

Epiststructural tension promotes protein associations

Ariel Fernández

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

Department of Computer Science, The University of Chicago, Chicago, IL 60637;

Phone: 608 609 4836; E-mail: ariel@uchicago.edu

Epiststructural tension is the reversible work per unit area required to span the aqueous interface of a soluble protein structure. The parameter accounts for the free-energy cost of imperfect hydration, involving water molecules with a shortage of hydrogen bonding partnerships relative to bulk levels. The binding hot spots along protein-protein interfaces are identified with residues that contribute significantly to the epiststructural tension in the free subunits. Upon association, such residues either displace or become deprived of low-coordination vicinal water molecules.

PACS numbers: 87.15.Aa, 61.20.Qg, 41.20.Cv

We introduce the term “epistructural” to refer to phenomenology at the interface between the solvent and the structure of a soluble protein. The term implies a nanoscale physical model of hydration that incorporates the structure of the protein as input [1-4]. Thus, epistructural interfacial tension (EIT) is the reversible work per unit area required to span the solvent envelope of the protein in a specific soluble conformation. This parameter represents the free-energy cost associated with locally imperfect hydration of the protein structure, involving water molecules with a shortage of hydrogen-bonding partnerships relative to bulk levels. The EIT has been recently estimated using molecular dynamics equilibration of water-embedded structures and surface integration procedures [5]. Its role as promoter of protein associations, the basic molecular events in the cell, is beginning to be recognized [3-5]. Complementing the treatment in [5], in this letter we present a variational derivation of the EIT that requires an assessment of the entropic cost associated with interfacial water confinement in inhomogeneous nanoscale cavities. This entropy contribution is a thermodynamic indicator of imperfect hydration. To reduce the free-energy cost of spanning the interface, the EIT promotes highly controlled associations of soluble proteins into complexes, while precluding proteins from precipitating into a separate phase. As shown in this work, these selective associations are determined by a variational principle that governs the minimization of EIT.

The EIT derivation involves an elastic term that accounts for local reductions in hydrogen-bond coordination of interfacial water, while allowing for compensations as water dipoles interact with pre-existing or dipole-confinement-induced electrostatic fields. The latter interactions yield a polarization term whose magnitude depends on local coordination restrictions of water molecules.

In this work we establish the crucial role of EIT as promoter of specific protein associations by adopting a variational approach to derive the interfacial tension. We contrast significant local contributions to the epistructural tension of free (uncomplexed) protein structures against an alanine-scanning dissection of binding hot spots along structurally reported protein-protein (PP) interfaces involving the same protein chains [6-14]. The alanine scanning procedure involves the

calorimetric determination of the change in association free energy $\Delta\Delta G_a$ resulting from a residue-specific alanine substitution, with substitution at hot spots along the PP interface bringing about the most significant increases in ΔG_a . In this work, we show that binding hot spots are residues that either displace or become deprived of low-coordination backbone-solvating water molecules upon protein association. Thus, hot spots are shown to be the dominant contributors to the reduction in epistructural tension that occurs upon protein association.

This solvent-centric approach complements previous approaches that predicted hot spots based on local compositional and geometric properties of the protein surfaces [15, 16]. These earlier approaches take into account shape complementarity, pair-wise polar interactions across the protein-protein interface and dehydration penalties associated with the burial of polar groups. These factors are in fact determinants of the structural quality of the interfacial water lattice around the free subunits. Thus, by focusing directly on the interfacial water structure, a unified physical picture is obtained, subsuming the attributes exploited in previous approaches and incorporating hitherto neglected features, especially the free-energy cost of imperfect hydration and the confinement-related dynamic propensity of interfacial water molecules.

To compute the EIT, we adopt the coordination scalar field $g=g(\mathbf{r})$, a water-structure 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\AA (thickness of a single water layer). Compared with bulk water ($g=4$), interfacial water has reduced hydrogen-bonding opportunities ($g<4$) that may be counterbalanced through interactions with polar groups on the protein surface or with induced electrostatic fields resulting from dipole alignments that in turn result from nanoscale confinement. Thus, the term ΔG_{if} represents the free-energy cost (reversible work) associated with spanning the protein-water interface and incorporates unfavorable local decreases in g as entropic cost associated with transferring water molecules from bulk solvent to confined environments. These local decreases in g describe sites of imperfect or insufficient hydration of the protein structure.

