

Communication: Episturational thermodynamics of soluble proteins

Ariel Fernández

Citation: *J. Chem. Phys.* **136**, 091101 (2012); doi: 10.1063/1.3691890

View online: <http://dx.doi.org/10.1063/1.3691890>

View Table of Contents: <http://jcp.aip.org/resource/1/JCPSA6/v136/i9>

Published by the [American Institute of Physics](#).

Additional information on *J. Chem. Phys.*

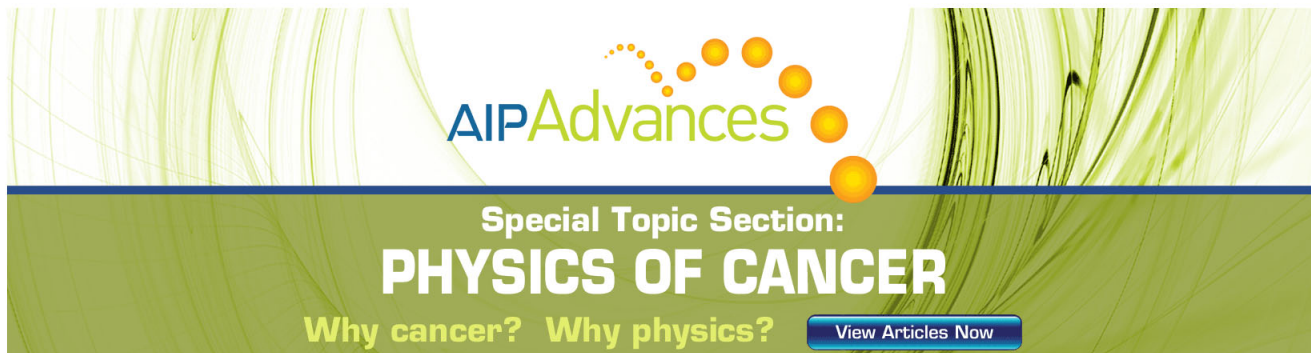
Journal Homepage: <http://jcp.aip.org/>

Journal Information: http://jcp.aip.org/about/about_the_journal

Top downloads: http://jcp.aip.org/features/most_downloaded

Information for Authors: <http://jcp.aip.org/authors>

ADVERTISEMENT



AIP Advances

Special Topic Section:
PHYSICS OF CANCER

Why cancer? Why physics? [View Articles Now](#)

Communication: Epistuctural thermodynamics of soluble proteins

Ariel Fernández^{a)}

Instituto Argentino de Matemática, CONICET (National Research Council), Saavedra 15, Buenos Aires 1083, Argentina

(Received 17 January 2012; accepted 14 February 2012; published online 1 March 2012)

The epistuctural tension of a soluble protein is defined as the reversible work per unit area required to span the interfacial solvent envelope of the protein structure. It includes an entropic penalty term to account for losses in hydrogen-bonding coordination of interfacial water and is determined by a scalar field that indicates the expected coordination of a test water molecule at any given spatial location. An exhaustive analysis of structure-reported monomeric proteins reveals that disulfide bridges required to maintain structural integrity provide the thermodynamic counterbalance to the epistuctural tension, yielding a tight linear correlation. Accordingly, deviations from the balance law correlate with the thermal denaturation free energies of proteins under reducing conditions. The picomolar-affinity toxin HsTX1 has the highest epistuctural tension, while the metastable cellular form of the human prion protein PrP^C represents the least tension-balanced protein. © 2012 American Institute of Physics. [<http://dx.doi.org/10.1063/1.3691890>]

