenging to explain the modulation of protein behavior by salt ions completely. Therefore, to understand the molecular mechanisms underlying the Hofmeister theory, we quantified the effects of ions on the folding and stability of our model proteins. We also investigated the relevance of the Hofmeister series to ionic hydration and perturbation by specific ion-pairing.
The Hofmeister series

The Hofmeister series orders ions on the basis of their ability to affect protein solubility (protein salting-in or salting-out) (154). Hofmeister, the person established this, tried to relate the series with ionic hydration capacity (137). More specifically, different salts withdraw water molecules from a solubilized protein to a different extent, causing it to precipitate (137). Eventually, the Hofmeister series was incorporated into the Kosmotropes/Chaotropes classification (144, 155). The potent hydration of Kosmotropes allows them to remove water from proteins, whereas weakly hydrated chaotropic cannot (156, 157). However, this explanation of the Hofmeister series faces recent challenges, since the consensus is that salting-in and salting-out cannot be fully explained by the bulk properties of water but are more closely related to specific ion-protein interactions (144). Accordingly, binding of ions to the major building blocks on the protein surface (i.e., the backbone and charged, hydrophobic and polar amino acid side-chains) have been extensively characterized (144). In addition to interacting with proteins and water molecules, oppositely charged ions with comparable hydration free energy tend to pair in water, weakening or even eliminating the effect of each individual ion (158, 159). Furthermore, investigation of salt effects in non-aqueous solvents has revealed the same basic ion-specific trends in all solvents, suggesting that the Hofmeister effect originates specifically from the ions themselves, rather than from their overall hydration (160, 161).

Protein solubility

Protein solubility is defined as the protein concentration at which the chemical potentials of the protein molecules in the liquid and solid phases are equal (in equilibrium) under any given ambient conditions (162). It is a physicochemical property, which is determined by protein amino acid sequence and ambient conditions (163). Protein solubility in an electrolyte solution is determined by a variety of interactions, i.e., protein-protein, protein-water, protein-ion, and ion-water interactions (164, 165).

Precipitation with ammonium sulfate is the most reliable procedure for determining protein solubility, providing accurate informations rapidly (164, 165). When precipitated in this manner, the protein is in an amorphous form and can be redissolved in the liquid phase by reducing [(NH₄)₂SO₄]. Moreover, the structure of the protein is not affected by ammonium sulfate precipitation (166). The logarithm of protein solubility normally declines linearly (Equation 19) with increasing [(NH₄)₂SO₄] (159):

\[
\log \text{protein solubility} = -k \cdot [(\text{NH}_4)_2\text{SO}_4]
\]
\[
\log s = \log s_0 - K_s[\text{NH}_4\text{SO}_4]_s
\]
(Equation 19)

where \(s\) is protein solubility, \(\log s_0\) is the logarithm of solubility at 0 M [salt], and \(K_s\) is a salting-out constant that is independent of pH and temperature, but dependent on the natures of both the salt and protein. Generally, salts with large Hofmeister effects exhibit a high \(K_s\) (159).

Screening electrostatic interactions

Electrostatic interactions can be represented simply by Coulomb interactions, the strengths of which in aqueous solution are almost always attenuated by nearby water molecules and ions. In a dielectric medium, the Coulomb interaction is reduced by a dielectric constant \(\varepsilon_r\), equal to the ratio of the permittivity of the medium to that of vacuum \(\varepsilon_0\). This constant indicates the extent to which a system concentrates electric flux and is thus proportional to the electric flux density. For example, at room temperature, the \(\varepsilon_r\) of pure water is 78.5, which means that the strengths of Coulomb interactions are reduced almost 80-fold compared to the strength in a vacuum. This reduction in interaction strength in a dielectric medium is sometimes referred to as charge “screening”. The Debye screening length \(l_D\) is helpful in connection with attempts to understand the screening of electrostatic interactions by free ions in solution. This \(l_D\) is the characteristic distance across with the net electrostatic effect of a charged molecule in solution is felt (167). The charged surfaces of bio-macromolecules trap counterions to form a charged layer of ions (approximately \(l_D\) in width) that can neutralize those surface charges and strongly attenuate their long-range electrostatic interactions (168). Such screening of electrostatic interactions is also referred to as the Debye screening effect.

