

Exploring the Free Energy Landscape and Thermodynamics of Protein-Protein Association: HIV-1 Integrase Multimerization Induced by an Allosteric Inhibitor

6/24/2020

*C. Tse, L. Wickstrom, M. Kvaratskhelia, E. Gallicchio, R. Levy, N. Deng**

Keywords: HIV-1 Integrase, Allosteric HIV-1 Integrase Inhibitor, Protein-Protein Association, Potential of Mean Force, Binding Enthalpy, Entropy.

Abstract

We report the free energy landscape and thermodynamics of the protein-protein association responsible for the drug-induced multimerization of HIV-1 integrase (IN). Allosteric HIV-1 integrase inhibitors (ALLINIs) promote aberrant IN multimerization by bridging IN-IN intermolecular interactions. However, the thermodynamic driving forces and kinetics of the multimerization remain largely unknown. Here we explore the early steps in the IN multimerization by using umbrella sampling and unbiased molecular dynamics simulations in explicit solvent. In direct simulations, the two initially separated dimers spontaneously associate to form near-native complexes that resemble the crystal structure of the aberrant tetramer. Most strikingly, the effective interaction of the protein-protein association is very short-ranged: the two dimers associate rapidly within tens of nanoseconds when their binding surfaces are separated by $d \leq 4.3 \text{ \AA}$ (less than two water diameters). Beyond this distance, the oligomerization kinetics appears to be diffusion controlled with a much longer association time. The free energy profile also captured the crucial role of ALLINI in promoting multimerization, and explained why several CTD mutations are remarkably resistant to the drug-induced multimerization. The results also show that at small separation the protein-protein binding process contains two consecutive phases with distinct thermodynamic signatures. First, inter-protein water molecules are expelled to the bulk resulting in a small increase in entropy, as the solvent entropy gain from the water release is nearly cancelled by the loss of side chain entropies as the two proteins approach each other. At shorter distances, the two dry binding surfaces adapt to each other to optimize their interaction energy at the expense of further protein configurational entropy loss. While the binding interfaces feature clusters of hydrophobic residues, overall, the protein-protein association in this system is driven by enthalpy and opposed by entropy.

Statement of Significance

Elucidating the energetics and thermodynamic aspects of protein-protein association is important for understanding this fundamental biophysical process. This study provided a more complete physical picture of the protein-protein association responsible for the drug-induced HIV-1 integrase multimerization. The results captured the critical role of the inhibitor, and accounted for the effects of mutations on the protein association. Remarkably, the effective range of the protein-protein attractive funnel is found to be very short, at less than two layers of water, despite the fact that the two binding partners carry opposite net charges. Lastly, entropy/enthalpy decomposition shows that the solvent release from the inter-protein region into the bulk is more than offset by the loss of the solute configurational entropy due to complexation.

Introduction

Noncovalent and specific protein-protein association plays a crucial role in a wide range of fundamentally important biological processes, such as signal transduction and antibody-antigen recognition¹. In recent years, protein-protein interactions are increasingly targeted for therapeutics development²⁻³. In this work, we characterize at an atomic level of detail the free energy landscape and the thermodynamic driving forces for the protein-protein association underlying the drug-induced HIV-1 integrase multimerization⁴⁻⁷. HIV-1 integrase (IN) is an important therapeutic target for the development of antiviral therapy², because of the vital roles IN plays in the life cycle of HIV. Allosteric HIV-1 integrase inhibitors (ALLINIs) bind at the dimer interface of the IN catalytic core domain (CCD) and inhibit HIV-1 by promoting aberrant IN multimers, which impairs both the catalytic activity of IN and virus core maturation.⁸⁻¹⁰ The mechanism by which ALLINIs promote aberrant multimerization has previously been revealed by our protein-protein docking simulations which suggested that the multimerization is mediated by ALLINI which bridges IN-IN interactions between the catalytic core domains (CCD) of one IN dimer and the C-terminal domain (CTD) of another dimer. The central prediction from the modeling study of the IN aberrant multimerization, i.e. the CCD and CTD intersubunit interaction bridged by ALLINI⁵, was confirmed by the experimental crystal structure later obtained containing full-length IN and ALLINI¹¹.

Despite the progress in understanding the ALLINI-induced aberrant multimerization, several important aspects of the molecular mechanism of the aberrant multimerization need to be better understood in order to inform rational inhibitor design. What is the predominant thermodynamic driving force for the intermolecular association of CTD and CCD bridged by ALLINI? How does the association free energy landscape of the IN multimerization look like? At the molecular level, why do specific amino acid substitutions in the CTD such as Y226D and K266A block the ALLINI-induced multimerization? What are the roles of the water molecules in the intervening space between the CCD and CTD in the association? To date, only limited, indirect experimental data regarding the drug-induced IN multimers are available.