Epistuctural attributes of soluble proteins refer to interfacial properties arising from the physical interaction between structure and solvent and have received comparatively little attention, partly because nanoscale models of interfacial water are still in development.^{1–6} The most significant epistuctural parameter is the solvent-structure interfacial tension (SSIT), that is, the reversible work per unit area required to span the solvent envelope of the protein structure. This epistuctural tension has been estimated only recently,⁶ despite its importance in determining protein associations, the basic molecular events in biological processes. A direct computation of the SSIT is difficult due to the confinement of the interfacial water within chemically and geometrically inhomogeneous nanoscale cavities^{3,6} and the complications arising from the assignment of entropic costs to such confinements. To reduce the free-energy cost of spanning the interface, the SSIT promotes a highly controlled association of soluble proteins into specific complexes, while precluding proteins from amorously precipitating to yield phase separations. This peculiar behavior sets SSIT apart from its thermodynamic homogeneous bulk-phase counterpart: phase separations are not feasible in a functional aqueous cellular context.⁴

In accord with a nanoscale description of the solvent envelope, a rigorous SSIT derivation involves an elastic term that penalizes local losses in hydrogen-bonding coordination of interfacial water. Furthermore, the treatment allows for compensations to this entropic cost as water dipoles interact with pre-existing or confinement-induced electrostatic fields,^{3,4} yielding a polarization whose magnitude depends on the local coordination restrictions of the water dipoles.

By evaluating the interfacial tension of an exhaustive set of nonhomologous monomeric soluble proteins and peptides with structure reported in the protein data bank (PDB), we

show that the SSIT is quantitatively counterbalanced by intramolecular disulfide bridges that are necessary to maintain the structural integrity of the proteins. Thus, a scaling law and related constants emerge from a tight SSIT-buttrussing linear correlation. Furthermore, deviations from the balance relation due to a lack or excess of disulfide buttressing are tightly correlated with the, respectively, lower and higher free energies of thermal denaturation under reducing conditions, providing a thermodynamic validation of the epistuctural parameter.

To compute the SSIT, we adopt the scalar field $g = g(\mathbf{r})$, a descriptor that assigns to each position vector \mathbf{r} the expected value of hydrogen-bond coordination of a water molecule situated within a sphere centered at position \mathbf{r} with radius 2.5 Å (thickness of a single water layer). A hydrogen bond is defined by the geometric constraints: O–O distance < 3.2 Å and O–H–O angle α_{HB} satisfying $120^\circ \leq \alpha_{\text{HB}} \leq 180^\circ$. The $g(\mathbf{r})$ value is computed as time average over solvent configurations determined by molecular dynamics over a 100 ns-period after the protein structure is equilibrated with the solvent, thus allowing for breathing motions of exposed atoms.⁶ Compared with bulk water ($g = 4$), interfacial water has reduced hydrogen-bonding opportunities ($g < 4$) and may counterbalance such losses by interacting with polar groups on the protein surface or with induced electrostatic fields resulting from preferred dipole alignments under confinement.^{3,4} In regards to $g(\mathbf{r})$ as computed in this work, only the first layer of the interface differs from bulk water, while water in outer layers invariably recapitulates bulk configurations. Thus, the usefulness of $g(\mathbf{r})$ as descriptor of water structure requires a local one-layer resolution. The term ΔG_{if} represents the free-energy cost of spanning the protein-water (structure-solvent) interface and incorporates unfavorable local decreases in g and favorable polarization contributions. This reversible work is given by the integral $\Delta G_{\text{if}} = \int \Xi(g(\mathbf{r}), \nabla g(\mathbf{r})) d\mathbf{r}$, where the integrand gives the free-energy cost of transferring water from bulk to volume $d\mathbf{r}$ at position \mathbf{r} . Explicitly, $\Xi(g(\mathbf{r}),$

^{a)}Electronic mail: ariel@uchicago.edu. Telephone: 608 609 4836.