Formation of water surface tension at the molecular level

The surface tension of water arises from its propensity to contract as much as possible at the liquid-air interface, allowing the surface to resist external forces. In the bulk phase, cohesive forces between the water molecules are exerted equally in all directions, resulting in null net force (Figure 26). However, water molecules locate at the liquid-air interface are not surrounded by other water molecules on all sides, and, therefore, cohesive forces cannot be exerted in all directions. An inward pulling force \(F\) thus arises at this surface (Figure 26), creating internal pressure that shrinks the surface area to a minimum, thereby forming a stretched elastic film (169). The consequenc-
es of water surface tension are visible everywhere in nature, e.g., in the formation of water droplets (Figure 26) and water striders on the surface of a pond (Figure 26). This tension is represented by $\gamma$, defined in terms of force per unit length (newton per meter, N/m), and depends on the intrinsic properties (e.g., composition and temperature) of the aqueous solution (170).

![Figure 26. A. Two examples of the consequences of water surface tension in nature. The upper panel shows the formation of water droplets and the bottom one a water strider on the surface of a pond. B. Formation of water surface tension at the molecular level. Water molecules tend to be attracted to other water molecules in their vicinity. In the bulk liquid phase, these cohesive forces act in all directions, resulting in a net attractive force (F) equal to 0. However, a water molecule located at the water-air interface, is pulled downward due to the absence of water molecules above it, creating a strongly bonded elastic layer on the water surface.]

The effects of ion-water interactions on water surface tension

Ions are classified as makers (Kosmotrope) or destroyers (Chaotropes) of water structure on the basis of their ability to hydrate (157), which is usually related to their charge density (171), (i.e., the ratio of ionic charge to volume). In the periodic table, the charge density in a group decreases from top to bottom, since the charge remains the same, but the size increases; while the charge density increases from left to right in a period, since the charge increases, but size decreases (e.g., Mg$^{2+}$ and Na$^+$) (172).
Table 1. Charge densities (C mm⁻³) of selected ions. Charge densities are calculated as \( \frac{ne}{4\pi r^3} \), where \( r \) is the six-coordinated ionic radii in millimeter (173), \( e \) is the electron charge \((1.60 \times 10^{-19} \text{ C})\) and \( n \) is the ionic charges.

<table>
<thead>
<tr>
<th>Ion</th>
<th>Charge density (C mm⁻³)</th>
<th>Ion</th>
<th>Charge density (C mm⁻³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ba²⁺</td>
<td>23</td>
<td>Zn²⁺</td>
<td>112</td>
</tr>
<tr>
<td>Cs⁺</td>
<td>6</td>
<td>Br⁻</td>
<td>6</td>
</tr>
<tr>
<td>Ca²⁺</td>
<td>52</td>
<td>CO₃²⁻</td>
<td>17</td>
</tr>
<tr>
<td>NH₄⁺</td>
<td>11</td>
<td>Cl⁻</td>
<td>8</td>
</tr>
<tr>
<td>Na⁺</td>
<td>24</td>
<td>F⁻</td>
<td>24</td>
</tr>
<tr>
<td>K⁺</td>
<td>11</td>
<td>I⁻</td>
<td>4</td>
</tr>
<tr>
<td>Li⁺</td>
<td>52</td>
<td>NO₃⁻</td>
<td>9</td>
</tr>
<tr>
<td>Mg²⁺</td>
<td>120</td>
<td>SO₄²⁻</td>
<td>5</td>
</tr>
</tbody>
</table>

Ions can affect the structure of water by interacting with water molecules, thereby attenuating water-water interactions. The balance between ion-water and water-water interactions determines the structure of water molecules in a hydration shell. Small ions (Kosmotropes) with high charge density cause strong electrostatic ordering of surrounding water molecules, thereby eliminating hydrogen bonding between water molecules. In contrast, large ions (Chaotropes) with a low charge density allow such hydrogen bonding to dominate the structure of water (157). Moreover, in general, anions order the structure of water more extensively than cations, since the negative end of the dipole of a water molecule is closer to its center than the positive end, i.e., charge distribution is asymmetrical (159).