The only experimental structure of an aberrant multimer was determined at a relatively low resolution of 4.4 Å¹¹. While our modeling correctly predicted the central mechanism of action of the drug-induced

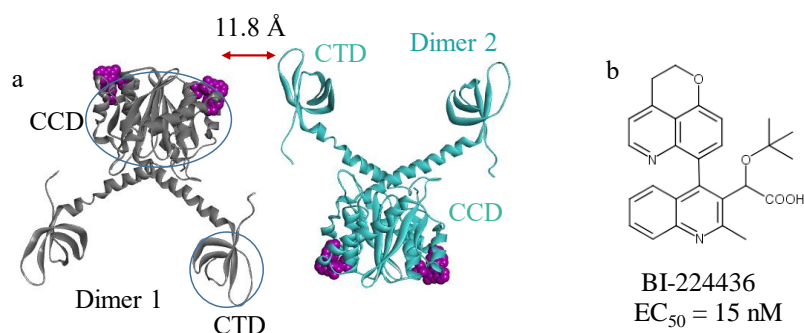


Figure 1. a. Two HIV-1 integrase dimers separated by interfacial gap $d = 11.8 \text{ \AA}$ are solvated in a box of water (not shown). b. Chemical structure of the ALLINI BI224436.

multimerization, the use of protein-protein docking precludes the characterization of the association mechanism directly influenced by solvent.

To address the mechanistic questions regarding the IN multimerization at an atomic level of detail, here we employ molecular dynamics (MD) simulations in explicit solvent to study the CCD and CTD association mediated by ALLINI. In recent years, because of the advances in methodology, algorithm and computer hardware, MD simulation and associated techniques have increasingly been successfully used to study the energetics and kinetics of protein-protein binding¹²⁻¹⁵. Here, we first directly simulate the formation of a tetramer starting from two dimers initiated from dissociated states with different separations between the CTD and CCD. The results show that the experimental multimer structure can be obtained using unbiased physical models in explicit aqueous solutions. We then quantitatively explore the free energy landscape of the protein-protein association using umbrella sampling with varying conditions to investigate the role of ALLINI and the different amino acid mutations that influence the multimerization. Entropy/enthalpy decomposition of the free energy reveals a detailed thermodynamic picture of protein-protein association that is dominated by favorable overall enthalpy, despite the presence of hydrophobic interactions traditionally believed to be favored by entropy. The study provides a wealth of atomistic information concerning the biophysics of drug-induced IN aberrant multimerization that may inform the design of new allosteric IN inhibitors.

Materials and Methods

The GROMACS program 2016-3¹⁶⁻¹⁷ was used to performed the MD simulations in this work. The starting structure of the IN dimers (Fig. 1) was taken from the structure of the aberrant tetramer containing the ALLINI BI224436 described in the previous report⁴. The protein was modeled by the AMBER parm99ILDN force field¹⁸, while the ligand by the Amber GAFF¹⁹ and the AM1-bcc charge model²⁰. Truncated octahedral boxes containing TIP3P water²¹ were used to solvate the solute, with the distance between any solute atom and the nearest wall of the box is ≥ 10 Å. Sodium and chloride ions are added to the solvent box to neutralize the net charges on the solute and to mimic the experimental ionic concentration of 0.1 M. The electrostatic interactions were treated using the particle-mesh Ewald (PME)²² method with a real space cutoff of 10 Å and a grid spacing of 1.0 Å. MD simulations were performed in the NPT ensemble with a time step of 2 fs.

The potentials of mean force were computed using umbrella sampling by applying harmonic biasing potential on the distance between the center of mass (COM) of the CTD (denoted as L1) and center of mass of the following CCD residues (denoted here as P1): 124-128, 131-132, 167-168 and BI224436 (Fig. 5). During the umbrella sampling simulation, weak angular harmonic restraints are also applied to maintain the relative orientation between the two proteins and to ensure that the pulling is done along the predefined axis \mathbf{r}_{P1-L1} , and to accelerate the convergence of the PMF calculation. For this purpose, two polar angles (θ , ϕ) and three Euler angles (Θ , Φ , Ψ)²³⁻²⁴ are defined using the two centers of masses P1 and L1, plus two atoms on each of the binding proteins. The following two CCD atoms (P2, P3) and two CTD atoms (L2, L3) are chosen together with P1 and L1 to define five angles (θ , ϕ , Θ , Φ , Ψ) that determine the ligand orientation: P2: T124-CA; P3: W131-C; L2: D253-CA; L3: V234-C. The range of the pulling distance is $11.0 \text{ \AA} \leq |\mathbf{r}_{P1-L1}| \leq 23.9 \text{ \AA}$. The equilibrium values of the five angles are: $\theta_0 = 58.29^\circ$, $\phi_0 = -70.92^\circ$, $\Theta_0 = 130.1^\circ$, $\Phi_0 = 77.27^\circ$, $\Psi_0 = -57.45^\circ$. A single force constant $k_r = 23.9$ kcal

$\text{mol}^{-1} \text{ \AA}^{-2}$ is used for the distance restraints in all the umbrella sampling windows. The force constants used in the angular restraints are: $k_{\theta} = k_{\phi} = k_{\Theta} = k_{\Phi} = k_{\Psi} = 0.0728 \text{ kcal degree}^{-2} \text{ mol}^{-1}$.

