Non-Debye frustrated hydration steers biomolecular association:

Interfacial tension for the drug designer

Ariel Fernández *

Argentine Mathematics Institute (I. A. M.)
National Research Council (CONICET), Buenos Aires 1083, Argentina

AF Innovation, Pharmaceutical Consultancy GmbH, Avenida del Libertador 1092,
Buenos Aires 1112, Argentina

Corresponding autor. Email: ariel@afinnovation.com

Cell functioning involves recruitment of biomolecular complexes, a process mediated by water that gets displaced as subunits bind. This process affects water frustration, that is, the number of unmet hydrogen-bonding opportunities at the protein-water interface. By searching for least-frustrated aqueous interfaces, this work delineates the role of frustration in steering molecular assemblage. The search entails a trajectory sampling using a functional measuring the gradient of frustration and computing the resulting non-Debye electrostatics within relaxation times for coupled protein-water systems. The minimal frustration principle is validated against spectroscopic measurements of frustration-dependent dielectric relaxation, affinity scanning of protein-protein interfaces, and NMR-inferred association propensities of protein-complex intermediates. The methods are applied to drug/ligand design, revealing the targetable nature of the aqueous interface while enabling the engineering of ultra-specific drugs.

Key Words: aqueous interface, molecular modeling, protein associations, structural biology, drug design

Introduction

With the recognition that no cogent theory fully reconciles thermodynamic and structural characterization of protein associations, and that hydrophobicity is hardly operative at Angstrom-level curvatures, the forces that steer the formation of protein complexes remain under intense scrutiny [1-6]. This work addresses this problem, investigating protein complex formation from a solvent-centric perspective that takes into account the unusual properties of water near biomolecular surfaces. The approach requires a variational procedure to optimize the frustration state [7] of the aqueous interface. Frustration refers to unmet opportunities for hydrogen-bonding that cause interfacial tension [6]. It is at present unclear whether force fields and solvation models adopted in molecular dynamics (MD) computations can properly model frustration or capture interfacial tension [8-12]. It is also unclear -and unlikely- that the global free energy minimum would be a state of minimal interfacial frustration and, while the biological relevance of the former is still debated, the latter is clearly decisive in driving complex formation, as this work shows [5,11].

As shown in this work, a special microstate sampling strategy steering MD computations towards the least-frustrated interface proves essential to identify the molecular factors that trigger complexation. Accordingly, we implement and experimentally validate an explicit-solvent MD algorithm with microstate sampling guided by a functional that defines a variational principle to enable the search for minimally frustrated interfaces. The functional form of the variational principle hinges on the *ansatz* that frustration creates a polarization component decoupled from the electrostatic field [5,6]. The functional adopted exploits the link between frustration and electrostatics and computes the non-Debye electrostatic energy equivalently stored as the frustration-related interfacial tension [5]. Experimental evidence supports this assumption, as

shown in the Supplementary Material (SM). Thus, the trajectories generated entail the computation of the gradient of frustration at each spatial location.

We demonstrate that frustration-related water polarization steers protein-protein (PP) associations with high specificity and drives the dynamics of complex formation. The validation of these assertions entails contrasting optimally frustrated states with affinity scanning of PP-interfaces [13] in complexes that assemble to mitigate frustrated hydration of free subunits, and with NMR information on the dynamics of complex formation [14]. Subsequently, we identify the optimal sequence of steps that maximize the mitigation of interfacial tension, funneling complex formation. Finally, we apply the solvent-centric approach to drug design in order to demonstrate the efficacy of the variational approach to engineer ultra-specificity, creating ligands that discriminate isoforms of target proteins.

Results

Water configurations with minimal frustration

To describe locally frustrated hydration, we introduce a "frustration scalar field" $\phi=\phi(\mathbf{r})$ that indicates the expected number of missed hydrogen bond opportunities of a test water molecule at spatial location $\mathbf{r}$ relative to the quasi-tetrahedral coordination associated with bulk water. The frustration field is defined as $\phi(\mathbf{r})=4-g(\mathbf{r})$, where $g(\mathbf{r}) \leq 4$ is the time-averaged number of hydrogen bonds engaging a water molecule visiting a sphere of radius r=4Å centered at position $\mathbf{r}$ for a minimum time period τ=1ps, a permanence time typically associated with the relaxation timescale for a decoupled water lattice [15]. The hydrogen bond is operationally defined via

geometric constraints [5,12]: O-O distance <3.2Å and O-H-O angle α_{HB} satisfying $120° \leq \alpha_{HB} \leq 180°$.

Water frustration is caused by partial nanoscale confinement at the protein surface and hence introduces a component $\overrightarrow{P^{\sharp}}$ of the polarization vector $\vec{P}$ that is uncoupled from the electrostatic field $\vec{E}$ and orthogonal to the Debye polarization component $\overrightarrow{P^{\parallel}}$ [5,6]. Thus, the interfacial energy is stored in the electrostatic energy term

$$\Delta U^{\sharp} = \frac{1}{2}\varepsilon_0^{-1} \int \left\| \overrightarrow{P^{\sharp}} \right\|^2 d\vec{r} = \frac{1}{2} \int \omega \left\| \vec{\nabla}\phi \right\|^2 \, d\vec{r} \tag{1}$$

where the elastic integral on the r.h.s. results from the *ansatz* [5]:

$$\overrightarrow{P^{\sharp}} = -\xi \nabla \phi \, , \quad \xi = (\omega \varepsilon_0)^{1/2} \tag{2}$$

pointing to water frustration, with its lack of partial charge cancellation, as causative of the anomalous non-Debye polarization. The conceptual link between frustration and non-Debye electrostatics defined by Equation (2) is experimentally validated as shown in SM, where the prior determination of the constant ω is also provided. Equation (2) implies that the non-Debye electrostatic energy may be equivalently stored as interfacial tension, which is frustration-related [5]. Thus, the link between frustration and electrostatics combined with a standard entropic model used in the derivation of a Poisson-Boltzmann equation [16] motivates the choice of a Lagrangian $\mathcal{L}(\phi)$ defining a free-energy functional $\Im(\phi) = \int \mathcal{L}(\phi) \, d\vec{r}$. This functional is thus introduced to compute the free energy cost $\Delta G_{if}(\phi)$ of generating the water frustration field ϕ when spanning an aqueous interface enveloping the solute:

$$\Delta G_{if}(\phi) = \Im(\phi) = \int \left[\frac{1}{2}\omega \left\| \vec{\nabla}\phi \right\|^2 - k_B T(\rho_0 - \rho) \ln\left(1 - \frac{\phi}{4}\right) \right] d\vec{r} \tag{3}$$

The elastic energy term $\int \frac{1}{2}\omega\left\|\vec{\nabla}\phi\right\|^2 d\vec{r}$ from Equation (1) is a Dirichlet-type functional [12,17] penalizing departures from bulk structure, and the interfacial entropy differential $dS_{if}(\phi) = k_B ln[\frac{4-\phi}{4}]^{(\rho_0-\rho)d\vec{r}} = [k_B(\rho_0-\rho)\ln\left(1-\frac{\phi}{4}\right)]d\vec{r}$ corresponds to the transference of a water molecule from bulk solvent (number density ρ_o = 0.033Å^{-3}) to a differential volume region $d\vec{r}$ located at position **r** with number density ρ(**r**) and frustration $\phi=\phi$(**r**). For small structural distortions at $0<\phi<<4$, a linear approximation yields $dS_{if}(\phi) \approx -[k_B(\rho_0-\rho)\,(\frac{\phi}{4})]d\vec{r}$.

The least-frustrated state of the aqueous interface is represented by an extremal frustration field ϕ determined as described in the Methods section. The minimization of interfacial frustration is shown to translate into a major factor steering protein associations (Figure 1). Frustration cannot be incorporated explicitly into the canonical-ensemble molecular dynamics of soluble proteins since there is no energetic penalty for missed interactions. Yet, as shown in Methods section, we can select at each stage of the simulation a representative from the set of *a-priori* equi-probable microstates to ensure a *likely* decrease in $\Delta G_{if}(\phi) = \Im(\phi)$ that would ultimately lead to an effective search for a state of minimal frustration.

Maximal reduction of water frustration steers highly specific protein-protein associations

To assess the steering power of the functional given in Equation (3) towards the minimally frustrated interfacial state, and illustrate the role of frustrated hydration as promoter of protein associations, we generated 500ps-trajectories for soluble proteins known to engage in complexes with structures reported in the protein data bank (PDB). A TIP3P explicit solvent [18] within an AMBER package [19,20] and Coulombic interactions evaluated with the Ewald summation

scheme [21] were adopted for MD simulation steered by the microstate sampling strategy described in Methods. An isothermal/isobaric ensemble at T=298K was adopted with initial structural coordinates for the protein chain obtained for the corresponding free subunit extracted *in silico* from the PDB-reported complex. The free (uncomplexed) protein subunits were solvated in a water box extending at least 10Å beyond any protein atom and were thermalized for 5ns. Following relaxation, positional restraints defined by force constant=5kcal/Å^2 were imposed on all protein nonhydrogen atoms.