\( \delta \gamma / \delta [\text{salt}] \), the slope of \( \gamma \) versus [salt] plot, is generally positive (174), with higher values indicating stronger ion-water interactions (Table 2). Salt ions displace water molecules from the air-water interface, thereby increasing \( \gamma \), the magnitude of which depends on the extent of hydration (174). Although the hydration of ions is generally consistent with the Hofmeister series (175–178), the correlation is not perfect. For example, although Ca²⁺ and Mg²⁺ are extensively hydrated, but cause a salting-in effect at high concentrations (Table 2). Salts with higher \( \delta \gamma / \delta [\text{salt}] \) values (e.g., sulfates) favor the reduction of protein SASA during folding and aggregation; whereas smaller \( \delta \gamma / \delta [\text{salt}] \) values (e.g., Gdm⁺) favor exposure of protein SASA (Table 2). There are often large differences in the \( \delta \gamma / \delta [\text{salt}] \) of salts with a common cation (or anion) and the paired anions (or cations) from the poles of the Scheme 2, e.g., the \( \delta \gamma / \delta [\text{salt}] \) values for Na₂SO₄ and (Gdm)₂SO₄ are 3 ± 0.07 and 0.9 ± 0.16, respectively.
Table 2. Partial values of $\delta_\gamma/\delta_{[\text{salt}]}$ from (174, 179, 180).

<table>
<thead>
<tr>
<th></th>
<th>Na$^+$</th>
<th>K$^+$</th>
<th>Li$^+$</th>
<th>NH$_4^+$</th>
<th>Cs$^+$</th>
<th>Gdm$^+$</th>
<th>Mg$^{2+}$</th>
<th>Ca$^{2+}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SO$_4^{2-}$</td>
<td>3 ± 0.07</td>
<td>2.6</td>
<td>3.1</td>
<td>2.3</td>
<td>3 ± 0.16</td>
<td>0.9 ± 0.05</td>
<td>2.6</td>
<td>3 ± 0.07</td>
</tr>
<tr>
<td>Cl$^-$</td>
<td>2 ± 0.07</td>
<td>1.7</td>
<td>1.7</td>
<td>1.5</td>
<td>1.6</td>
<td>0.8 ± 0.11</td>
<td>3.8</td>
<td>4.1</td>
</tr>
<tr>
<td>Br$^-$</td>
<td>1.8 ± 0.04</td>
<td>1.4</td>
<td>1.3</td>
<td>1.3</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO$_3^-$</td>
<td>1.2 ± 0.01</td>
<td>1.1</td>
<td>1.2</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I$^-$</td>
<td>1.1 ± 0.09</td>
<td>1.2</td>
<td>0.8</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Scheme 2. The relative hydration strengths of selected cations and anions (174, 179, 181).

Our model proteins

We employed ribosomal protein S6 from *Thermus thermophilus* to examine the effects of salt on protein folding and unfolding for the following reasons: First, like bacterial TTHA and human HAH1, this $\alpha/\beta$ protein is composed of four $\beta$ sheets and two $\alpha$ helices (Figure 27). Furthermore, it carries numerous charges, 16 positively charged and 16 negatively charged residues, altogether constituting 32% of the entire sequence (Figure 27). In addition, S6 shows a strong pH resistance, allowing modulation of surface charge by pH titration. And last but not least, S6 demonstrates two-state folding (182).

Figure 27. Like bacterial TTHA and human HAH1, S6 is an $\alpha/\beta$ protein with 16 positive (blue) and 16 negative (red) charges on its surface.
When refolding and unfolding of our model proteins in the presence of increasing [salt] were monitored with a rapid mixing stopped-flow technique, the logarithm of the relaxation rate \( \log k_{\text{obs}} \) was found to be linearly related to [salt] (Figure 28). \( \delta(\log k_{\text{obs}})/\delta[\text{salt}] \) (also called the m-value), which reflects the \( \Delta \text{SASA} \) during protein folding and unfolding (Equations 20), provides an ideal measure of the effects of various salts with respect to salting-out (stabilization) and salting-in (destabilization).

\[
\log k_{\text{obs}} = \log k_{\text{obs}}^{\text{H}_2\text{O}} + m[\text{salt}]
\]  

(Equation 20)

where \( \log k_{\text{obs}}^{\text{H}_2\text{O}} \) is the \( \log k_{\text{obs}} \) at 0 M [salt], \( m \) is the m-value.