Through variational calculus, we construct the water-coordination field $g(\mathbf{r})=g^*(\mathbf{r})$ that minimizes ΔG_{if} , and identify the “hot” interfacial ($g^*<4$)-water molecules, showing that they become displaced upon protein association. In this way, we enable a computational identification of binding hot spots, validated by comparison with alanine-scanning results.

The reversible work required to span the epistructural envelope of the protein is an extremal of the g -dependent functional 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} and $g=g^*(\mathbf{r})$ is the coordination field that minimizes ΔG_{if} . Explicitly, $\Xi(g(\mathbf{r}), \nabla g(\mathbf{r})) = \frac{1}{2} \{ \lambda |\nabla g|^2 - |\mathbf{P}(g(\mathbf{r}))|^2 \}$, where the elastic term $\frac{1}{2} \lambda |\nabla g|^2$ accounts for tension-generating reductions in water coordination ($|\nabla g| > 0$) and vanishes everywhere except at the epistructural interface, while the polarization term $\mathbf{P}(g(\mathbf{r}))$ accounts for dipole-electrostatic field alignments at the interface [5]. By singling out imperfect water configurations that depart from bulk coordination, the elastic term accounts for the thermodynamic cost of imperfect protein hydration that leads to distortion of local ($g=4$)- states. The elasticity is compatible with the variational nature of the problem, whereby the extremal scalar field seeks to minimize coordination distortion, and is mechanically justified since there is no tension without elasticity. The compatibility with water macroscopic surface tension in the large curvature radius limit has been established [5]. Thus, 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 $\lambda = 9.0\text{mJ/m} = \lim_{\theta/1\text{nm} \rightarrow \infty} [\gamma(4\pi\theta^2)/\int \frac{1}{2} |\nabla g|^2 d\mathbf{r}]$, where $\gamma=72\text{mJ/m}^2$ is the macroscopic surface tension of water at 298K.

To determine the g -dependent polarization term $\mathbf{P}=\mathbf{P}(\mathbf{r})$, we adopt the Fourier-conjugate wavenumber space (\mathbf{v} -space) and represent the dipole correlation kernel $K_p(\mathbf{v})$ and the electrostatic field $\mathbf{E}=\mathbf{E}(\mathbf{r})$ in this space [3,5,17]. This representation is required to capture the entire dielectric loss spectrum occurring mostly within the microwave range ($10^{-3}\text{m} \leq \mathbf{v}^{-1} \leq 0.3\text{m}$).

In \mathbf{v} -space we get

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

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

The polarization $\mathbf{P}(\mathbf{r})$ satisfies $\nabla(\epsilon_o \mathbf{E} + \mathbf{P})(\mathbf{r}) = \rho(\mathbf{r})$ ($\rho(\mathbf{r})$ =charge density distribution) [18] and thus Eq. (1) yields the following equation in \mathbf{r} -space:

$$\nabla[\int F^{-1}(\mathbf{K})(\mathbf{r}-\mathbf{r}') \mathbf{E}(\mathbf{r}') d\mathbf{r}'] = \rho(\mathbf{r}), \quad (2)$$

with $\mathbf{K}(\mathbf{v}) = \epsilon_o + K_p(\mathbf{v})$. The convolution $\int F^{-1}(\mathbf{K})(\mathbf{r}-\mathbf{r}') \mathbf{E}(\mathbf{r}') d\mathbf{r}'$ gives the dipole correlation induced by the electrostatic field.

Upon water confinement, the dielectric relaxation undergoes a frequency redshift arising from the reduction in hydrogen-bond partnerships that translates in a reduction in dipole orientation possibilities ($g(\mathbf{r}) < 4$, $\tau(\mathbf{r}) > \tau_b$) [2]. Thus, at position \mathbf{r} , the relaxation time is $\tau = \tau_b \exp(B(g(\mathbf{r}))/k_B T)$, where the kinetic barrier $B(g(\mathbf{r})) = -k_B T \ln(g(\mathbf{r})/4)$ yields $\tau(\mathbf{r}) = \tau_b (g(\mathbf{r})/4)^{-1}$.