In Figure 2A we report the time evolution of $\Delta G_{if}(\phi) = \Im(\phi)$ for the free human growth hormone (hGH) receptor subunit hGHbp that forms the receptor/hGH complex, adopting the structural coordinates for the individual subunits provided in PDB entry 3HHR [13] with the structures thermalized as previously indicated. The thin solid line reports a single trajectory, the thick solid grey line, the average over 20 trajectories following the iterative scheme given in Methods, and the dashed line represents the control consisting of an average over 20 trajectories of 500ps-MD simulation without enforcing the microstate sampling strategy. The minimal $\Delta G_{if}(\phi)$-value is $\Delta G_{if}{}^{*}(\phi) = 121 kJ/mol$, corresponding to interfacial tension γ=2.13mJ/m^2. A significant relaxation is observed in the 350-400ps range (Figure 2A), in agreement with experimental relaxation times for the coupling of protein motion and hydration pattern [21]. It should be noted that such values are incommensurably larger than the relaxation time for the decoupled water lattice (~1ps) previously incorporated [15].

To assert the role of interfacial water frustration as promoter of PP association (cf. Figure 1), we determine the effect of site-directed mutations on the interfacial free energy increment $\Delta G_{if}(\phi)$, and compare the $\Delta\Delta G_{if}(\phi)$ values with the change in affinity quantified by the

calorimetrically measured $\Delta\Delta G_a = RTln[\frac{K_d(m)}{K_d(0)}]$, where $\Delta\Delta G_a$ is the change in association free energy that results from a site-specific mutation [13], with K_d(m), K_d(0) indicating the equilibrium dissociation constants for mutant (m) and wild-type (0), respectively. Hot spots in PP associations were experimentally identified by alanine scanning of the PP interface [13]. This technique involves site-specific amino acid substitution for alanine (effectively, a side-chain truncation). The most comprehensively scanned interface corresponds to the 1:1 hGH-hGHbp complex [13]. Thus, the changes in affinity measured by $\Delta\Delta G_a$ and obtained from alanine scanning were contrasted with *in-silico* substitution effects computed as changes in interfacial free energy $\Delta\Delta G_{if} = \Delta G_{if}(m) - \Delta G_{if}(0)$. A tight correlation ($R^2$=0.88) between $-\Delta\Delta G_{if}$ and $\Delta\Delta G_a$ corresponding to the scanning of the receptor hGHbp (Figure 2B), supports the relation $-\Delta\Delta G_{if} \approx \Delta\Delta G_a$, *and upholds the view that the complexes assemble to reduce the interfacial tension of the free subunits.*

To further support this assertion, other protein associations were considered where $\Delta\Delta G_a$-values were available [13, 23-26]. For example, the p53 trans-activation domain, that forms a complex with MDM2 (PDB.1YCR) [26], has been thermalized and equilibrated in free form following the protocol previously indicated to identify hot spots representing contributions ≥3kcal/mol to ΔG_{if}. Such hot spots correspond to the frustrated water partially confined in the cavities formed by packing defects in the protein structure (Figure 1) [5,6]. These defects consist of solvent-exposed backbone hydrogen bonds, known as *dehydrons* (Figure 2C). The backbone hydrogen bonds in the MDM2 fragment are induced upon binding and do not constitute hot spots, since no stable structure can be sustained by the MDM2 peptide in free form. Upon forming the complex, residues F19, W23 and L26 in MDM2 have been identified as removing frustrated water from the vicinity of the hot-spot p53 dehydrons shown in Figure 2C. For example, L26 displaces

frustrated water from hot-spot dehydrons M50-L54, L54-G58, L94-H96, H96-Y100 and Y100-Y104, as shown in Fig. 2D. These considerations lead us to identify MDM2 residues F19, W23 and L26 as the significant contributors, yielding $-\Delta\Delta G_{if} \geq 3kcal/mol$ upon *in silico* scanning of the PP interface. This result is in accord with the experimental alanine scanning yielding $\Delta\Delta G_a \geq 3kcal/mol$ solely for these three residues in MDM2 [26].

To extend and simplify the analysis, we classified residues according to $\Delta\Delta G_a$-ranges and independently according to $\Delta\Delta G_{if}$-ranges. Adopting the affinity classifier, hot-spot residues are labeled (upper row for each PP interface, Figure 2E) according to the ranges $\Delta\Delta G_a \geq 3kcal/mol$ (red), $1kcal/mol \leq \Delta\Delta G_a < 3kcal/mol$ (white), and $\Delta\Delta G_a < 1kcal/mol$ (blue). Adopting the interfacial tension classifier (lower row for each PP interface), residues are labeled according to the ranges $-\Delta\Delta G_{if} \geq 3kcal/mol$ (red), $\frac{1kcal}{mol} \leq -\Delta\Delta G_{if} < 3kcal/mol$ (white), and $-\Delta\Delta G_{if} < 1kcal/mol$ (blue). Using these two classifiers, the PP interfaces for five complexes with scanning data available were examined: hGH/hGHbp receptor (3HHR) [13], HIV-1-CD4/GP120 (1GC1) [23], trypsin inhibitor/beta-trypsin (2PTC) [24], colicin E9 immuno-protein/colicin E9 DNase domain (1BXI)[25] and p53/MDM2 (1YCR)[26]. *The significant correlation between the $\Delta\Delta G_a$ and $-\Delta\Delta G_{if}$ classifiers (Figure 2E, P-value<10^{-5}) upholds the view that interfacial tension generated by frustrated hydration drives protein association.*

Interfacial frustration and association propensity in complex-denaturation intermediates

To further validate the variational frustration principle we analyze the complex dissociation intermediates arising from progressive cold denaturation of homodimeric repressor protein CgIR2, a system for which the dynamic intermediates have been resolved at atomic resolution

[14]. Thus, the frustration-related interfacial tension is given in Figure 3A for PDB entries (absolute temperatures in brackets) 2LYJ (298K), 2LYK (270K), 2LYL (266K, dimer), 2LYP (266K, monomer), 2LYQ (262K), 2LYR (259K) and 2LYS (257K). The open disks correspond to the computation of interfacial tension= $\Delta G_{if}(\phi)/A$ for the free subunit with A=solvent-accessible surface area, while the filled disks represent the interfacial tension corresponding to one half of the $\Delta G_{if}(\phi)$-value for dimeric states, a valid estimation of the interfacial free-energy contribution for each subunit within a homodimer. It is clear that the inflexion point at 266K represents a critically high interfacial tension beyond which the dimeric state prevails to mitigate the higher thermodynamic cost of spanning the interface as the temperature for complex renaturation is approached. A tight correlation (R^2=0.921, Figure 3B) is obtained between dimer/monomer molar ratio X, a measure of association propensity obtained from average values of hydrodynamic radius [14], and the interfacial tension. This correlation validates the variational frustration principle and its dynamic steering role in complex formation. The protein surface regions causative of interfacial frustration [5] are shown in green on Figure 3C. Thus, hot spots in interfacial frustration (green, cf. Figure 1) are shown for the most denatured intermediate (PDB.2LYS, 257K), and the inflexion-point intermediate whose critically high interfacial frustration promotes significant dimerization at 266K, with monomeric state in PDB.2LYP and dimeric state in PDB.2LYL. The inflexion-point intermediate is attributed two $\Delta G_{if}(\phi)$-values, one as free subunit (PDB.2YLP) and the other as subunit-within-dimer (PDB.2LYL).

Finally, the non-Debye frustration pathway leading to maximal reduction of interfacial tension conducive to protein subunit association is generated following the flow chart (Figure 5) in Methods at renaturation T=298K. We computed the path to minimally frustrated hydration adopting the structural coordinates of two fully dissociated CgIR2 subunits (PDB.2LYS) as

initial condition. Four runs reveal a unique nucleating step at ~1μs, ultimately leading to formation of the fully dehydrated PP interface. This step consists of intermolecular dehydration of dehydron Leu109-Ala112 and Leu109-Leu113 by Leu43 in the complementary binding subunit (Figure 3D). The pathway leads to optimal mitigation of interfacial tension and ultimate formation of the dimer with RMSD<0.80Å relative to the homodimer crystal in PDB entry 2LYJ.

Discussion

The results presented reveal a protein "epistructure" defined by the frustration state of the aqueous interface [5]. This concept is likely to impact drug design geared at impairing protein function through target-drug association whereby the interfacial tension of the target is reduced. As shown subsequently, such a design strategy is expected to enable an ultra-selective control of the drug impact with discriminatory molecular recognition even at the level of isoforms of the target protein (>80% sequence identity). Homologous proteins typically possess high structural similarity but the differences in composition translate into differences at the epistructural level which may be harness to design selective drugs that can tell them apart.