$\nabla\mathbf{g}(\mathbf{r}) = (1/2)\{\lambda|\nabla\mathbf{g}|^2 - |\mathbf{P}(\mathbf{g}(\mathbf{r}))|^2\}$, where the elastic term $(1/2)\lambda|\nabla\mathbf{g}|^2$ accounts for tension-generating reductions in water coordination ($|\nabla\mathbf{g}| > 0$) and vanishes everywhere except at the first water layer of the solvent-structure interface (SSI), while the counterbalancing polarization term $\mathbf{P}(\mathbf{g}(\mathbf{r}))$ accounts for dipole-electrostatic field alignments at the interface. The scaling parameter λ is obtained from the interfacial tension of a nonpolar sphere with radius θ in the macroscopic limit $\theta/1 \text{ nm} \rightarrow \infty$. We get $\lambda = 9.0 \text{ mJ/m} = \lim_{\theta/1 \text{ nm} \rightarrow \infty} [\gamma(4\pi\theta^2)/\int (1/2)|\nabla\mathbf{g}|^2 d\mathbf{r}]$, where $\gamma = 72 \text{ mJ/m}^2$ is the macroscopic surface tension of water at 298 K.

The g -dependent polarization term $\mathbf{P} = \mathbf{P}(\mathbf{r})$ requires that we adopt the Fourier-conjugate wavenumber space (ω -space) and represent the dipole correlation kernel $K_p(\omega)$ and the electrostatic field $\mathbf{E} = \mathbf{E}(\mathbf{r})$ in this space. This representation is essential to capture the entire dielectric loss spectrum occurring mostly within the microwave range ($10^{-3} \text{ m} \leq \omega^{-1} \leq 0.3 \text{ m}$).^{7,8} In ω -space we get^{6,9,10}

$$F(\mathbf{P})(\omega) = K_p(\omega)F(\mathbf{E})(\omega) \quad (1)$$

where F denotes 3D-Fourier transform $F(\mathbf{f})(\omega) = (2\pi)^{-3/2} \int \exp(i\omega \cdot \mathbf{r})\mathbf{f}(\mathbf{r})d\mathbf{r}$, and the kernel $K_p(\omega)$ is the Lorentzian $K_p(\omega) = (\varepsilon_b - \varepsilon_o)/[1 + (\tau(\mathbf{r})c)^2|\omega|^2]$, with $\tau(\mathbf{r})c =$ local dielectric relaxation length, $c =$ light speed, $\varepsilon_b =$ bulk permittivity constant, $\varepsilon_o =$ vacuum permittivity. For bulk water, we get $\tau = \tau_b \approx 100 \text{ ps}$, and relaxation length $\tau_b c = \omega_b^{-1} \approx 0.03 \text{ m}$, the microwave wavelength yielding the best fit with experimental data.⁶ Through the Lorentzian, the frequency dependence of bulk permittivity is subsumed into the normal distribution factor $[1 + (\tau(\mathbf{r})c)^2|\omega|^2]^{-1}$ with $\tau(\mathbf{r}) \equiv \tau_b$.

For a generic charge distribution

$$\rho(\mathbf{r}) = \sum_{m \in L} 4\pi q_m \delta(\mathbf{r} - \mathbf{r}_m) \quad (2)$$

with charge spatial locations $\{\mathbf{r}_m\}$ and $L =$ set of charges on the protein surface labeled by index m , we get

$$\begin{aligned} \mathbf{P}(\mathbf{r}) &= \int F^{-1}(K_p)(\mathbf{r} - \mathbf{r}')\mathbf{E}(\mathbf{r}')d\mathbf{r}' \\ &= (2\pi)^{-3} \sum_{m \in L} \int d\mathbf{r}' F^{-1}(K_p)(\mathbf{r} - \mathbf{r}')\nabla_{\mathbf{r}'} \\ &\quad \int d\omega \exp[-i\omega \cdot (\mathbf{r}' - \mathbf{r}_m)] 4\pi q_m / [|\omega|^2 K(\omega)] \quad (3) \end{aligned}$$

with $K(\omega) = \varepsilon_o + K_p(\omega)$.