For a generic charge distribution:

$$\rho(\mathbf{r}) = \sum_{m \in L} 4\pi q_m \delta(\mathbf{r} - \mathbf{r}_m), \quad (3)$$

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

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

The epistructural interface 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. 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 $\Delta G_j = \frac{1}{2} \int_{D_j} \{ \lambda |\nabla g|^2 - |\mathbf{P}(g(\mathbf{r}))|^2 \} d\mathbf{r}$ is the interfacial surface tension associated with spanning contact region j . The EIT is then given by $\gamma_{if} = \Delta G_{if} / \Omega$, where Ω is the solvent-exposed surface area [19,20].

From variational calculus, the field $(g^*, \nabla g^*)$ is the extremal of $\Delta G_{if} = \int \Xi(g(\mathbf{r}), \nabla g(\mathbf{r})) d\mathbf{r}$ if it satisfies the spatial version of the Euler-Lagrange equation [21]:

$$\nabla [\partial \Xi / \partial \nabla g] = \partial \Xi / \partial g, \quad (5)$$

with boundary condition $g(\mathbf{r}) = 4$ for $\|\mathbf{r}\| = R$. This condition is defined with R sufficiently large so the protein molecule is contained in the ball $\{\|\mathbf{r}\| < R', R' \ll R - 5d\}$, with $d = 2.5 \text{ \AA} \approx$ thickness of a water layer. Thus, $g^*(\mathbf{r})$ satisfies the Poisson-type equation:

$$\nabla^2 g = - (1/\lambda) \partial \mathbf{P} / \partial g, \text{ with } g(\mathbf{r}) = 4 \text{ for } \|\mathbf{r}\| = R, \quad (6)$$

Thus, the water-coordination field $g^*(\mathbf{r})$ that minimizes the EIT is obtained from numerical integration of Eq. (6).

The $g^* < 4$ water molecules have been shown to solvate backbone hydrogen bonds that are partially exposed to solvent, known as *dehydrons* [5]. Thus, water is able to form hydrogen bonds with the paired polar groups of the backbone, the amide ($>NH$) and carbonyl ($>C=O$), only at the expense of reducing its g -value below bulk ($g=4$) levels. This reduction in coordination is the result of partial confinement as the water molecule fills the cavity that results from the incomplete burial of the dehydron [4]. The position of interfacial water molecules with a significant ($g^* < 3.5$) reduction of coordination in the two free (uncomplexed) subunits of human insulin is shown in Fig. 1. The “hot” water molecules solvate dehydrons (green lines) that become well-shielded hydrogen bonds (grey lines) within the insulin complex. Thus, the hot water molecules are displaced upon association, resulting in the intermolecular shielding of pre-formed dehydrons, and this displacement reduces the EIT due to the reduction in the term $\frac{1}{2} \int \lambda |\nabla g|^2 d\mathbf{r}$.

We now establish the fact that binding hot spots are sites responsible for the most significant reduction in EIT upon association. To this effect, we analyze patterns of interfacial water exclusion resulting upon complex formation. We focus on complexes for which the residue contribution to the association free energy ΔG_a has been dissected through alanine scanning of the PP interface. The residue substitution for alanine amounts to a truncation of the side chain at the β -carbon. By calorimetrically determining $\Delta\Delta G_a$ for the wild-type \rightarrow mutant transformation, we can identify the residues that most contribute to the association free energy.

The contention of this paper is that binding hot spots are residues that most significantly contribute to reduce the epistructural tension upon association by either displacing or becoming deprived of hot ($g^<4$) vicinal water upon association.*

Fig. 2 validates this assertion, focusing on alanine scanning analysis of PP interfaces for well-studied complexes, and contrasting the experimental results with the coordination quality (g^* -value) of water excluded upon complexation. To identify the location of hot ($g^*<4$) water molecules relative to hot-spot residues, we define *hydration vicinity* of a residue by two 4Å-radius spheres centered at the backbone amide nitrogen and carbonyl oxygen of the residue. Then a water molecule is said to be vicinal when the oxygen atom, roughly at the barycenter of the molecule, lies within the hydration vicinity of the residue. Note that a water molecule may be vicinal to more than one residue.