For example, it has been recognized that the phosphoinositide 3-kinase (PI3K) family is implicated in human cancer through its role in the regulation of diverse cellular processes [27]. Yet, pan-PI3K inhibitors that bind indiscriminately with nanomolar (nM) affinity to isoforms PI3K α,β,δ (class IA) are not desirable therapeutically due to off-target side effects [28]. It has become imperative to design drugs that can tell apart such isoforms. It would be desirable to re-engineering lead compounds to improve selectivity towards the β–isoform recognized as a

crucial target in PTEN (phosphatase and tensin homologue)-deficient cancers [29]. In this regard, the lead compound *2-[2-(2,3-dihydro-indol-1-yl)-2-oxo-ethyl]-6-morpholin-4-yl-3H-pyrimidin-4-one* (Figure 4A) binds with nM affinity to PI3K β and δ isoforms (IC_{50}=4 and 28nM, respectively) [29]. The specificity towards the β-isoform may be significantly enhanced by comparing the epistructures of both isoforms and modifying the lead compound to target epistructural differences (Figures 4B,C). A first substitution (R-isomer, Figure 4A) would in principle be suitable to selectively remove upon binding a pre-formed tension hot spot of interfacial tension present only in the β-isoform (cf. Figs. 4B,C). Indeed, there is an interfacial tension hot spot in drug target PI3Kβ corresponding to frustrated water partially hydrating dehydron Ala856-Leu859 (Figure 4B). The crystal structure of the complex reveals that the methylation of the lead compound displayed in Figure 4A targets dehydron Ala856-Leu859 upon drug-target association. However, the R-substitution *does not* endow the drug with higher discriminatory power towards the β-isoform [29]. This is a-priori surprising because the Ala856-Leu859 dehydron in PI3Kβ aligns with solvent-shielded backbone hydrogen bond Ala836-Leu839 in PI3Kδ (Figure 4 B,C) which does not frustrate water nearby (cf. Figure 1). The reason for the lack of specificity is that PI3Kδ features *in lieu* of a dehydron the exposed salt bridge Asp832-Asn836 (Figure 4D) which gets dehydrated upon binding with the R-ligand, thereby becoming stabilized to an extent comparable to the dehydron [5]. The interfacial thermodynamics for ligand association reflect this observation, yielding $\Delta\Delta G_{if}(\beta,\delta,R) \approx 0 \pm \frac{0.22kcal}{mol}$ for the difference in the interfacial free energy change between the δ– and β– isoforms upon binding to the R-ligand. To achieve specificity one may instead *retain* the frustrated interfacial water molecule upon ligand binding, adopting the methylated S-ligand to that effect (Figures 4B,C). In the respective apo-states, the $\langle\phi\rangle$=1.7 water frustration of water around

dehydron 856-859 in the β-isoform becomes <ϕ>=0.6 for the water molecule closest to Asn836 in the δ-isoform. The resulting local difference in interfacial tension between PI3Kβ and PI3Kδ suggests adopting the S-ligand with retention of interfacial water upon binding, as shown in Figures 4B,C in order to enhance specificity towards the β-isoform. Thus, the already frustrated water molecule around dehydron 856-859 simply rearranges to accommodate the S-ligand with <ϕ>=1.7→<ϕ>=1.9 upon binding, whereas the <ϕ>=0.6-water hydrating salt bridge Asp832-Asn836, becomes severely frustrated, with <ϕ>=0.6→<ϕ>=1.8, upon binding to the S-ligand (Figure 4E,F). This trend is reflected in the thermodynamics: $\Delta\Delta G_{if}(\beta,\delta,S) \approx 2.8 \pm \frac{0.61 kcal}{mol}$, indicative of a significantly higher affinity of the S-ligand towards the β-isoform. Experimental results uphold this prediction [29]. With IC_{50}=23nM and 468nM for the β and δ-isoforms, respectively, the S-ligand is three times more selective towards PI3Kβ than the parental compound (IC_{50}=4nM and 28nM, respectively).

This example illustrates the power of variational computations of frustrated hydration as a guide to design ultra-selective drugs that could not be rationally conceived adopting standard structure-based approaches.

To reconcile thermodynamics and structural data on protein complexation, this work focused on the steering role of frustrated hydration in biomolecular associations. To make frustration at the aqueous interface amenable to variational methods, we introduced a “frustration field” identifying locally unmet hydrogen-bonding opportunities due to sub-nanoscale confinement at each point in space. The frustration field was optimized variationally, and the dynamics resulting from the computational search were shown to drive protein associations with great specificity in order to mitigate frustration-related interfacial tension. Our computations entailed a microstate

sampling method implemented to search for the state of minimal interfacial frustration within relaxation times for protein-water dynamic coupling. Spectroscopic measurements of dielectric relaxation upheld the basic tenets of the theory, while the crystallographic, affinity scanning and NMR record on protein assemblages attested to the validity of the minimal frustration principle.

The frustration model reported could be improved to deal with more realistic solvents reflective of physiological conditions in the cytosol. For example, ions in micromolar concentrations are likely to titrate surface charges on the protein surface or be engaged in coordination complexes that will affect the hydration patterns and interfacial frustration thereof.

The solvent-centric approach introduced in this work was shown to impact the design of ultra-selective drugs, as it highlighted the targetable nature of the aqueous interface enabling the discrimination of closely related protein isoforms that could not be operationally distinguished using structure-based designs. In this regard, the role of dehydrons as selectivity filters has been recognized by this author [5] and others [30], yet the physical underpinnings of the concept are only provided in this paper.

Methods

Computing the least frustrated interface

The least frustrated state of the aqueous interface is represented by an extremal frustration field ϕ that satisfies the Euler-Lagrange equation [17]:

$$\nabla\, \partial \frac{\mathcal{L}(\phi)}{\partial \nabla \phi} = \partial \frac{\mathcal{L}(\phi)}{\partial \phi} \ , \tag{4}$$

written explicitly as

$$\nabla^2 \phi = \left(\frac{k_B T}{4\omega}\right)(\rho_0 - \rho)\left(1 - \frac{\phi}{4}\right)^{-1} \tag{5}$$

Equation (5) constitutes the fundamental relation controlling water frustration at protein interfaces. For small frustration 0<ϕ<<4, the linear approximation $dS_{if}(\phi) \approx -[k_B(\rho_0 - \rho)\,(\frac{\phi}{4})]d\vec{r}$ yields the simplified equation $\nabla^2\phi = \left(\frac{k_BT}{4\omega}\right)(\rho_0 - \rho)$.

As shown in Supplementary Material (SM), to obtain ω, we use Equation (3) applied to an equilibrating computation of ϕ obtained as a function of the distance X to a rigid nonpolar interface [6,31] consisting of a concave region of fixed curvature radius. The equilibration is carried out for 1ns, a time incommensurably larger than the relaxation timescale for the decoupled water lattice (~1ps) [15]. Thus, at T= 298K, Equation 3 yields $\omega = 9.18 \times 10^{-20}\, mJ\text{Å}^{-1}$, an estimation validated by contrasting it with the Young-Laplace computation of interfacial free energy [32].

Microstate sampling strategy to steer the search for the least frustrated state

As shown in the microstate sampling scheme in Figure 5, we can select at each stage of the simulation a representative from the set of *a-priori* equi-probable microstates to ensure a *probable* decrease in $\Delta G_{if}(\phi) = \Im(\phi)$, thus steering the search towards a state of minimal frustration. Let $[\mathbf{Q_n}, \mathbf{P_n}]$ denote the state of the system at stage *n*, specified by $\mathbf{Q_n} = \mathbf{Q_n^p} \oplus \mathbf{Q_n^w} =$ atomic coordinates vector, and $\mathbf{P_n} = \mathbf{P_n^p} \oplus \mathbf{P_n^w}$ = momentum vector, both decomposable as direct sums of protein (p) and water (w) contributions. The next state $[\mathbf{Q_{n+1}}, \mathbf{P_{n+1}^p} \oplus \mathbf{P_{n+1}^w}]$ resulting after 5ps of MD simulation is retained depending on a comparison of the scalar fields $\phi_n = \phi_n(\mathbf{Q_n})$ and $\phi_{n+1} = \phi_{n+1}(\mathbf{Q_{n+1}})$. State $[\mathbf{Q_{n+1}}, \mathbf{P_{n+1}^p} \oplus \mathbf{P_{n+1}^w}]$ is retained provided that either one of the following conditions holds:

$$\Im(\phi_{n+1}) \leq \Im(\phi_n),$$

$$\Im(\phi_{n+1}) > \Im(\phi_n) \text{ and } r \leq \exp\{-\beta[\Im(\phi_{n+1}) - \Im(\phi_n)]\}$$

with r=realization value of random variable adopting real values in the interval [0,1] and $\beta=[k_BT]^{-1}$. If $\Im(\phi_{n+1}) > \Im(\phi_n)$ and $r > exp\{-\beta[\Im(\phi_{n+1}) - \Im(\phi_n)]\}$, we perform 5ps MD runs on 1000 initial randomly selected *a-priori* equi-probable microstates $[\mathbf{Q_n}, \mathbf{P_n^p} \oplus \mathbf{P_n^{w(j)}}]$, j = 1,2, ...,1000, with randomly chosen water momenta $\mathbf{P_n^{w(j)}}$ ($\mathbf{P_n^{w(1)}} = \mathbf{P_n^w}$) subject to the constant kinetic energy constraint, and select the state $\left[\mathbf{Q_{n+1}^*}, \mathbf{P_{n+1}^{p*}} \oplus \mathbf{P_{n+1}^{w*}}\right]$ as the (*n+1*)-destiny state that realizes the lowest value $\Im(\phi_{n+1}\left(\mathbf{Q_{n+1}^{(j)}}\right))$.