The SSI may be covered by a minimal covering set W of water-confining osculating spheres D_j s, $j \in W$. These spheres make first-order contact with the solvent envelope obtained by sliding a water molecule along the surface of the soluble protein.¹¹ As a minimal covering, W contains all interfacial water molecules and this property no longer holds if any osculating sphere is excluded from the set. Interfacial tension arises in D_j when $\Delta G_j > 0$, where $\Delta G_j = (1/2)\int_{D_j} \{\lambda|\nabla\mathbf{g}|^2 - |\mathbf{P}(\mathbf{g}(\mathbf{r}))|^2\} d\mathbf{r}$ is the interfacial surface tension associated with spanning contact region j . The SSIT is then given by $\gamma_{if} = \Delta G_{if}/\Omega$, where Ω is the solvent-exposed surface area.^{11,12}

The SSIT of soluble nonhomologous and monomeric proteins and peptides with structures reported in the PDB has been computed according to the protocol previously described⁶ for an exhaustive dataset comprised of 11 963 PDB

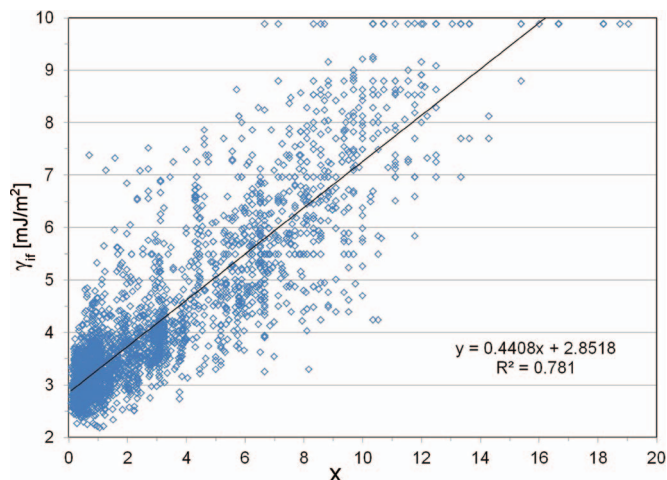


FIG. 1. Structure-solvent (epistuctural) interfacial tension *versus* normalized number of disulfide bridges (X) for structure-reported 11 963 nonhomologous monomeric soluble proteins with no prosthetic groups. The list is provided in the supplementary material.¹³ The epistuctural tension is derived from the water-coordination field $\mathbf{g}(\mathbf{r})$, in turn obtained by equilibrating the water-embedded PDB-reported structures in accord with the following tenets: The PDB structures are immersed in a pre-equilibrated truncated octahedral cell of TIP3P explicit water molecules that provide four water layers of solvent envelope.¹⁴ Counter ions are added to neutralize the systems. Protein atoms are described with the parm99SB force field parameterization, whereby the energetic criterion for hydrogen bonding (binding energy $< -1/2 \text{ kT}$) fits at 298 K the geometric criterion given in the main text.¹⁵ Water molecules extended at least 8 Å from the surface of the protein. Simulations are performed in the NPT ensemble, employing periodic boundary conditions. Ewald sums¹⁶ and an 8 Å distance cutoff are used for treating long-range electrostatic interactions. A Shake algorithm is employed to keep bonds involving hydrogen atoms at their equilibrium length,¹⁷ which allowed us to employ a 2 fs time step for the integration of Newton's equations. Constant pressure of 1 atm and temperature of 298 K are maintained using a Berendsen coupling scheme.¹⁸ The optimized systems are heated to 298 K and pre-equilibrated for 200 ps. The resulting structures are the starting point of the generation of 100 ns-MD thermalization trajectories, where the protein backbone coordinates are constrained according to the Shake scheme and only side chains are allowed to explore conformation space.