To contrast epistructural thermodynamics against alanine-scanning results for different complexes, we group residues according to the $\Delta\Delta G_a$ for wild-type \rightarrow mutant transformation and also according to the coordination quality of the interfacial water they displace or are deprived of upon protein-protein association. Hot-spot residues are classified according to the ranges $\Delta\Delta G_a \geq 3\text{kcal/mol}$, $1\text{kcal/mol} \leq \Delta\Delta G_a < 3\text{kcal/mol}$ and $\Delta\Delta G_a < 1\text{kcal/mol}$ (upper rows for each PP interface, Fig. 2). According to a second classifier (lower rows, Fig. 2), residues are grouped according to the ranges for vicinal water: $g^*<3$, $3 \leq g^*<4$, $g=4$. The PP interfaces for complexes

with available alanine scanning data were examined, classifying residues according to the $\Delta\Delta G_a$ and independently according to the g^* -parameter obtained by integration of Eq. 6 on free subunits. Thus, the following complexes were examined (PDB entries are given in brackets): human growth hormone/hGH receptor (3HHR) [6], HIV-1-CD4/GP120 (1GC1) [11], barnase and barstar in barnase/barstar complex (1BRS) [14], P53/MDM2 (1YCR) [10], trypsin inhibitor/beta-trypsin (2PTC) [9], ribonuclease inhibitor/angiogenin (1A4Y) [12], ribonuclease inhibitor/ribonuclease A (1DFJ) [12], colicin E9 immuno-protein/colicin E9 DNase domain (1BXI) [13].

A statistically significant correlation exists between the $\Delta\Delta G_a$ classifier and the g^* -classifier of interfacial residues (Fig. 2). *This correlation enables us to assert ($P\text{-value} < 10^{-5}$) that protein association is driven by displacement of hot interfacial water that promotes an EIT reduction.*

The prediction of binding hot spots is a long-standing biophysical problem. Here we provide a solution exploiting calculus of variations to determine the epistructural tension as extremal of a free-energy functional and computing the reduction in tension resulting from protein associations. The computational results are validated by contrasting them against the dominant contributions to the association free energy obtained from alanine scanning of protein-protein interfaces.

Protein-water interfaces are physically heterogeneous at multiple scales of water confinement. These singular features make thermodynamic concepts like interfacial tension, used to characterize homogeneous phase separations, not applicable in a biological context. For soluble proteins, the epistructural interfacial tension defined in this work must be determined following a surface integration procedure. The variational principle described enables such integration. The approach also enables the identification of hot interfacial water molecules, i. e. those with reduced hydrogen-bonding capability, singled out as we solve the variational problem of minimizing the epistructural tension. Thus, the identification of hot interfacial water molecules that are removed upon protein-protein association enables the prediction of binding hot spots and

may serve as a guidance to design drugs capable of disrupting protein associations to achieve therapeutic effects [3].

Epistuctural tension is mainly caused by local solvent exposure of the protein backbone, and the latter is indicative of structural disruption [5], therefore our approach incorporates conformational plasticity in its dissection of binding interfaces, fulfilling the imperatives of reference [16].

The variational principle outlined in this work may be exploited in any context of nanoscale water confinement, including water penetration into carbon nanotubes [22, 23]. For diameter $\sim 1\text{nm}$ we expect significant polarization effects resulting from induced electrostatic fields generated by wire-like aligned water confinement and from a low-dielectric enhancement of hydrogen bonds. These effects offset the elastic contribution ($|\nabla g| > 0$) that arises at the open boundaries of the nanotube, as water becomes deprived of hydrogen-bonding partnerships. Thus, the variational approach to interfacial tension can also address non-biological contexts, the subject of a future contribution.

Figure Captions

Figure 1. Interfacial water exclusion upon complex formation in human insulin (PDB.3E7Y). The ribbon rendering of the chain conformations (upper left) are an aid to the eye. The free (uncomplexed) chains (A in magenta, B in blue) are shown on the lower left. The backbone is represented as virtual bonds joining the α -carbons of residues along the chain, with well-shielded backbone hydrogen bonds and dehydrons shown as segments sustained between the paired residues in grey and green, respectively. The location of hot interfacial water molecules is marked by asterisks and relative to the hydration vicinity of dehydrons in the free chains. The hot interfacial water molecules are excluded upon complex formation, as shown on the right, and the dehydrons vicinal to hot water molecules in the free chains turn into well-shielded hydrogen

bonds (green to grey) within the complex. The intermolecular water exclusion patterns are marked by thin blue lines joining the residue that excludes water upon binding and the center of the hydrogen bond whose vicinity is deprived of interfacial water upon association. All hot interfacial water molecules with $g^* < 3.5$ are excluded upon complexation.