Identification of packing defects in proteins

To identify packing defects (dehydrons) in soluble proteins, a dehydron calculator may be used. One such calculator that uses structural PDB coordinates as input is obtained as a Pymol plugin and has been deposited in the Pymol script repository at the url:

https://github.com/Pymol-Scripts/Pymol-script-repo/blob/master/plugins/dehydron.py

Supplementary Material: The following Supplementary Material is available for this article: computational estimation of constant ω, experimental validation of Equation (2).

Acknowledgments: The author is indebted to his former students at Rice University, Drs. Jianping Chen and Xi Zhang, for implementing the microstate sampling strategy on the Rice Computational Research Cluster.

References

[1] Cheng YK & Rossky PJ (1998) Surface topography dependence of biomolecular hydrophobic hydration. *Nature* **392**, 696-699.

[2] Wang L, Berne BJ & Friesner RA (2011) Ligand binding to protein binding pockets with wet and dry regions. *Proc Natl Acad Sci USA* **108**, 1326-1330.

[3] Giovanbattista N, Lopez CF, Rossky PJ & Debenedetti PG (2008) Hydrophobicity of protein surfaces: Separating geometry from chemistry. *Proc Natl Acad Sci USA* **105**, 2274-2279.

[4] Snyder PW, Mecinović J, Moustakas DT, Thomas III SW, Harder M, Mack ET, Lockett MR, Héroux A, Sherman W & Whitesides GM (2011) Mechanism of hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc Natl Acad Sci USA* **108**, 17889-17894.

[5] Fernández Stigliano A (2015) *Biomolecular Interfaces: Interactions, Functions and Drug Design*, pp. 1-51. Springer, Berlin.

[6] Fernández Stigliano A (2013) Breakdown of the Debye polarization ansatz at protein-water interfaces. *J Chem Phys* **138**, 225103.

[7] Remsing RC, McKendry IG, Strongin DR, Klein ML & Zdilla MJ (2015) Frustrated Solvation Structures Can Enhance Electron Transfer Rates. *J Phys Chem Lett* **6**, 4804−4808.

[8] Lee M, Salsbury F & Olson MA (2004) An efficient hybrid explicit/implicit solvent method for biomolecular simulations. *J Comp Chem* **25**, 1967-1978.

[9] Allen MP & Tildesley DJ (1989) *Computer Simulation of Liquids*. Clarendon Press, Oxford, UK.

[10] Bachmann M (2014) *Thermodynamics and Statistical Mechanics of Macromolecular Systems*. Cambridge University Press, Cambridge, UK.

[11] Lee MS & Olson MA (2013) Comparison of volume and surface area nonpolar solvation free energy terms for implicit solvent simulations. *J Chem Phys* **139**, 044119.

[12] A. Fernández (2011) Nanoscale thermodynamics of biological interfacial tension. *Proc Roy Soc A* **467**, 559-568.

[13] Clackson T & Wells JA (1995) A hot spot of binding energy in a hormone-receptor interface. *Science* **267**, 383-386.

[14] Jaremko M, Jaremko L, Kim H, Cho M, Schwieters C, Giller K, Becker S & Zweckstetter M (2013) Cold denaturation of a protein dimer monitored at atomic resolution. *Nature Chem Biol* **9**, 264-270.

[15] Pal SK, Peon J & Zewail AH (2002) Biological water at the protein surface: dynamical solvation probed directly with femtosecond resolution. *Proc Natl Acad Sci USA* **99**, 1763-1767.

[16] Li B, Liu P, Xu Z & Zhou S (2013) Ionic size effects: generalized Boltzmann distributions, counterion stratification and modified Debye length. *Nonlinearity* **26**, 2899-2922.

[17] Evans LC (2010) *Partial Differential Equations*. American Mathematical Society, Providence, RI.

[18] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW & Klein ML (1983), Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926-938.

[19] Sierra MB, Accordino SR, Rodriguez-Fris A, Morini MA, Appignanesi GA & Fernández Stigliano A (2013) Protein packing defects "heat up" interfacial water. *Eur Phys J E* **36**, 62.

[20] Wang J, Wolf RM, Caldwell JW, Kollman PA & Case DA (2004) Development and testing of a general amber force field. *J Comput Chem* **25**, 1157-1174.

[21] Darden T, York D & Pedersen LJ (1993) Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J Chem Phys* **98**, 10089-10092.

[22] Zhang L, Wang L, Kao Y, Qiu W, Yang Y, Okobiah O & Zhong D (2007) Mapping hydration dynamics around a protein surface. *Proc Natl Acad Sci USA* **104**, 18461-18466.

[23] Ashkenazi A, Presta L, Marsters S, Camarato J, Rosenthal K, Fendly B & Capon D (1990) Mapping the CD4 binding site for human immunodeficiency virus by alanine-scanning mutagenesis. *Proc Nat Acad Sci USA* **87**, 7150-7154.

[24] Castro MJ & Anderson S (1996) Alanine point-mutations in the reactive region of bovine pancreatic trypsin inhibitor: effects on the kinetics and thermodynamics of binding to beta-trypsin and alpha-chymotrypsin. *Biochemistry* **35**, 11435-11446.

[25] Kuhlmann UC, Pommer A, Moore GR, James R & Kleanthous C (2000) Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. *J Mol Biol* **301**, 1163-1178.

[26] Bottger V, Bottger A, Garcia-Echeverria C, Chène P, Hochkeppel HK, Sampson W, Ang K, Howard SF, Picksley SM & Lane DP (1997) Molecular characterization of the hdm2-p53 interaction. *J Mol Biol* **269**, 744-756.

[27] Maira SM, Voliva C & Garcia-Echeverria C (2008) Class IA phosphatidylinositol 3-kinase: from their biologic implication in human cancers to drug discovery. *Expert Opin Ther Targets* **12**, 223-238.

[28] Ihle NT & Powis G (2010) The biological effects of isoform-specific PI3-kinase inhibition. *Curr Opin Drug Discovery Devel* **13**, 41-49.

[29] Certal V, Carry JC, Halley F, Virone-Oddos A, Thompson F, Filoche-Romme B, El-Ahmad Y, Karlsson A, Charrier V, Delorme C, Rak A, Abecassis PY, Amara C, Vincent L, Bonnevaux H, Nicolas JP, Mathieu M, Bertrand T, Marquette JP, Michot N, Benard T, Perrin MA, Lemaitre O, Guerif S, Perron S, Monget S, Gruss-Leleu F, Doerflinger G, Guizani H, Brollo M, Delbarre L, Bertin L, Richepin P, Loyau V, Garcia-Echeverria C, Lengauer C & Schio L (2014) Discovery and Optimization of Pyrimidone Indoline Amide PI3Kβ Inhibitors for the Treatment of Phosphatase and Tensin Homologue (PTEN)-Deficient Cancers. *J Med Chem* **57**, 903−920.

[30] Robinson DD, Sherman W & Farid R (2010) Understanding Kinase Selectivity through Energetic Analysis of Binding Site Waters. *ChemMedChem* **5**, 618−627.

[31] Fernández A (2015) Quantum theory of interfacial tension quantitatively predicts spontaneous charging of nonpolar aqueous interfaces. *Phys Lett A* **379**, 2405-2408.

[32] Matsumoto M & Tanaka K (2008) Nano bubble - Size dependence of surface tension and inside pressure. *Fluid Dyn Res* **40**, 546-553.

Figure Legends

Figure 1. Scheme of frustrated hydration funneling protein association: A solvent-exposed backbone hydrogen bond, known as *dehydron* [5,6], constitutes a packing defect that generates frustration at the interface (g<4) as a water molecule hydrogen-bonds to the backbone carbonyl (exploiting the lone an unused electron pair in the carbonyl oxygen) and thus gets confined in the dehydron cavity. The reader is reminded that the carbonyl oxygen is a double proton acceptor because of its two lone electron pairs in the outer-shell orbitals. The transference of the frustrated molecule (marked by the asterisk) to the bulk quasi-tetrahedral lattice of hydrogen bonds is thermodynamically favored and prompts the PP association as an expedient to reduce the interfacial tension.