In connection with data analysis, the nomenclature of parameters and definition of their corresponding physical properties are important (Figure 28). The log \( k_{\text{obs}} \) of the refolding and unfolding curves is linearly dependent on [Na\(_2\)SO\(_4\)], with slopes of \( m_f^+ \), \( m_f^- \) and \( m_u^- \), respectively. More specifically, the \( m_f^+ \) for stabilizing salts (e.g., sulfates and phosphates) is positive, since these favor the folded state and thus accelerate refolding \( (k_f) \). Conversely, this value is negative for destabilizing salts (e.g., GdmCl), which shifts the equilibrium to the denatured state, slowing down \( k_f \). Conversely, the \( m_u^+ \) is negative (slow down the \( k_u \)) for stabilizing salts (Figure 28) while positive (speed up the \( k_u \)) for destabilizing salts. The value of \( m_f^- \) is negative for both stabilizing and destabilizing salts. We demonstrated previously that the negative \( m_f^- \) for stabilizing salts is due to the formation of off-pathway intermediates at relatively high concentrations, slowing down \( k_f \)\(^{(182, 183)} \). It is worth noting the absolute m-values for salts with similar properties represent their efficacy at stabilizing or destabilizing proteins (i.e., the larger this value, the greater the efficacy). In addition, \( M_p(k_f) \) indicates the [salt] at which the limb of refolding curve begins to rollover.
Figure 28. Nomenclature of the related parameters ($m^+, m^-, m^-_{u}$ and $M(p(k))$) involved in titrated refolding and unfolding with stabilizing salts. The upper panel illustrates the typical chevron plot, with refolding and unfolding limbs. The titration curve is obtained by titrating protein refolding/unfolding by salt at varying concentrations.
Results and Discussion

The Debye-screening effect at low [salt]

We previously reported that the $k_f$ of refolding of S6 rises with increasing $[\text{Na}_2\text{SO}_4]$ (182, 183) (Figure 29). Intriguingly and conversely, the $k_f$ of SOD1$^{\text{barrel}}$ is retarded in the presence of 0-0.04 M $[\text{Na}_2\text{SO}_4]$ (Figure 29) (182, 183). We propose that this phenomenon reflects the generic Debye screening effect at low [salt], which reduces protein stability by screening intramolecular electrostatic interactions (Paper V). When the unfolding of SOD1$^{\text{I35A16}}$ was titrated with $\text{Na}_2\text{SO}_4$, $k_u$ increased slightly (Figure 29). Because stabilizing electrostatic interactions within the native structure were screened, thereby reducing the stability of this protein. Indeed, these electrostatic interactions were screened fully by 0.04 and 0.02 M $\text{Na}_2\text{SO}_4$ for refolding and unfolding respectively. Unexpectedly, there was no the Debye screening effect upon titration of the curves for S6 with salt (Figure 29). Unlike urea, the more potent denaturant guanidine hydrochloride (GdmCl), used to unfold the conformationally stable S6, is itself a salt and thus alters the ionic strength (182).