For umbrella sampling used to compute the 1D PMF, 24 umbrella windows are used to cover the full range of the distance space. The spacing between adjacent umbrella windows is 0.5 - 1.0 \AA , which we find provide good overlap between the sampled distances. In each umbrella window, a reasonable starting configuration is generated by first running a short 1.2 ns umbrella sampling simulation starting from the last snapshot of the short umbrella sampling simulation in the previous umbrella window. Next, 30 ns umbrella sampling simulations are run in parallel in each window starting from the initial configurations generated above. The last 20 ns trajectory was used to compute the PMF using the WHAM²⁵ program implemented in the GORMACS 2016.3 package. For each protein-protein complex, four independent umbrella sampling simulations were run for estimating the statistical uncertainties in the calculated PMF. Thus, a total of $\sim 2.88 \mu\text{s}$ simulation time is used to compute the PMF for a single protein-protein complex. The binding entropy is estimated using the temperature difference of the free energy changes at 275 K, 300 K and 320 K. The error bar in the ΔS is estimated using the formula for the standard error of the slope of linear regression, i.e. $s(b_1) = \sqrt{\frac{1}{n-2} \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (x_i - \bar{x})^2}}$. The error bar in the ΔH is estimated using error propagation from the statistical error in the $T\Delta S$ and ΔG .

Results

Direct simulation of the formation of aberrant tetramer from dimers

In a previous report, we have used protein-protein docking guided by indirect experimental information from hydrogen-deuterium exchange (HDX) data as restraints to derive the first structural model of the drug-induced HIV-1 integrase multimer⁵. The model accurately predicted that the intermolecular CCD-CTD interaction mediated by bound ALLINIs is the most important feature of the drug-induced IN

multimerization. While the structural model predicted using protein-protein docking was confirmed by the subsequently reported crystal structure¹¹, the docking study was carried out without the aqueous environment and therefore precluded the characterization of the thermodynamics and kinetics of the drug-induced IN aggregation. To explore the association mechanism of the IN multimerization process, here we perform multiple independent MD simulations in explicit solvent without any restraints for a total of 23 μ s starting from two dissociated HIV-1 integrase dimers containing ALLINI BI224436²⁶⁻²⁷ with varying starting interfacial separations d (Fig. 1). The starting structure of two dimer is taken from the structure of the aberrant tetramer containing BI224436 reported previously⁴. The starting interface gap distance d is obtained by increasing the dimer-dimer separation from the structure of the aberrant tetramer complex by d . The setup and results of these simulations are summarized in Table 1. In simulation set-1, the initial separation distance between the surfaces of CTD and CCD of the two dimers is $d \approx 4.3$ Å, while in the simulation set-2, they are initially separated by between 7.8 Å and 11.8 Å. All the simulations are repeated with different initial velocities and each lasts on average 500 ns. The structures sampled in the MD trajectories are compared with the crystal structure of the aberrant IN tetramer reported by Gupta *et al*¹¹.

Table 1. Summary of the direct simulations of aberrant tetramer formation starting from dissociated IN dimers.

Simulation Set	Starting Configuration	Successful association/total trajectories	MFPT ^a (ns)	Average trajectory length (ns)
Set 1	$d \approx 4.3$ Å	8/10	~20	500
Set 2	$7.8 \text{ Å} \leq d \leq 11.8 \text{ Å}$	0/10	> 500	500

a. Mean First Passage Time to the formation of near-native complex.