Figure 2. Water frustration as steer of protein associations probed by alanine scanning: **A**) Steering power of microstate sampling defined by the functional in Equation (3) (Methods). The search for the minimally frustrated interface state is evidenced by the time evolution of $\Delta G_{if}(\phi) = \Im(\phi)$ during 500ps-MD trajectories. The single trajectory is in solid thin line, the average over 20 trajectories in thick grey line, while the dashed grey line represents an average over 20 control 500ps-trajectories with no microstate sampling. Initial structural coordinates were obtained for the corresponding hGHbp receptor chain in the free state that forms a 1:1 complex with the human growth hormone (hGH) (PDB.3HHR). The minimal $\Delta G_{if}(\phi)$-value is reached within the 350-400ps range. **B)** Correlation (R^2=0.88) between $-\Delta\Delta G_{if}$ and $\Delta\Delta G_a$ for site-directed mutational substitution of interfacial residues in hGHbp, yielding the approximate relation $-\Delta\Delta G_{if} \approx \Delta\Delta G_a$. **C)** Interfacial tension reduction in the MDM2-p53-transactivation domain complex (PDB.1YCR). Hot spots are identified as preformed backbone hydrogen bonds exposed to water (dehydrons) in the free p53 subunit, undergoing displacement of vicinal water upon association. The displacement results from the approach of a side chain of the binding partner and is marked by a thin line from the α-carbon of the displacing side chain to the center of the dehydron. The dehydrons from the MDM2 peptide are induced upon binding. Hydrogen bonds are represented as lines joining paired residues in a virtual-bond-α-carbon representation. The ribbon rendering on the right is an aid to the eye. **D)** Displacement by MDM2 residue L26 of frustrated water around preformed hot-spot dehydrons in the p53-transactivation domain upon association. **E)** Residue classification for PDB-reported protein complexes with available affinity scanning data according to $\Delta\Delta G_a$-values and $-\Delta\Delta G_{if}$-values. In the upper row for each PP interface, the classification is established according to the ranges $\Delta\Delta G_a \geq 3kcal/mol$ (red), $1kcal/mol \leq \Delta\Delta G_a < 3kcal/mol$ (white), and $\Delta\Delta G_a < 1kcal/mol$ (blue). In the lower row residues are labeled following the second classifier ($-\Delta\Delta G_{if}$), according to the ranges $-\Delta\Delta G_{if} \geq 3kcal/mol$ (red), $\frac{1kcal}{mol} \leq -\Delta\Delta G_{if} < 3kcal/mol$ (white), and $-\Delta\Delta G_{if} < 1kcal/mol$ (blue).

Figure 3. Interfacial frustration and dimerization propensity of cold denaturation intermediates of a protein complex: **A)** Free-energy cost of interfacial frustration per unit surface area for the free monomeric subunit (open disks) and the subunit within the dimer (filled disks) for intermediate states resulting from progressive cold denaturation of the native homodimeric repressor protein CγIR2. The PDB files for the intermediates (absolute temperature in brackets) are 2LYJ (298K), 2LYK (270K), 2LYL (266K, dimer), 2LYP (266K, monomer), 2LYQ (262K), 2LYR (259K) and 2LYS (257K). **B)** Correlation between molar dimer/monomer ratio and free energy cost of interfacial frustration per unit surface area for the free (uncomplexed) subunit in the cold-denaturation intermediates. **C)** Hot spots of interfacial frustration (shown in green on the protein structure) for denaturation intermediate at 257K (PDB.2LYS), and the inflexion-point intermediate with monomeric state in PDB.2LYP and dimeric state in PDB.2LYL. The inflexion point intermediate is attributed two $\Delta G_{if}(\phi)$-values, one as free monomer and the other as subunit within the dimer. Hot spots are shown as solvent-exposed backbone hydrogen bonds or *dehydrons* (cf. Figure 1). The chain conformation featuring hot-spot location is schematically represented by virtual bonds joining α-carbons. **D)** Optimal frustration pathway leading to protein subunit association generated

following the microstate sampling strategy in Methods (T=298K, computation time: 0.1ms, number of iterations: 2-21×10^9). Initial conditions were provided by the structural coordinates of two dissociated CgIR2 subunits (PDB.2LYS). The chains are represented as virtual bonds joining α-carbons and intermolecular water exclusion from a PP interfacial dehydron is indicated by a line from the α-carbon of the approaching residue that promotes water removal to the center of the interfacial hydrogen bond. The latter are displayed as lines joining the α-carbons of the paired residues. Four runs reveal a unique nucleating configuration (**1**) at ~1μs (±0.31 μs). Other reproducible configurations (**2, 3**) occur at ~19μs and ~90μs, with the latter (**3** and ribbon representation) presenting the fully dehydrated PP interface with RMSD<0.80Å relative to the homodimer crystal in PDB entry 2LYJ.

Figure 4. Drug design guided by the interfacial tension hot spots of the target protein: **A)** Chemical structure of lead compound with nM affinity towards the β and δ-isoforms of PI3K. Two methylation substitutions intended to improve selectivity are shown: R isomer in red, S-isomer in grey. The latter is predicted to endow the compound with discriminatory power, enhancing its selectivity towards the β-isoform. The R-isomer is named using IUPAC convention. **B**) Hot spot of interfacial tension in drug target PI3Kβ corresponding to dehydron Ala856-Leu859 (inset, PDB.4BFR). Dehydrons are displayed as green segments joining the α-carbons of residues paired by solvent-exposed backbone hydrogen bonds (cf. Figure 1). The structure of the complex (panel with space-filling atoms) reveals that the S-isomer methylation shown in Figure 4A targets dehydron Ala856-Leu859 with retention of hydrating molecule (red circle). The ligand binding does not appreciably increase the frustration of the water molecule that was already frustrated due to sub-nanoscale confinement within the dehydron cavity and re-accommodates upon binding to the S-ligand. **C)** Dehydron pattern for PI3Kδ (PDB.4V0I). The Ala856-Leu859 dehydron in PI3Kβ aligns with backbone hydrogen bond Ala836-Leu839 in PI3Kδ, which is not solvent-exposed and hence it does not cause significant water frustration (ϕ<1). This local difference in interfacial tension between PI3Kβ and PI3Kδ suggests the modifications featured in Figure 5A as a means to enhance drug selectivity towards the β-isoform. **D)** Salt bridge (dashed line) in PI3Kδ absent in PI3Kβ. **E)** Scheme of frustration effects on dehydron-vicinal water upon binding the R- and S-isomer. Upon binding, the R-isomer excludes vicinal water, thus stabilizing the dehydron, while the S-isomer retains a water molecule but does not increase its frustration relative to the "apo" form. The frustrated loner pair on the oxygen is marked by an asterisk. **F)** Scheme of frustration effects on water vicinal to a salt bridge upon binding the R- and S-isomer. Upon binding, the R-isomer excludes vicinal water, thus stabilizing the salt bridge, while the S-isomer retains a water molecule significantly increasing its frustration relative to the "apo" form.

Figure 5. Scheme of microstate sampling strategy to steer the search towards the state of minimal frustration. Retention or further selection step for a 5ps-MD iteration controlled by interfacial frustration functional $\Im(\phi)$. The upper box indicates flow under which the destiny microstate $[\mathbf{Q_{n+1}}, \mathbf{P^{p}_{n+1}} \oplus \mathbf{P^{w}_{n+1}}]$ is retained in the n→n+1 iteration based on a comparison of scalar fields $\phi_n = \phi_n(\mathbf{Q_n})$ and $\phi_{n+1} = \phi_{n+1}(\mathbf{Q_{n+1}})$. The lower box indicates the sampling procedure that takes place when the destiny state is not retained and further exploration is required.