Figure 2. Comparison between alanine scanning and variational analysis of protein-protein interfaces combining the $\Delta\Delta G_a$ classifier (upper entries) and the g^* -classifier (lower entries) of interfacial residues. Hot-spot residues marked in dark grey (bold font), white and light grey correspond to the ranges $\Delta\Delta G_a \geq 3 \text{ kcal/mol}$, $1 \text{ kcal/mol} \leq \Delta\Delta G_a < 3 \text{ kcal/mol}$ and $\Delta\Delta G_a < 1 \text{ kcal/mol}$, respectively (upper rows for each PP interface). According to the second classifier (lower rows), residues are marked in dark grey (bold font), white and light grey if their vicinal water lies in the ranges $g^* < 3$, $3 \leq g^* < 4$, $g^* = 4$, respectively. The interfacial effect of the association upon the hydration vicinity of a residue is represented by “X” if water is displaced, “#” if the residue displaces water intermolecularly as a result of its side-chain penetration into the vicinity of another residue across the PP interface, “S” if an intermolecular salt bridge is formed across the PP-interface, and “blank space” if the hydration vicinity is unaltered.

References

1. N. Giovambattista, C. F. Lopez, P. Rossky and P. G. Debenedetti, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2274 (2008).
2. F. Despa, *Ann. N. Y. Acad. Sci.* **1066**, 1 (2006).
3. A. Fernández, *Transformative concepts for drug design: Target wrapping* (Springer-Verlag, Heidelberg, Germany, 2010).
4. P. W. Fenimore, H. Frauenfelder, B. H. McMahon and R. D. Young, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 14408 (2004).
5. A. Fernández and M. Lynch, *Nature* **474**, 502 (2011).

6. T. Clackson and J. A. Wells, *Science* **267**, 383 (1995).
7. K. S. Thorn and A. A. Bogan, *Bioinformatics* **17**, 284 (2001).
8. Z. Li and J. Li, *Bioinformatics* **78**, 3304 (2010).
9. M. J. Castro and S. Anderson, *Biochemistry* **35**, 11435 (1996).
10. V. Bottger, A. Bottger, C. Garcia-Echeverria, Y. Ramos, A. van der Eb, A. G. Jochemsen and D. P. Lane, *J. Mol. Biol.* **269**, 744 (1997).
11. A. Ashkenazi, L. Presta, S. Marsters, J. Camarato, K. Rosenthal, B. Fendly and D. Capon, *Proc. Nat. Acad. Sci. U. S. A.* **87**, 7150 (1990).
12. C. Chen and R. Shapiro, *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1761 (1997).
13. U. C. Kuhlmann, A. Pommer, G. R. Moore, R. James, C. Kleanthous, *J. Mol. Biol.* **301**, 1163 (2000).
14. G. Schreiber and A. R. Fersht, *J. Mol. Biol.* **248**, 478 (1995).
15. T. Kortemme and D. Baker, *Proc. Natl. Acad. Sci. USA* **99**, 14116 (2002).
16. W. DeLano, *Curr. Op. Struct. Biol.* **12**, 14 (2002).
17. R. Scott, M. Boland, K. Rogale and A. Fernández, *J. Phys. A* **37**, 9791 (2004).
18. P. Debye, *Polar Molecules* (Dover Publications, New York, USA, 1929).
19. A. G. Street and S. L. Mayo, *Folding & Design* **3**, 253 (1998).
20. N. Zhang, C. Zeng and N. S. Wingreen, *Proteins Struct. Funct. Bioinf.* **57**, 565 (2004).
21. J. S. Rowlinson and B. Widom, *Molecular Theory of Capillarity* (Oxford University Press, New York, USA, 1982).
22. T. Pascal, W. A. Goddard and Y. Jung, *Proc. Natl. Acad. Sci. USA* **108**, 11794 (2011).
23. A. Fernández, *J. Chem. Phys.* **119**, 5315 (2003).