entries. The epistuctural ($\gamma_{if} = \Delta G_{if}/\Omega$) and structural parameters (chain length, $X =$ number of disulfide bonds per 100 aminoacids) for each individual PDB entry are provided in the supplementary material.¹³ Of the 11 963 proteins examined, 2988 have disulfide bridges.¹³ The protein-water interface is decomposed into a minimal covering of osculating spheres making first-order contact with the solvent-accessible envelope, and the free-energy cost of spanning each interfacial sphere, ΔG_j , is computed as described above. The field \mathbf{g} used in the numerical integration of the ΔG_{if} expression $\int \{\lambda|\nabla\mathbf{g}|^2 - |\mathbf{P}(\mathbf{g}(\mathbf{r}))|^2\} d\mathbf{r}$ is obtained from molecular dynamics, equilibrating each water-embedded PDB structure within an isothermal-isobaric NPT ensemble.^{6,14-18}

Across the exhaustive set of natively monomeric proteins and peptides,¹³ the SSIT varies in the range $2.10 \text{ mJ/m}^2 \leq \gamma_{if} \leq 9.90 \text{ mJ/m}^2$ (Fig. 1). The toxin HsTX1 (PDB.1QUZ) with the highest known (picomolar) affinity for Kv1.3 potassium channel¹⁹ has, as expected, the highest epistuctural tension at $\gamma_{if} = 9.90 \text{ mJ/m}^2$. As we plot the epistuctural tension SSIT against the number X of disulfide bridges normalized to a chain length of 100 amino acids, the scaling law $\gamma_{if} = (0.44X + 2.85) \text{ mJ/m}^2$ emerges from linear regression (Fig. 1), and its

TABLE I. Epistuctural (η) and thermodynamic (ΔG) parameters of soluble monomeric proteins under specified conditions (T, pH) for thermal denaturation.

PDB entry	η (mJ/m ²)	ΔG (kJ/mol)	T (C)	pH	Reference
1BSQ	-0.85	46.46	40.00	7.00	22
1RTB	-0.42	42.28	25.00	8.40	23
4LYZ	-0.28	37.76	26.85	7.00	24
1CX1	-0.27	22.52	24.85	7.09	25
1QG5	0.01	36.84	40.00	7.00	22
2AIT	0.22	28.05	25.00	5.00	26
3SSI	0.79	17.04	20.00	7.00	27
1HIC	2.25	21.01	25.00	7.00	28
1PMC	3.36	4.60	20.00	3.00	29

statistical significance is assessed by the Pearson coefficient $R^2 = 0.78$. This law reflects a balance principle whereby the free-energy cost of spanning the interfacial solvent envelope is quantitatively counterbalanced by intramolecular disulfide bridges, with scaling factor 0.44 mJ/m² and baseline epistuctural tension 2.85 mJ/m².

The extent to which proteins deviate from the balance equation is estimated by the parameter $\eta = \gamma_{if} - (0.44X + 2.85)$ mJ/m². Strikingly, at $\eta = 4.21$ mJ/m², the metastable cellular form of the human prion protein PrP^C (PDB.1QM0) (Ref. 20) has the lowest disulfide-bond buttressing counterbalance to its epistuctural tension (supplementary material),¹³ attesting to the metastability of the structure and to its propensity to relinquish the soluble state in favor of a fibrillogenic aggregate.⁴ Indicative of an aberrant dysfunctional state, this spontaneous aggregation illustrates an instance of the only known type of phase separation in biology.⁴

A structure-based balance law was previously unraveled by Fernández and Berry, where disulfide bridges were shown to buttress structures with spots vulnerable to the disruptive effect of backbone hydration.²¹ However, unlike the principle expounded in this work, the previous correlation did not involve thermodynamic parameters but was based purely on a quantitative balance between structural strengths and deficiencies.