FIGURE 1

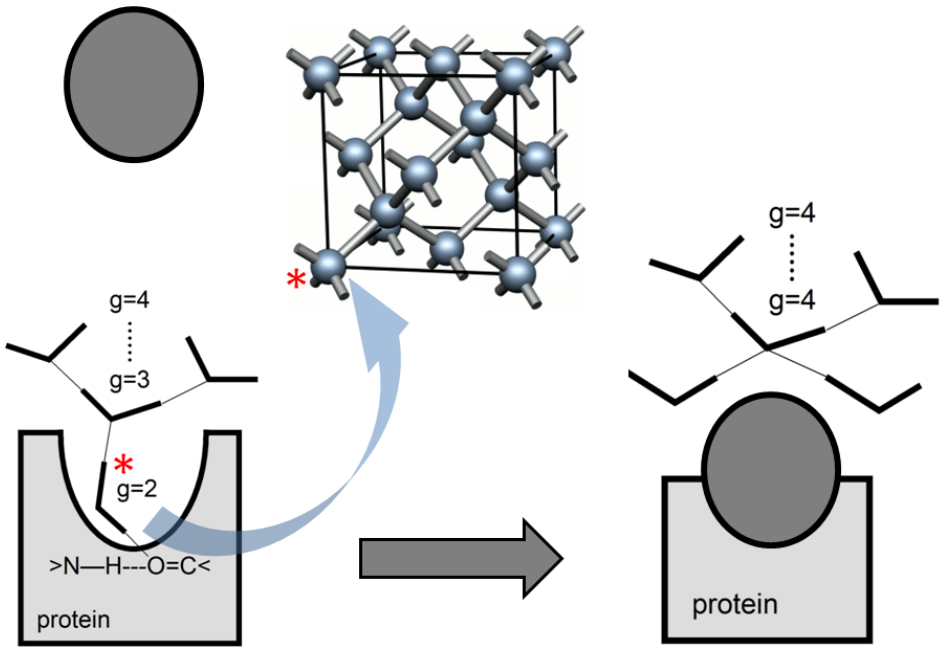


FIGURE 2 A

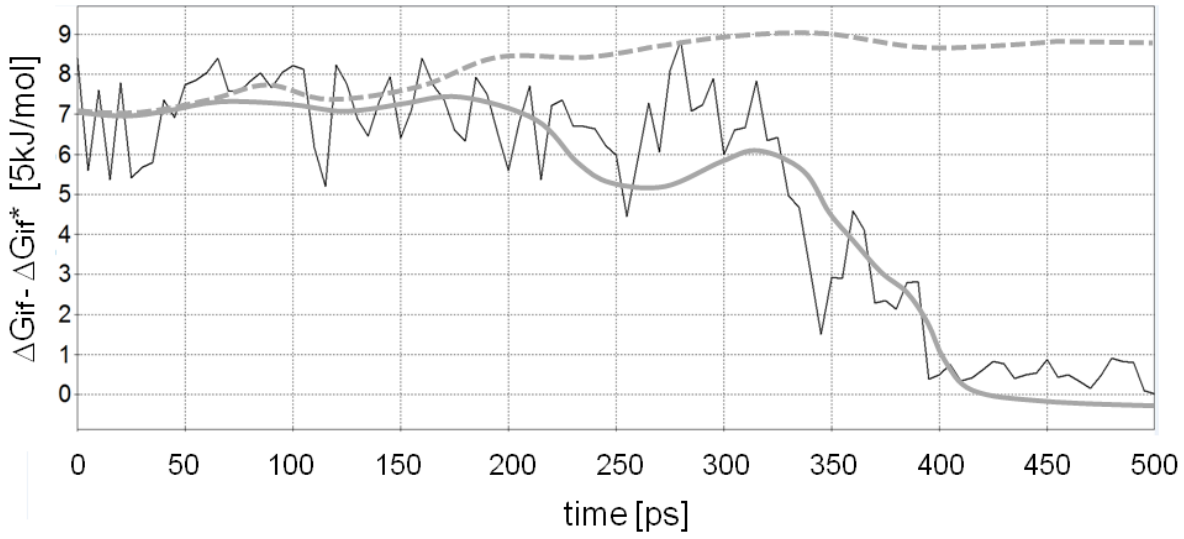


FIGURE 2B

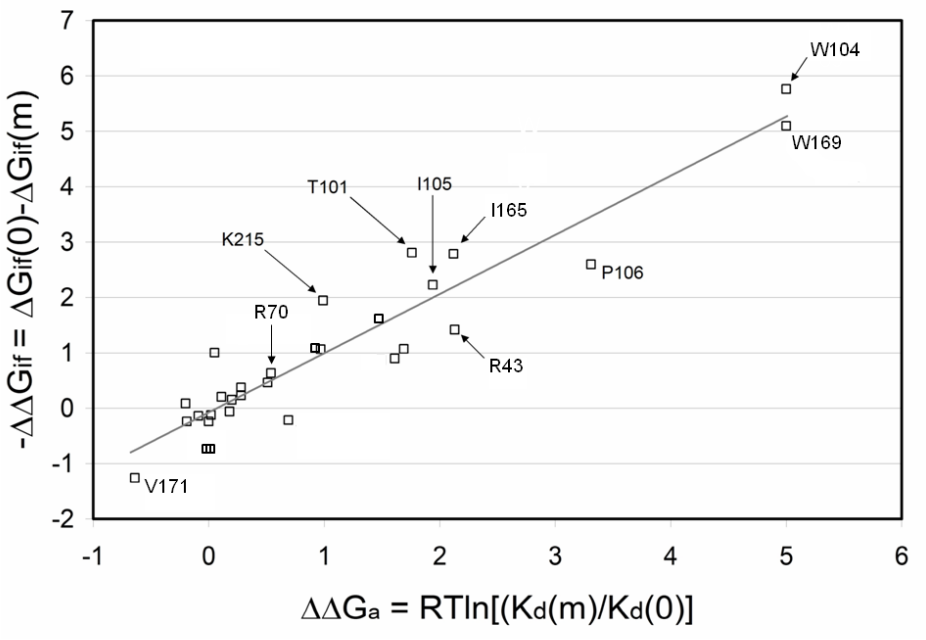


FIGURE 2C

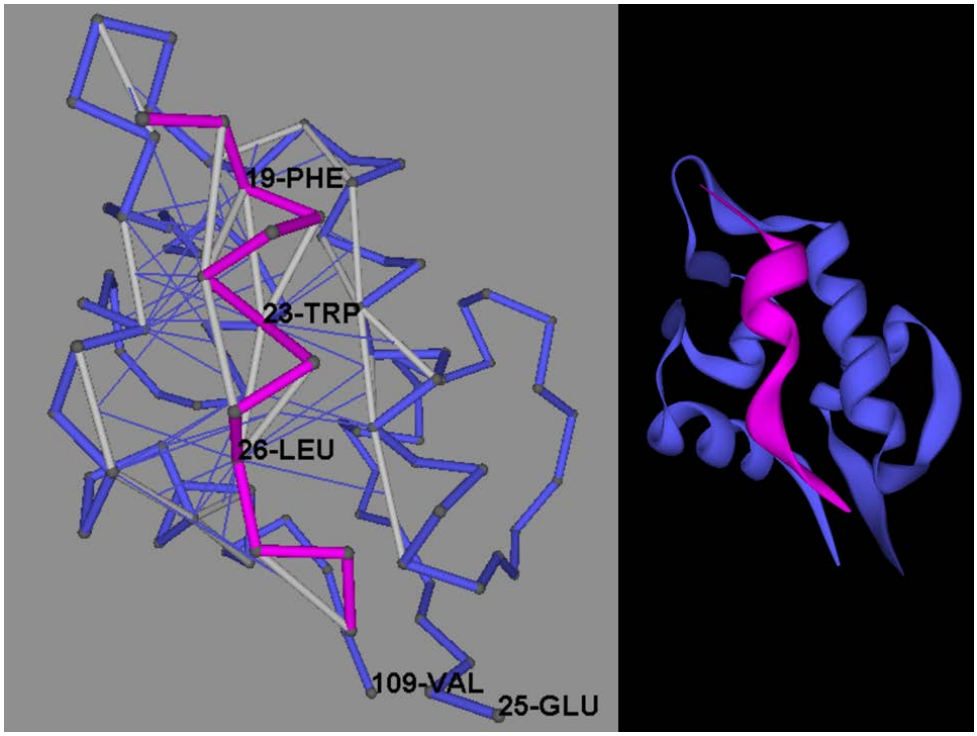


FIGURE 2D

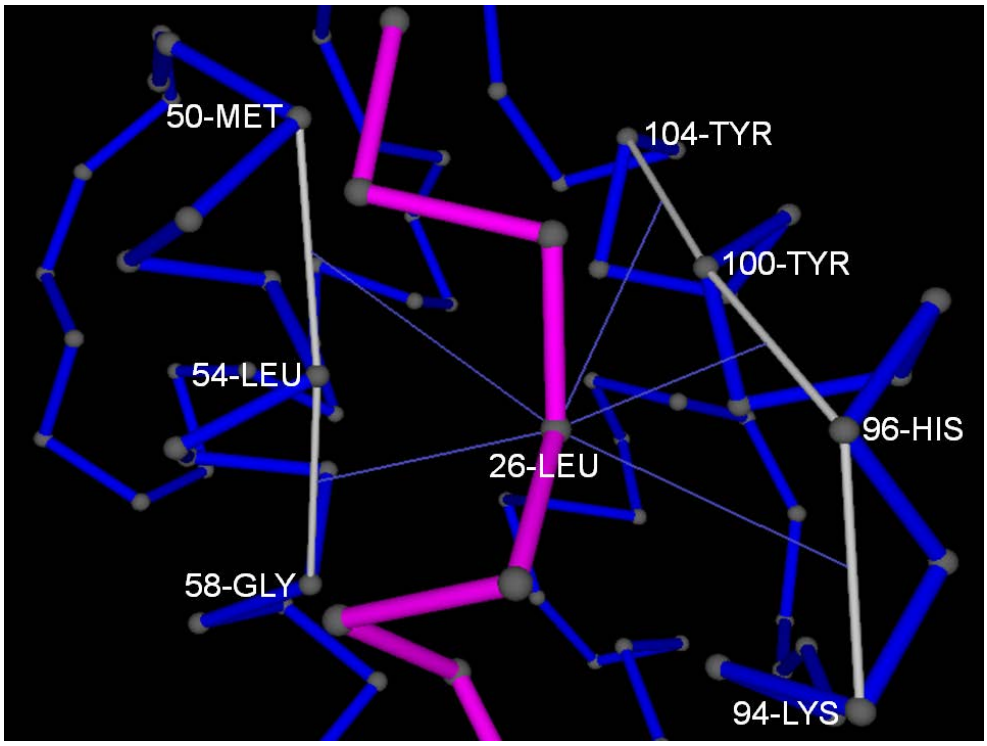