![Figure 29. The effects of Na$_2$SO$_4$ on the refolding and unfolding of S6 (A) and SOD1$^{\text{barrel}}$ (B).](image)

$^{16}$ To avoid the usage of a large amount of denaturant, the marginally stable variant of SOD1$^{\text{barrel}}$ (SOD1$^{\text{I35A}}$), with similar folding behaviour, was used for our kinetic studies of the effects of salt on unfolding.
The relevance of $M_p(k_f)$ to protein solubility

The details concerning refolding of our two model proteins differ (Figure 30). Specifically, rollover for S6 occurs with 0.2 M [Na$_2$SO$_4$], but at 1 M in the case of SOD$_1^{\text{barrel}}$ (Figure 30). In other words, the refolding curve for S6 exhibits a long downward and negligible upward limb, while the corresponding curve for SOD$_1^{\text{barrel}}$ has a longer upward and less pronounced downward limb (Figure 30). Moreover, a shift of $M_p(k_f)$ to a higher [Na$_2$SO$_4$] was observed for the extremely negatively charged variant of S6, as well as in the presence of a high [GdmCl] (182). These phenomena suggest a correlation between $M_p(k_f)$ and protein solubility. To verify this, the solubility of S6 was adjusted utilizing pH titration to protonate/deprotonate charged amino acid side-chains (182), resulting in a linear relationship ($R^2 = 0.92$) between $M_p(k_f)$ and protein solubility (log $s_0$) (Paper V). Thus, more soluble proteins have higher $M_p(k_f)$ values, and, consequently, less tendency to form aggregate. Indeed, SOD$_1^{\text{barrel}}$ is more soluble (400 mg/ml) under quiescent conditions than S6 (100-200 mg/ml). This observation indicates further that the salt effect on protein folding depends on physicochemical properties of the particular protein itself.

![Figure 30. $M_p(k_f)$ exhibits a positive dependence on protein solubility (log $s_0$).](image)

Ionic hydration cannot rationalize the Hofmeister effect fully

This study shows that higher [salt] enhances the Hofmeister effect significantly and that the m-value is a valid measure of the extent to which an ion can stabilize (salting-out) or destabilize (salting-in) proteins (Paper V). In this respect, both cations and anions follow the standard Hofmeister series. When we explored the relationship between ionic hydration and the Hofmeister effect on the folding-unfolding of proteins, the $\delta y/\delta [\text{salt}]$ values (a measure of the capacity for ionic hydration (174, 179, 180)) were found to be linearly correlated to their m-values for monovalent cations (i.e., CaCl$_2$ and MgCl$_2$), with
an extensive capacity for hydration, deviate in this context from other salts, tending to destabilize at high concentrations (Figure 31). This disagrees with the interpretation of the Hofmeister effect based on the Kosmotropes/Chaotropes classification, since, indeed, Mg$^{2+}$ and Ca$^{2+}$ are “Kosmotropic” cations that cause salting-in according to the Hofmeister series. However, the recent consensus is that salting-in and salting-out cannot be explained fully by the capacity for ionic hydration, but are instead, more closely related to specific ion-protein interactions (144). Furthermore, highly hydrated divalent cations associate more strongly with carboxylate sites on biomacromolecules, as well as carbonyl groups on the amide backbone of proteins than weakly hydrated monovalent cations (144, 150, 184), which could explain the unexpected behaviour of CaCl$_2$ and MgCl$_2$, including their destabilizing effects.

![Figure 31. Correlations between $\gamma/\delta[\text{salt}]$ and the m-values ($m_\ell$ and $m_u$) for SOD1$^{\text{barrel}}$ and SOD1$^{\text{I35A}}$. $m_\ell^*$, $m_u^*$ and $m_u$ of SOD1$^{\text{barrel}}$ and SOD1$^{\text{I35A}}$ are all linearly dependent on $\gamma/\delta[\text{salt}]$ with the exceptions of $m_u^*$ (purple) and $m_u$ (green) for the divalent cationic salts MgCl$_2$ and CaCl$_2$.](image-url)