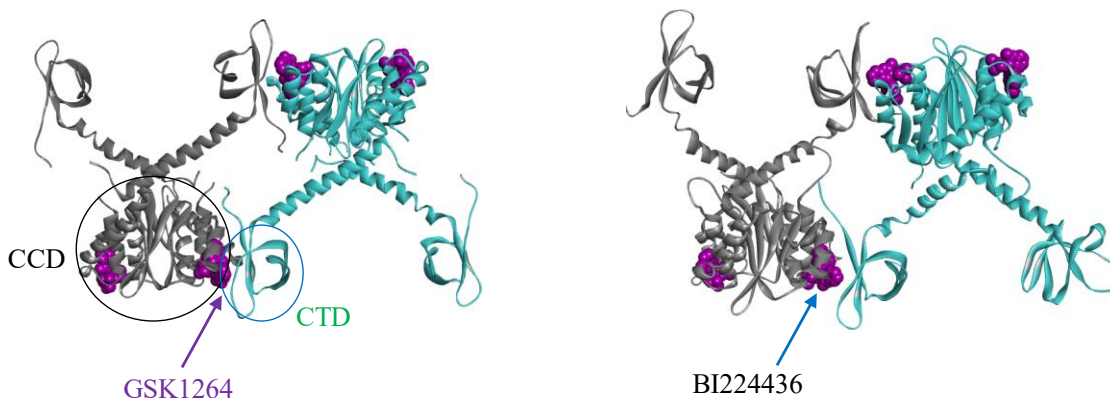


Figure 2. Left: Crystal structure of the IN aberrant tetramer containing the GSK compound GSK1264. Right: A representative MD snapshot taken at 28 ns (starting from two dissociated dimers, each containing the Gilead compound BI224436).

The experimental aberrant tetramer structure is captured in unbiased MD simulations in explicit solvent starting from separated integrase dimers

We found that in the majority of the trajectories in Set 1, the two initially separated dimers ($d \approx 4.3 \text{ \AA}$) spontaneously associate to form complexes that resemble the experimental structure of the native aberrant multimer: see Fig. 2, which compares a representative MD snapshot with the experimental structure. A snapshot is considered to have a near native multimer interface if the following criteria is met: when the central interface residues in the CCD (T124, A128, E170, H171 and T174) are superimposed onto the corresponding residues of the CCD in the crystal structure, the root mean square deviation (IRMSD) between the interface residues of the CTD (Y226, W235, K266 and I268) in the MD structure and those in the crystal structure is less than 3.5 \AA . Importantly, the outcome of the simulation is highly sensitive to the starting separation of the two dimers: in contrast to the trajectories in Set 1, none of the trajectories in the Set 2 leads to successful association and the formation of aberrant tetramer. Note that the only difference between the two sets of simulations is that the two dimers in Set 2 have larger initial separation compared with those in Set 1 (Table 1).

The probability for observing the native aberrant tetramer within a timescale of hundreds nanoseconds is highly sensitive to initial dimer-dimer separation

These results therefore show that there is a very sharp increase in the mean first passage time (MFPT) to association (Table 1) as the initial separation between the CTD and CCD on the two dimers is increased from about 4.3 Å to 7.8 Å: when the two dimers are separated by a gap of ≤ 4.3 Å, which is less than two water diameters, the association is fast, with MFPT ~ 20 ns for reaching the near native complex (Fig. 3A). Beyond the critical separation gap, the association kinetics is characterized by a much longer association time scale (Fig. 3B) beyond the microsecond range explored in this work.