To provide a thermodynamic validation of the scaling law described in Fig. 1, we collected an exhaustive set of monomeric nonhomologous proteins for which both epistuctural and thermal denaturation thermodynamic parameters are readily available (Table I). Proteins with $\eta > 0$ do not provide sufficient intramolecular compensation for the free-energy cost of spanning their interfacial solvent envelope, while those with $\eta < 0$ overcompensate for their epistuctural tension. In consonance with the balance principle, the free-energy changes associated with thermal denaturation under reducing conditions are expected and in fact do correlate tightly and linearly with the η -values ($R^2 = 0.73$, Fig. 2), so that under-compensated proteins are more favorably denatured than over-compensated ones. This tight correlation between the epistuctural parameter η and the denaturation free-energy change validates the treatment of the interfacial tension put forth in this work.

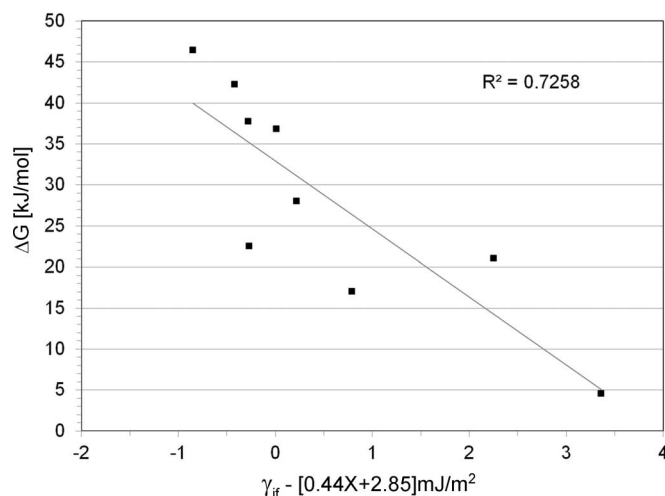


FIG. 2. Free-energy changes for thermal denaturation versus epistuctural parameter η for the monomeric soluble proteins indicated in Table I.

Biomolecular interfaces are physically heterogeneous and geometrically confined at multiple scales. For these reasons, macro-thermodynamic concepts like interfacial tension, broadly used in the context of homogeneous phase separation, are not readily applicable or even meaningful in a biological context. For soluble proteins, their epistuctural counterpart must be determined following a surface integration procedure whereby local reductions in water coordination become important contributors to the interfacial free energy. Since this free energy increment translates into reversible work to span the interface, our *ansatz* is compatible with the energetic enhancement of hydrogen bonds that correlates with depletion of nearest water neighbors.³⁰ Based on these premises, this work provides the physical framework that enables the surface integration and identifies the disulfide bridges as the structural features responsible for counterbalancing the epistuctural tension. This thermodynamic parameter is unique to a biological context, since biological solutes may reproducibly form complexes under highly controlled conditions but never precipitate under physiological conditions. The only distinct phase separation encountered in biology appears to be the amyloidogenic aggregation of prions,^{4,20} which as noted in this work, are endowed with a soluble structure that generates the least counterbalanced interfacial tension. It is likely that the exquisite biological control of phase separation will become amenable to a statistical mechanical treatment of crowding effects once epistuctural thermodynamics is incorporated into the Hamiltonian random matrix treatment of multicomponent systems.³¹

This work highlights the role of epistuctural tension as a destabilizer of protein structure, emphasizing the importance of this thermodynamic parameter in providing the physical underpinnings for the control of protein associations as a means to avoid phase separation.

This research was supported in part by CONICET, the National Research Council of Argentina.

¹M. Gerstein and C. Chothia, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10167 (1996).