**Strong ion pairing by Gdm$^+$ and SO$_4^{2-}$ at high concentrations**

To examine whether the cations and anions in an electrolyte solution exert independent effects on protein stability, we investigated the refolding and unfolding of S6 in the presence of Na$_2$SO$_4$, NaCl, (Gdm)$_2$SO$_4$ and GdmCl. Of these four ions, Na$^+$ and Cl$^-$ are generally considered “neutral”, SO$_4^{2-}$ as strongly stabilizing and Gdm$^+$ as strongly destabilizing. The refolding curves in the presence of Na$_2$SO$_4$ and NaCl almost have the shape of an inverted V (Figure 32); this curve is linear, with a negative $m_\ell$ value, in the case of GdmCl (Figure 32); and, unexpectedly, (Gdm)$_2$SO$_4$ destabilizes S6 up to a concentration of 1.8 M (Figures 32 and 33), and at higher levels begins to stabilize S6, as do NaCl and Na$_2$SO$_4$, but all with relatively weak potencies (Figures 32 and 33). The question then raised as to whether the weak effect
of (Gdm)$_2$SO$_4$ on the stability of S6 (Figure 32) stems simply from the balance between stabilization by SO$_4^{2-}$ and destabilization by Gdm$^+$ or are other molecular mechanisms in play? At high concentrations, both Gdm$^+$ and SO$_4^{2-}$ are strongly ion-pairing and almost all of these ions are clustered, thereby attenuating their effects on protein stability (185, 186). Moreover, ion-pairing at high concentrations affects the properties of the electrolyte solution in general (161, 185, 186). Therefore, the increased stability of S6 at high concentration of these ions probably reflects superposition of the effects of ions not involved in ion-pairing, i.e., the stabilizing effect of free SO$_4^{2-}$ is stronger in this condition. In summary, the effect of ion-pairing on electrolyte properties at high [salt] cannot be ignored when attempting to explain the effects of specific ions. Furthermore, attenuation of these effects by ion-pairing reveals the limitation of the Hofmeister series, which considers cationic and anionic effects separately (185, 186).

However, when the first limbs of the four curves in Figure 32 A were integrated in Figure 32 B, the Hofmeister efficacies of Na$^+$ and Gdm$^+$, Cl$^-$ and SO$_4^{2-}$ could be determined separately (Figure 32 B). For example, the differences between m$\ell$(Na$_2$SO$_4$) and m$\ell$((Gdm)$_2$SO$_4$) and between m$\ell$(NaCl) and m$\ell$(GdmCl) were 1.59 and 1.41, respectively. These similar $\Delta m$ values at low [salt] indicate that the cationic and anionic effects of these salts on the $\Delta$SASA of S6 are largely additive, which is in line with the Hofmeister series. Thus, with a low concentration of (Gdm)$_2$SO$_4$, the destabilizing effect of Gdm$^+$ is stronger than stabilization by SO$_4^{2-}$. Taken together, these discoveries reveal that the salt effect is concentration-dependent: there is no significant Gdm$^+$-SO$_4^{2-}$ pairing at low [(Gdm)$_2$SO$_4$] and, thus, these ions exert independent impacts on the stability of S6; whereas at high[(Gdm)$_2$SO$_4$], Gdm$^+$-SO$_4^{2-}$ pairing is significant and the traditional Hofmeister theory can no longer explain the effects observed.
Figure 32. A. Refolding of S6 in the presence of NaCl, Na$_2$SO$_4$, GdmCl or (Gdm)$_2$SO$_4$. B. Superposition of the titration curves shown in A at low [salt], i.e., the upward limbs of the curves with Na$_2$SO$_4$ and NaCl and the destabilizing limbs for (Gdm)$_2$SO$_4$ and GdmCl (Paper V).
Figure 33. Chevron plots and corresponding energy diagrams of S6 in the presence of (Gdm)_2SO_4. A. Refolding (left panel) and unfolding (right panel), as well as the corresponding chevron plots (middle panel) in the presence of varying concentrations of (Gdm)_2SO_4. B. The energy profiles of the chevron plots shown in A. (Gdm)_2SO_4 exerts a weak effect on the conformational stability of S6, as indicated by the log K^{D-N}_{ON} even though SO_4^{2-} and Gdm^+ are strongly stabilizing and destabilizing, respectively. Even though the impact of (Gdm)_2SO_4 on the kinetics of S6 unfolding is negligible, the effect on m_u is more pronounced than that on m_f.
Conclusions