FIGURE 2E

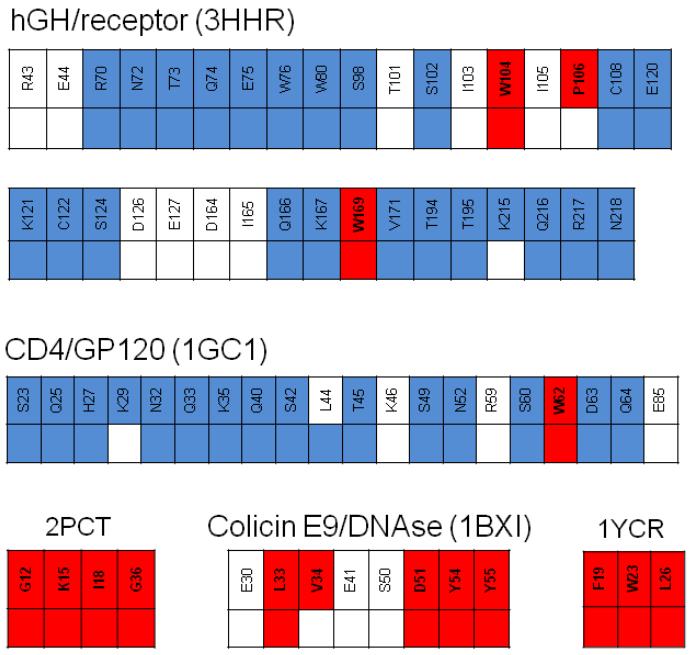


FIGURE 3A,B

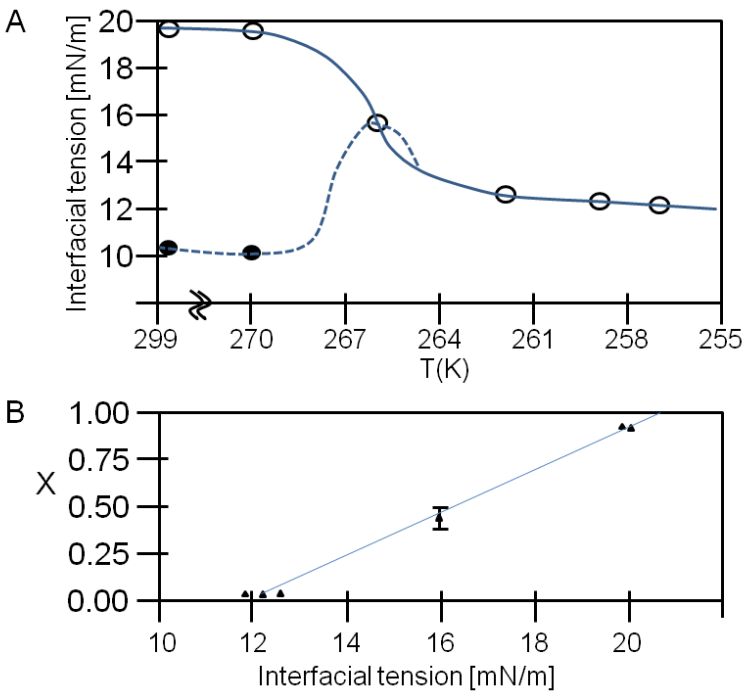


FIGURE 3C

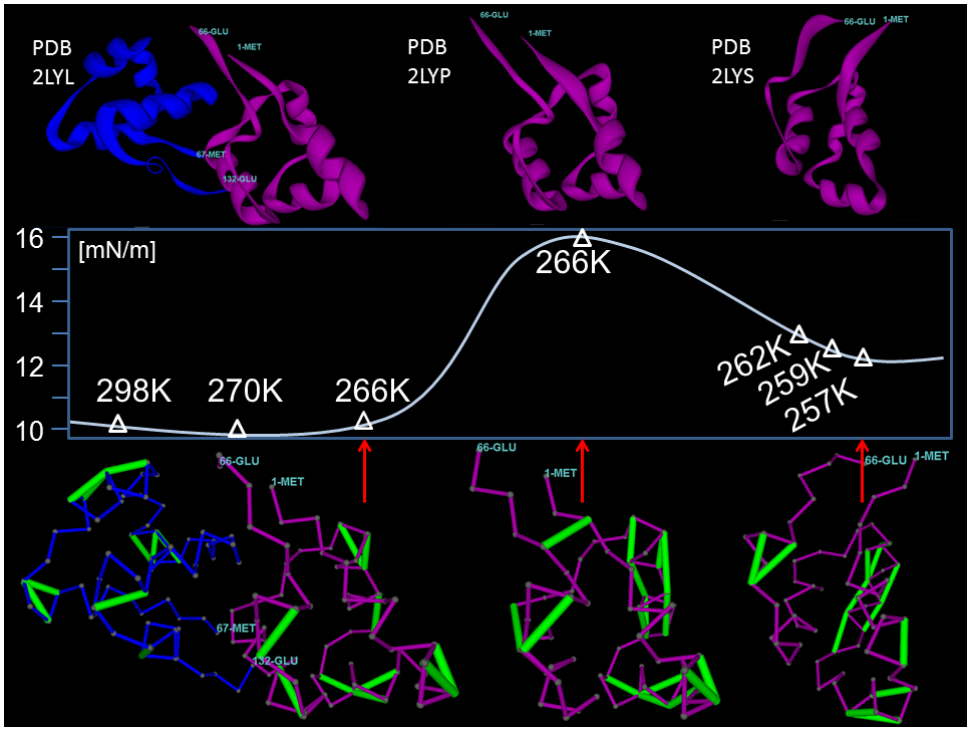


FIGURE 3D

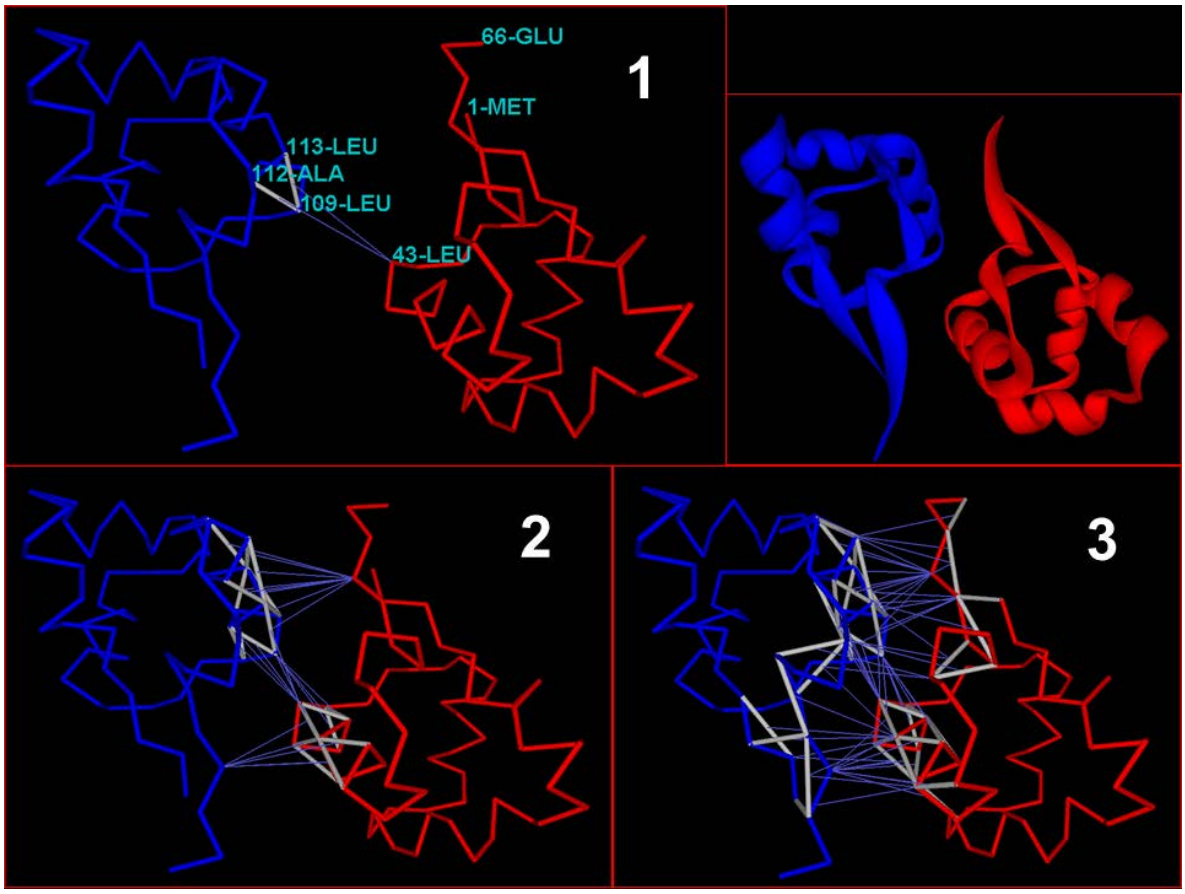
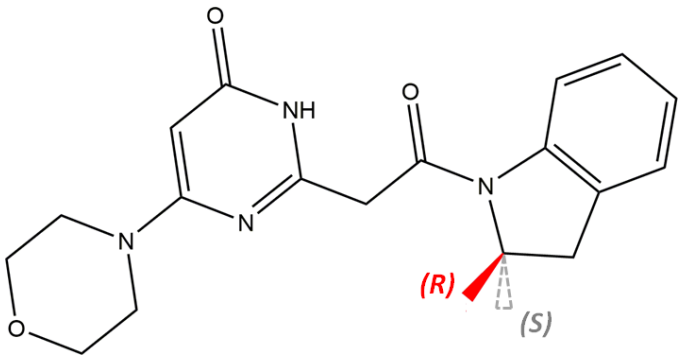