- ²N. Giovambattista, C. F. Lopez, P. Rosicky, and P. G. Debenedetti, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2274 (2008).
- ³E. Schulz, M. Frechero, G. Appignanesi, and A. Fernández, *PLoS ONE* **5**, e12844 (2010).
- ⁴A. Fernández, *Transformative Concepts for Drug Design: Target Wrapping* (Springer-Verlag, Germany, 2010).
- ⁵P. W. Fenimore, H. Frauenfelder, B. H. McMahon, and R. D. Young, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 14408 (2004).
- ⁶A. Fernández and M. Lynch, *Nature (London)* **474**, 502 (2011).
- ⁷J. B. Hasted, *Water: A Comprehensive Treatise*, edited by F. Franks (Plenum, New York, 1972), Vol. 1, pp. 255–309.
- ⁸R. Buchner, J. Barthel, and J. Stauber, *Chem. Phys. Lett.* **306**, 57 (1999).
- ⁹P. Debye, *Polar Molecules* (Dover, New York, 1929).
- ¹⁰R. Scott, M. Boland, K. Rogale, and A. Fernández, *J. Phys. A* **37**, 9791 (2004).
- ¹¹A. G. Street and S. L. Mayo, *Folding Des.* **3**, 253 (1998).
- ¹²N. Zhang, C. Zeng, and N. S. Wingreen, *Proteins: Struct., Funct., Bioinf.* **57**, 565 (2004).
- ¹³See supplementary material at <http://dx.doi.org/10.1063/1.3691890> for table of exhaustive set of 11,963 nonhomologous monomeric proteins with respective PDB entries, structural and epistructural parameters.
- ¹⁴W. L. Jorgensen, J. Chandrasekhar, J. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).
- ¹⁵V. Hornak, R. Abel, A. Okur, B. Strockbine, A. E. Roitberg, and C. L. Simmerling, *Proteins: Struct., Funct., Bioinf.* **65**, 712 (2006).
- ¹⁶T. Darden, D. York, and L. Pedersen, *J. Chem. Phys.* **98**, 10089 (1993).
- ¹⁷J. P. Ryckaert, G. Ciccotti, and H. J. Berendsen, *J. Comput. Phys.* **23**, 327 (1977).
- ¹⁸H. J. Berendsen, J. P. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak, *J. Chem. Phys.* **81**, 3684 (1984).
- ¹⁹R. McKinnon, P. H. Reinhart, and M. N. White, *Neuron* **1**, 997 (1988).
- ²⁰R. Zahn, A. Liu, T. Luhrs, R. Riek, C. von Schroetter, F. Lopez Garcia, M. Billeter, L. Calzolari, G. Wider, and K. Wuthrich, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 145 (2000).
- ²¹A. Fernández and R. S. Berry, *J. Proteome Res.* **9**, 2643 (2010).
- ²²A. Divsalar, A. A. Saboury, A. A. Moosavi-Movahedi, and H. Mansoori-Torshizi, *Intl. J. Biol. Macrom.* **38**, 9 (2006).
- ²³T. F. O'Connor, P. G. Debenedetti, and J. D. Carbeck, *Biophys. Chem.* **127**, 51 (2007).
- ²⁴Y. M. Efimova, S. Haemers, S. Wiercinski, W. Norde, and A. A. van Well, *Biopolymers* **85**, 264 (2007).
- ²⁵M. Rajavel, D. Lalo, J. W. Gross, and C. Grubmeyer, *Biochemistry* **37**, 4181 (1998).
- ²⁶M. Renner, H. J. Hinz, M. Scharf, and J. W. Engels, *J. Mol. Biol.* **223**, 769 (1992).
- ²⁷A. Tamura and J. M. Sturtevant, *J. Mol. Biol.* **249**, 625 (1995).
- ²⁸A. Otto and R. Seckler, *Eur. J. Biochem.* **202**, 67 (1991).
- ²⁹G. Mer, H. Hietter, and J. F. Lefevre, *Nature Str. Biol.* **3**, 45 (1996).
- ³⁰S. Matysiak, P. Debenedetti, and P. J. Rosicky, *J. Phys. Chem. B* **115**, 14859 (2011).
- ³¹R. P. Sear and J. A. Cuesta, *Phys. Rev. Lett.* **91**, 245701 (2003).