First, these observations indicate that salts influence the salting-in/salting-out of proteins in a variety of ways, including via the Debye screening effect at low concentrations and the dominant Hofmeister effect at high concentrations. Secondly, the Kosmotropes/Chaotropes classification (144, 155) cannot explain the basis for the Hofmeister series fully and more attention must be focused on specific ion-protein interactions in connection with Hofmeister effect on protein salting-out and salting-in (144). Third, strong hetero-ion pairing for specific salts, which can largely eliminate the effects of the individual ions involved, cannot be ignored when attempting to understand effects of salts. Finally, the Hofmeister effect is dependent not only on the nature of the salt, but also on the identity of the protein involved. Consequently, no unifying theory can explain the salt effects. To understand these effects on the thermodynamic behavior of proteins, the physicochemical properties of all participants (i.e., the ions involved, the target protein, and the solvent), as well as the interactions between these must be taken into consideration (144, 161). Although the generic Hofmeister series appears powerful, its categorization of ions based on assumptions concerning underlying molecular mechanisms are only approximations, and a more detailed systematic investigation of the effects of specific ions on the biological functions of macromolecules is absolutely necessary.

Future perspectives

To understand the effects of salts on protein salting-out and salting-in, much attention has been paid to specific interactions of cations and anions with groups on the protein surface (144, 178). The effects observed do, indeed, involve all of the local interactions of all components of the system. In the context of ionic interactions, the protein surface can be divided into several major building blocks, namely, the backbone and charged, hydrophobic and polar amino acid side-chains. The different effects of salts on the stability of our model proteins S6 and SOD1-barrel, with distinct surface characteristics, provide a rationale for the complexity of the Hofmeister series. To further explore the mechanisms by which interactions between salts and various groups on the protein surface disturb or even reverse this series, we will systematically investigate the interactions of charges (i.e., purely positive or negative) and hydrophobic groups (i.e., no charges) on the surface of proteins with a variety of salts.
Populärvetenskaplig Sammanfattning


En enklare typ av celler, och därigenom något lättare att hantera och studera, är bakteriecellen. En av de allra vanligaste cellmodellsystemen är också just

Vidare kunde vi observera att proteinerna påverkas något annorlunda i mänskliga celler jämfört med i bakterierna, även om den generella trenden är liknande. Detta kan härledas till en lägre koncentration av proteiner och andra molekyler i de mänskliga cellerna. Den totala koncentrationen i *E. coli* har uppskattats till 300–450 mg/ml (allså upp till 40 % av tillgänglig volym) och från laddningsberoendet på rotationsdiffusionen i mänskliga celler kunde vi uppskatta koncentrationen i dessa till ca 50–75 mg/ml. Kopplingen mellan den totala koncentrationen makromolekyler och rotationsdiffusion kunde vi verifiera och kvantifiera genom att aktivt öka eller minska volymen på *E. coli* celler och bestämma effekten på rotationsdiffusionen av våra modellproteiner.

Trots att miljön inne i *E. coli* celler är trängre och mer tätpackad än t.ex i humana celler, så kan bakteriella proteiner alltså diffundera relativt fritt där. Men, den trånga miljön i bakterien gör också att små förändringar i ytegenskaper kan ge massiva följder, vilket i sin tur utgör ett kraftfullt evolutionärt verktyg: Även små förändringar kan leda till att nya interaktionsvägar öppnas, där priset är att ’fel’ förändringar leder till att proteinet inte kan dif-
fundera alls i cellen. På motsvarande sätt ger den lägre koncentrationen i humana celler ett större spelrum för evolutionära föränderingar på ytan av proteinet, utan att det riskerar att förlora diffusibilitet. Resultaten som presenteras i den här avhandlingen kan ge ledtrådar till hur molekylär igenkänning fungerar också i komplexa miljöer, hur dessa miljöer påverkar proteiners grundläggande egenskaper som stabilitet och diffusibilitet samt ge insikter i hur proteiner ifrån olika organismer optimeras till rådande intracellulär miljö.
Acknowledgements

I would like to extend my sincere gratitude to the following individuals:

My supervisor, Mikael Oliveberg, for providing me with the opportunity to work and study in your research group. You are certainly one of the most brilliant scientists I have ever met. Thank you for your support, inspiration, and encouragement, as well as for allowing me considerable freedom while doing this research. This thesis could not have been finished without your expertise in biophysics, excellent guidance, critical suggestions concerning experiments and valuable interpretation of data. I hope that I can become as good a scientist as you in the future.