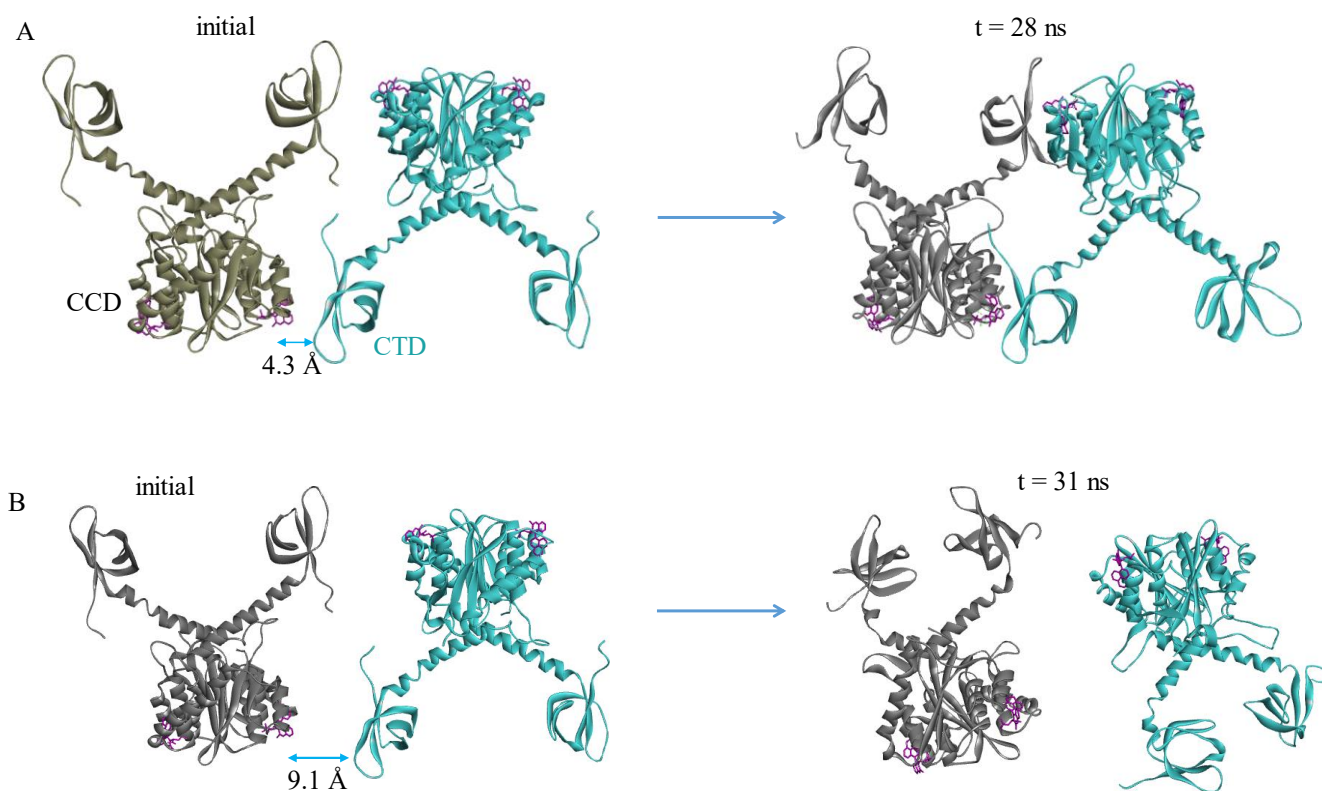


Figure 3. (A) Representative results of simulations Set 1 where the CTD and CCD on the two dimers are initially separated by interfacial gap $d \geq 4.3$ Å. (B) A trajectory from simulations Set 2 where the closest distance between CTD and CCD in the starting structure is 9.1 Å.

The calculated potential of mean force (PMF) reveals that the association free energy landscape is a narrow funnel with the downhill association beginning at a critical separation of about 4.1 Å.

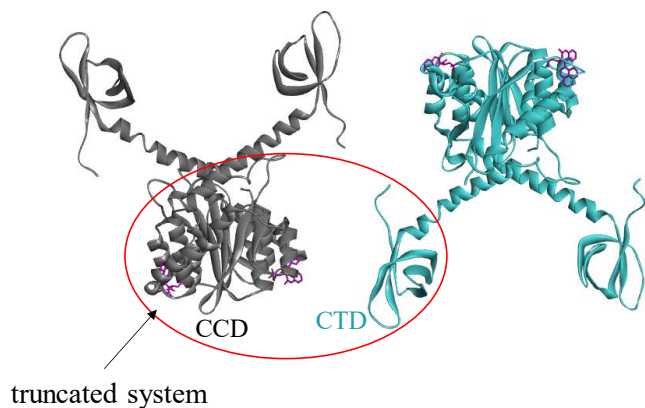


Figure 4. The truncated system (red oval) used to compute the PMF of CTD-CCD association using umbrella sampling.

To quantitatively characterize the energy landscape, the strength and the molecular mechanism of the drug-induced IN aberrant association, we compute the association free energy profile or potential of mean force (PMF)^{13, 28-31} as a function of the separation distance of the

two binding partners. The solvated system of the two dimers has ~180,000 atoms, which makes the

PMF calculation computationally very costly. From Fig. 2, it is noted that in the complexed state, two dimers make two symmetrical contact interfaces that each involves one CTD and one CCD dimer, bridged by one drug molecule. Since the two symmetrical contact interfaces are equivalent, a truncated system of one CCD dimer and one CTD (Fig. 5) provides a reasonable model system for investigating the free energy landscape of the association of the full dimer-dimer system. This truncated system (~70,000 atoms) used in the umbrella sampling calculation of the PMF comprises one CCD dimer with a bound ALLINI and one CTD. We show below that the validity of the truncated computational model is supported by the

comparison of the onset of the downhill association from the truncated system with that of the full dimer-dimer system.

Fig. 5 shows the PMF of the CTD association with an ALLINI-bound CCD dimer, computed from umbrella sampling by applying harmonic bias on the COM distance, i.e. the distance between the center of mass (COM) of the CTD and that of the following interfacial residues on the CCD dimer: 124-128, 131-132, 167-168 and BI224436. The interface separation gap d between the two binding surfaces is also shown in the left panel of Fig. 5. Here the gap d is calculated by subtracting the COM distance corresponding to the bound complex $d_{bound} = 11.4 \text{ \AA}$ from the current COM distance. During umbrella sampling the reversible pulling is done along the axis defined by the two COMs defined above, which is