FIGURE 4A



(*R*)-2-(2-(2-methylindolin-1-yl)-2-oxoethyl)-6-morpholinopyrimidin-4(3*H*)-one

FIGURE 4B

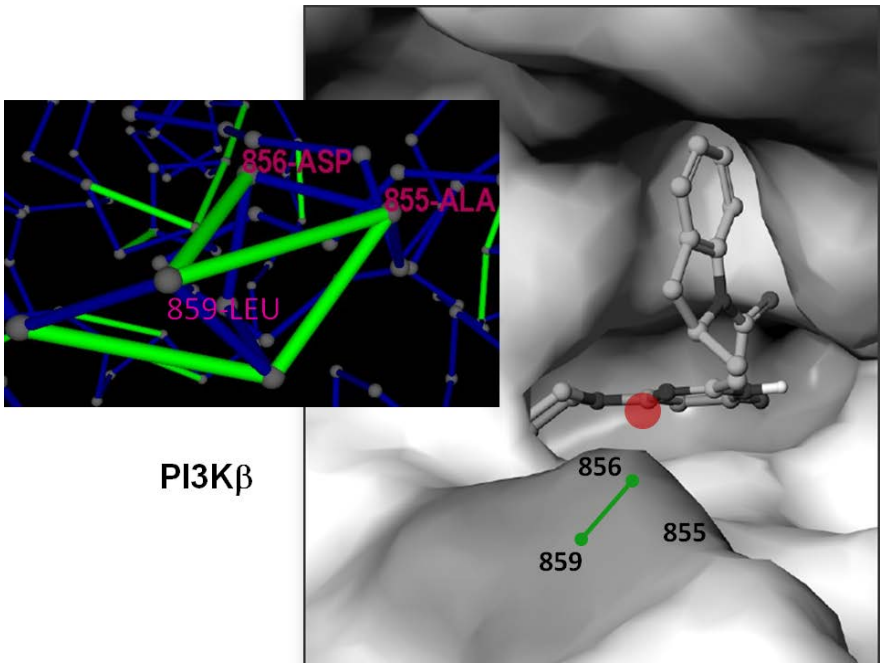


FIGURE 4C

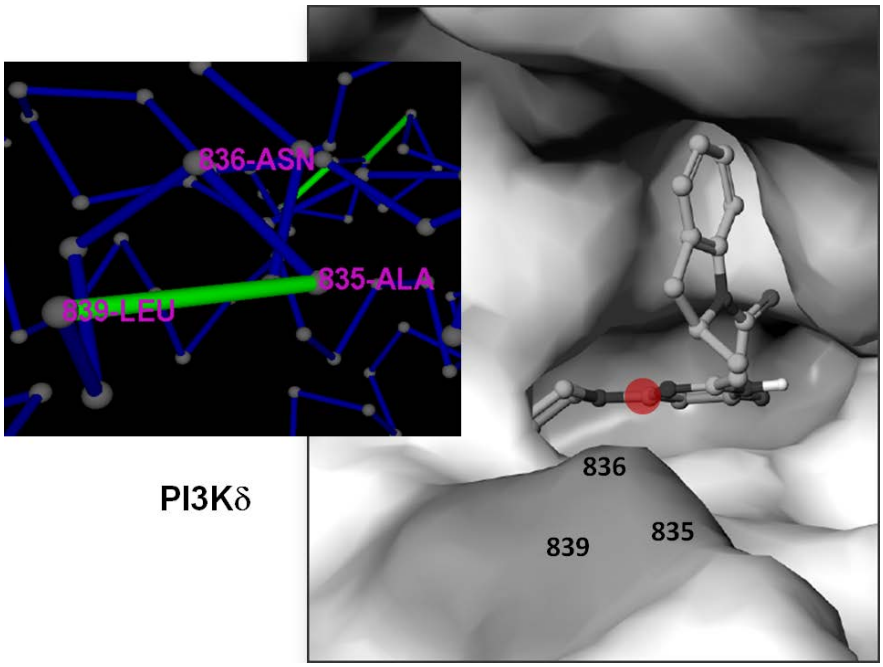


FIGURE 4D

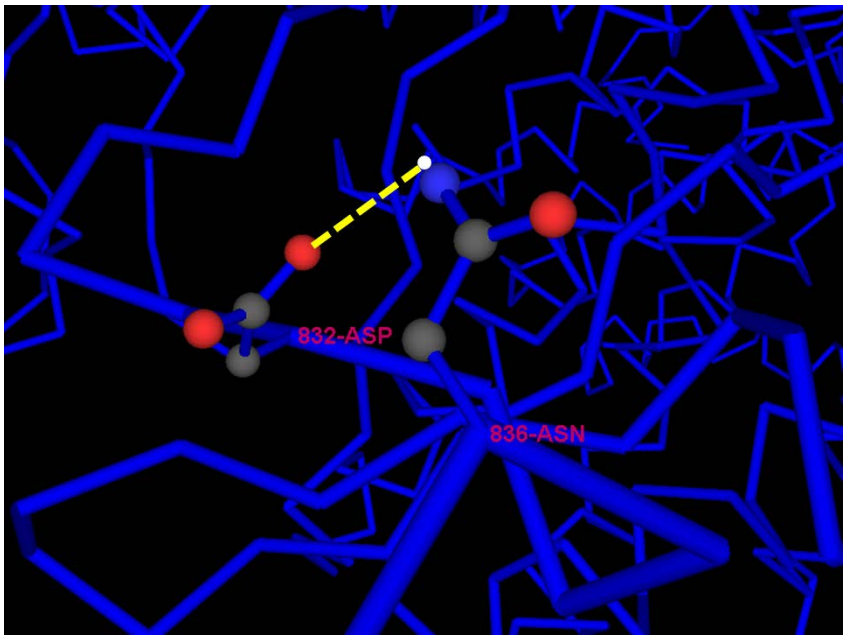


FIGURE 4E

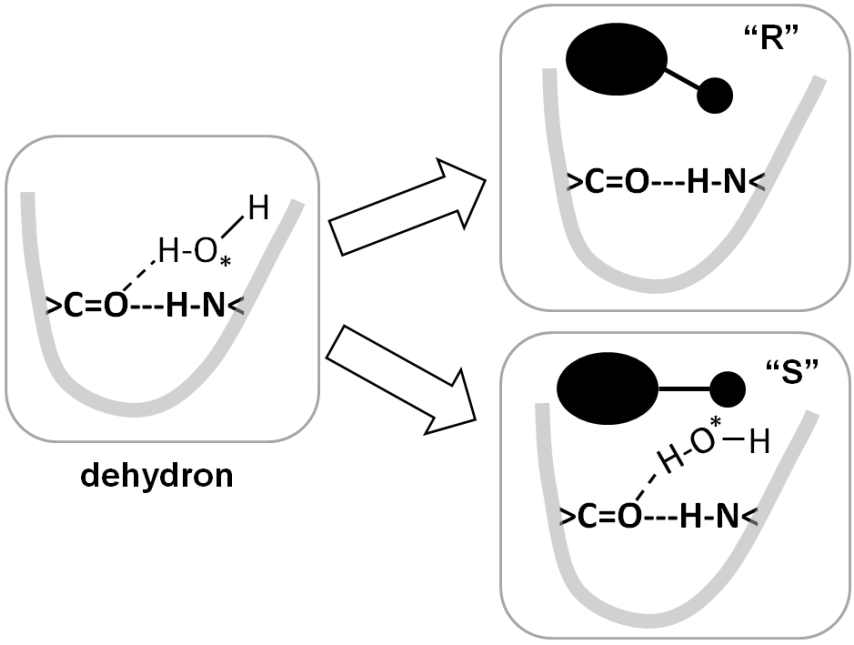


FIGURE 4F

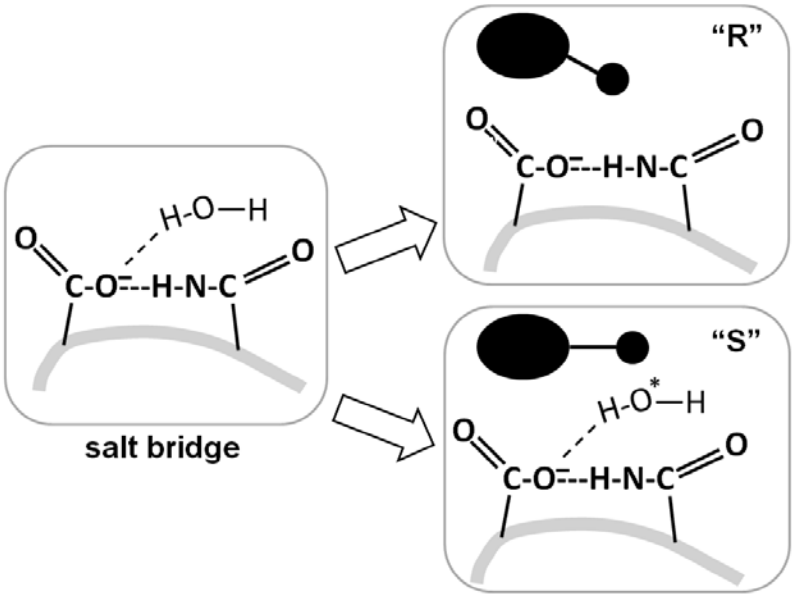


FIGURE 5