My co-supervisor for my PhD studies, Jens Danielsson, for introducing me to the world of NMR and helping me meticulously with the first and second projects. You can always explain research problems in a way that is most easy to understand. You provided me with precious and efficient methods for data analysis, which have really benefited me a lot!

Seongil Choi, previous postdoc in our group, for being such a great advisor whenever I encountered problems. You have been very patient with me and your suggestions are better than great! I wish you all the best in the future.

Lisa Lang, Huabing Wang, Yang Fan, Sarah Leeb, Therese Grönlund, Eloy Vallina Estrada and Mingming Xu for being excellent of my research partners.

All past and present members of the Mikael Oliveberg group for all kinds of help and support.

Håkan Wennerström, Philipp Selenko, Derek T. Logan, Nikolay V. Dokholyan, David Mowray, Andres Binolfi, Francois-Xavier Theillet, Beata Bekei, my coauthors outside of this group, for their invaluable contributions to these projects.

Daniel Daley, for agreeing to be my co-supervisor and for helpful discussions. Pia Ådelroth, Stefan Nordlund, Gunnar von Heijne, Lena Måler, Martin Högbom and Pia Harryson for helping and supporting me during my PhD, so that they could be completed successfully.
Mikael Oliveberg, Jens Danielsson, Joseph W DePierre, Pia Ädelroth, Bin Yang, Eloy Vallina Estrada, Sarah Leeb, Therese Grönlund for proof-reading this thesis and providing valuable comments.

Bin Yang, for helping me draw the cover to and two sketches involved in this thesis.

Jens Danielsson and Therese Grönlund, for helping me translate the popular science summary.

Maria Sallander Hellman, Johansson Elisabeth, Alexander Tuuling, Astlind Torbjörn, Bennett Matthew, Ann Nielsen, Malin Kamb, Filip Malec, Richard Risberg, Charlotte Andersson, Kristina Linkyte, Karin Häggbom Sandberg, Sylvia Trunk, Eleonora Matisic, Liselotte Antonsson, My Jegenberg, for helping me deal with a lot of problems and keeping the department running safely and smoothly.

Friends in the Departments of Biochemistry and Biophysics and Organic Chemistry, Can Zhu, Dan Dou, Hao Wang, Weihua Ye, Biao Fu, Youai Qiu, and Tuo Jiang for your encouragement, fruitful discussions, and support. I hope we can collaborate even more in the future.

The agencies whose funding made this research possible: the Swedish Research Council, the Swedish Foundation for Strategic Research, Hjärnfonden, the Magnus Bergwall Foundation, the Bertil Hållsten Foundation, the National Institutes of Health Grant and the Knut and Alice Wallenberg Foundation.

The foundations and agencies who provided me scholarships, for attending conferences: the Knut and Alice Wallbergen Foundation, and C F Liljevalch Scholarship Fund. Thanks to the embassy of China in Sweden and the China Scholarship Council for their interest in my research and praise.

The National Academy of Sciences for granting me permission to reprint articles here.

Per Jemth at Uppsala University, for accepting me as a Master’s student in your group. Jakob Dogan for excellent supervision during my Master’s project.

And last but not least, my husband Dr. Bin Yang, Thank you for your love and unconditional support, both in my life and my career. Meeting you was the luckiest happening in my life. Thanks also to my grandfather, Huaguo Mu, for encouraging me to study abroad, I hope that you are well; my parents, Xiaojun Mu and Biao Zhang for covering my back and supporting me no matter what difficulties I have encountered; my son, Boyan Yang, you are the most beautiful gift in my life, thank you for coming to me, mom loves you; and other family members for their support and encouragement.
References

33. Fersht AR (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability,


118. Majumder S, et al. (2015) Probing protein quinary interactions by in-
cell nuclear magnetic resonance spectroscopy. *Biochemistry* 54(17):2727–2738.


185. Mason PE, et al. (2009) Specificity of ion-protein interactions: Complementary and competitive effects of tetrapropylammonium,