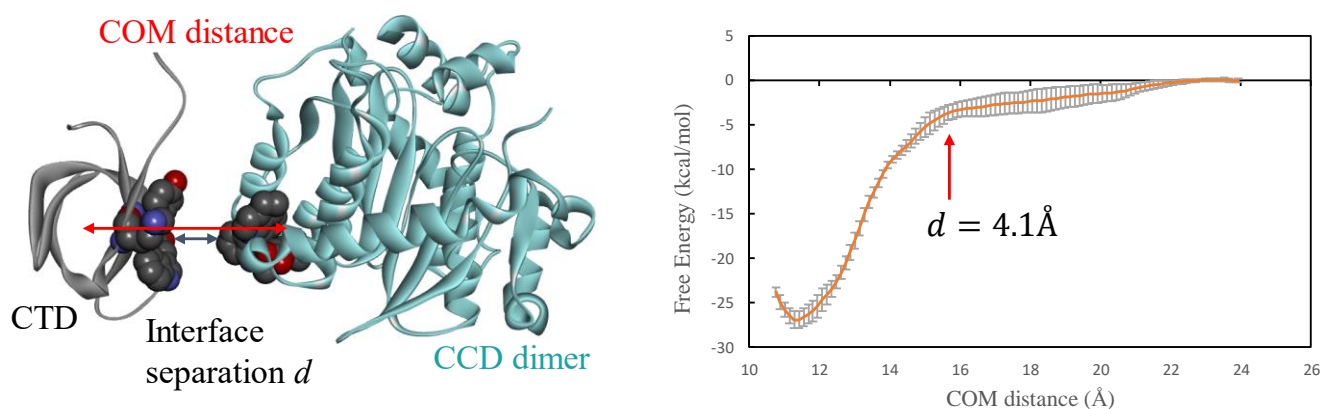


Figure 5. Left: The distance between the center of mass of the CTD and that of the interface residues on the CCD. Also shown is the interface separation distance d between the two binding surfaces. The residues at the centers of the two binding interfaces (Tyr226 and Trp235 on CTD, and BI224436 that binds with CCD) are shown as spheres to illustrate the interface separation d . Right: The PMF of the CTD and BI224436-bound CCD computed using umbrella sampling on the COM distance.

perpendicular to the binding interface. This ensures that no intermolecular entanglement or collision between the two binding partners occurs along this pulling pathway. To accelerate the convergence in computing the PMF of protein-protein association, weak angular harmonic restraints on the center of the CTD and its relative orientation with respect to the CCD are applied to the polar angles (θ , ϕ) and three Euler angles (Θ , Φ , Ψ)²³⁻²⁴ during umbrella sampling: see Methods.

The PMF is a monotonically decreasing function of the separation. Starting at a critical separation distance of about 4.1 Å the landscape of the protein-protein association becomes a steep downhill funnel. As shown in Fig. 5, the free energy of bringing the two proteins from far apart to $d = 4.1$ Å, i.e. $\Delta G(d: \infty \rightarrow 4.1 \text{ \AA}) = -3.6 \pm 0.86 \text{ kcal/mol}$, whereas going from $d = 4.1$ Å to $d = 0$ Å (corresponds to the bound complex), the free energy drop is much larger, $\Delta G(d: 4.1 \text{ \AA} \rightarrow 0 \text{ \AA}) = -23.3 \pm 0.43 \text{ kcal/mol}$. This result coincides with the observations from the unrestrained molecular dynamics simulations of the full dimer-dimer system, where the spontaneous association is found to occur at an initial separation distance of $d = 4.3$ Å (Table 1 and Fig. 3). The result therefore confirms that the effective attractive interaction involved in the multimerization is very short-ranged, i.e. it becomes attractive at a distance that can accommodate less than two layers of water molecules. Beyond this critical short separation, the association free energy landscape is almost flat and the corresponding association kinetics is largely diffusive. The short-ranged nature of the effective attraction is likely to have contributed to the slow kinetics for multimerization; for example, as shown by the study from Kvaratskhelia group the timescale for the growth of the ALLINI-induced aggregates is on the order of minutes⁷.

Justification for using PMF alone as a measure of the strength of association

While the strength of molecular association is normally measured by the absolute binding free energy of the two binding partners, for reasons given below, we only need to consider the PMF as a measure for the strength of the protein-protein association. The absolute binding free energy can be written as²³⁻²⁴

$$\begin{aligned}
 G_{bind}^{\circ} &= -\Delta G_{(\theta, \phi, \Theta, \Phi, \Psi)}^{bound} - w(r^*) - k_B T \ln \frac{\int_{bound} e^{-\frac{w(r)}{k_B T}} dr}{\left(\frac{2\pi k_B T}{k_r}\right)^{\frac{1}{2}}} - k_B T \ln \frac{C^{\circ} r^{*2} \sin\theta_0 \sin\Theta_0 (2\pi k_B T)^3}{8\pi^2 (k_r k_{\theta} k_{\phi} k_{\Theta} k_{\Phi} k_{\Psi})^{\frac{1}{2}}} \\
 &= -\Delta G_{(\theta, \phi, \Theta, \Phi, \Psi)}^{bound} - w(r^*) - \Delta G_{bound-bulk}^{vibr} + \Delta G_{restr}^{bulk} \quad (1)
 \end{aligned}$$

where ΔG_{restr}^{bulk} and $-\Delta G_{(\theta, \phi, \Theta, \Phi, \Psi)}^{bound}$ are the free energy of switching on the polar and orientational restraints $(U_{\theta}, U_{\phi}, U_{\Theta}, U_{\Phi}, U_{\Psi})$ for the ligand (in this case, CTD) in the bulk solution and that from and

switching them off in the bound state, respectively. $w(r^*)$ is the reversible work (PMF) of pulling the ligand from the bound state to the bulk location r^* , computed using umbrella sampling in the presence of the angular restraints $(U_\theta, U_\phi, U_\Theta, U_\Phi, U_\Psi)$. $-\Delta G_{bound-bulk}^{vibr}$ accounts for the ligand positional fluctuation in the bound state versus when it is harmonically restrained at a bulk location with the force constant k_r .²³ $k_\theta, k_\phi, k_\Theta, k_\Phi,$ and k_Ψ are the force constants for the two polar angles (θ, ϕ) and three Euler angles (Θ, Φ, Ψ) respectively, for the ligand in the receptor reference frame²³⁻²⁴. Table 2 lists the different components of the absolute binding free energy for the CTD association with an ALLINI-bound CCD dimer. The calculated $\Delta G_{bind}^\circ = 18.68 \pm 0.96$ kcal/mol has the same order of magnitude as the experimental affinity for tight protein-protein complexes such as the barnase-barstar ($\Delta G_{bind}^\circ = -19.02$ kcal/mol)³².

Table 2. Components of the absolute binding free energy for the CTD association with an ALLINI-bound CCD dimer. (kcal/mol)

$-\Delta G_{(\theta,\phi,\Theta,\Phi,\Psi)}^{bound}$	$-w(r^*)$	$-\Delta G_{bound-bulk}^{vibr}$	ΔG_{restr}^{bulk}	ΔG_{bind}°
-1.73 ± 0.11	-26.9 ± 0.96	-0.29 ± 0.03	10.24	-18.68 ± 0.96

As can be seen from Table 2, among the different contributions to the absolute binding free energy G_{bind}° , $-\Delta G_{(\theta,\phi,\Theta,\Phi,\Psi)}^{bound}$ and $-\Delta G_{bound-bulk}^{vibr}$ are small compared with the PMF term $-w(r^*)$. In addition, ΔG_{restr}^{bulk} is a constant term for the various combinations of protein-protein association calculations considered in this work. Therefore, in this work we only need to consider the PMF contribution $-w(r^*)$ as a measure of the absolute association free energy G_{bind}° .

Entropy-enthalpy decomposition of the PMF reveals a two-step association process with distinct thermodynamic signatures

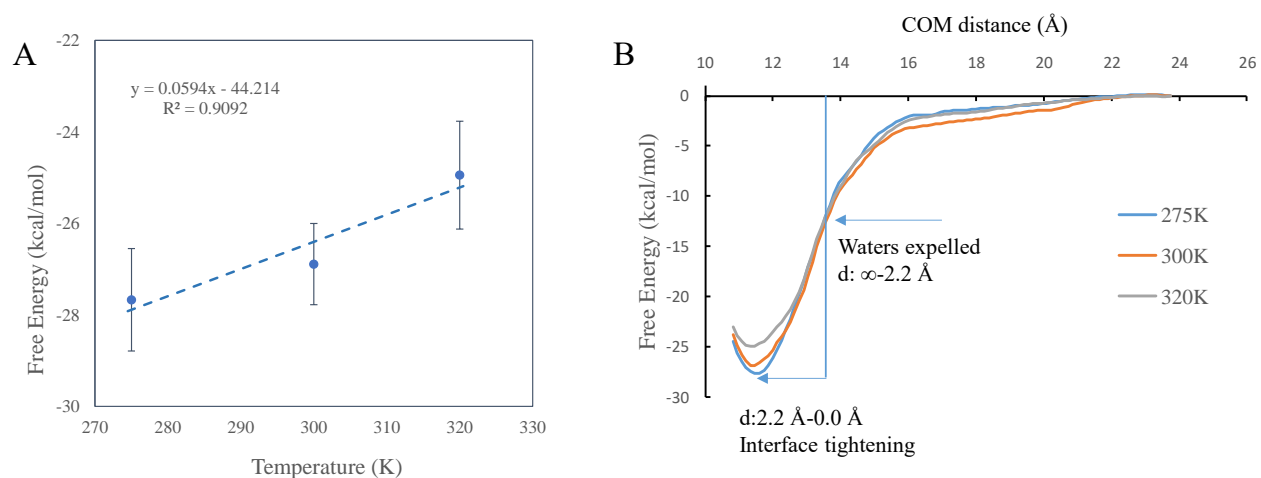


Figure 6A. Temperature dependence of the free energy of association. B. The range of the PMF can be divided into two parts: the water expulsion phase and the interface tightening phase.

To elucidate the thermodynamic driving forces for the association, we decomposed the free energy of association into enthalpic and entropic contributions. Here the free energy of association is taken to be the value at the contact minimum in the PMF near $d = 0 \text{ \AA}$ that corresponds to the bound complex relative to the separated state. The entropy of association is estimated from the temperature derivative of the free energy, i.e. $\Delta S = -\left(\frac{\partial \Delta G}{\partial T}\right)_P$, where the ΔG is obtained from the PMF at three temperatures 275 K, 300 K and 320 K. As seen from Fig. 6A, the temperature dependence of the free energy of association is approximately linear.

Table 3 gives the thermodynamic quantities of the protein-protein association at different separations between the two binding partners. The column “d: ∞ to 0 \AA ” shows the thermodynamic quantities for bringing the CTD and the ALLINI-bound CCD from infinitely apart to the contact minimum. This column therefore gives the total association free energy, entropy and enthalpy. As seen from Table 3, overall, the association $\Delta G(\infty \rightarrow 0 \text{ \AA}) = -26.9 \pm 0.96 \text{ kcal/mol}$ is driven by a large favorable enthalpic contribution $\Delta H(\infty \rightarrow 0 \text{ \AA}) = -44.7 \pm 5.54 \text{ kcal/mol}$, and opposed by an unfavorable entropy change $-T\Delta S(\infty \rightarrow 0 \text{ \AA}) = 17.8 \pm 5.46 \text{ kcal/mol}$. Note that the estimation of the enthalpy and

entropy components has much larger statistical uncertainty than that in the free energy, which is expected since the entropy and enthalpy are related to the derivative of free energy³³⁻³⁴.

Given the fact that protein-protein association is accompanied by releasing water molecules to the bulk solution from the intervening space between the two proteins, the large unfavorable entropy of association is puzzling. To examine the physical origins of the entropic and enthalpic changes, we divide the entire range of the PMF into two segments and compute the corresponding thermodynamic quantities: see Fig. 6B. The first phase consists in bringing the dissociated proteins to a separation of $d = 2.2 \text{ \AA}$ where all the water molecules from the intervening space are expelled to the bulk. As shown in Table 3, during this phase the corresponding entropy change is weakly favorable, with $-T\Delta S(\infty \rightarrow 2.2 \text{ \AA}) = -1.29 \pm 6.08 \text{ kcal/mol}$. This small favorable entropy results from releasing the interfacial water molecules to the bulk, and its small magnitude is attributable to fact that the increase in solvent entropy is largely offset by the loss of side chain entropies as the two binding surfaces approach each other to establish intermolecular contact during this phase. For example, during this phase of the association the Lys266 side chain starts to form a salt bridge with the carboxylate group of the BI224436 at $d \approx 3.8 \text{ \AA}$. Such intermolecular interactions are also reflected by the favorable $\Delta H(\infty \rightarrow 2.2 \text{ \AA}) = -10.26 \pm 6.10 \text{ kcal/mol}$, see Table 3.

The second phase in the association process is from $d = 2.2 \text{ \AA}$ until the contact minimum at $d = 0 \text{ \AA}$. This phase is characterized by a small change in the separation distance of just 2.2 \AA while the two binding surfaces adapt to each other to optimize their interaction energy, which results in a large reduction in enthalpy $\Delta H(2.2 \text{ \AA} \rightarrow 0 \text{ \AA}) = -34.45 \pm 6.09 \text{ kcal/mol}$, which is gained at the expense of a large entropy loss of $-T\Delta S(2.2 \text{ \AA} \rightarrow 0 \text{ \AA}) = 19.11 \pm 6.03 \text{ kcal/mol}$. Therefore, the large entropy penalty originates from the loss of conformational entropy of the binding surface residues, since all the intervening waters have already been expelled to the bulk in the previous phase $d: \infty \rightarrow 2.2 \text{ \AA}$. Thus, these results

indicate that the large unfavorable entropy change for the overall association is caused by configurational entropy loss partially offset by entropic gains due to the expulsion of interfacial water to the bulk.

Table 3. Thermodynamic properties of the CTD-CCD association at different separations (300 K)

Thermodynamic quantity	d: ∞ to 0 Å	d: ∞ to 2.2 Å	d: 2.2 Å to 0 Å
$-T\Delta S$	17.8 ± 5.46	-1.29 ± 6.08	19.11 ± 6.03
ΔH	-44.7 ± 5.54	-10.26 ± 6.10	-34.45 ± 6.09
ΔG	-26.9 ± 0.96	-11.55 ± 0.48	-15.34 ± 0.85

The calculated PMF captures the crucial role of ALLINI in promoting IN multimerization

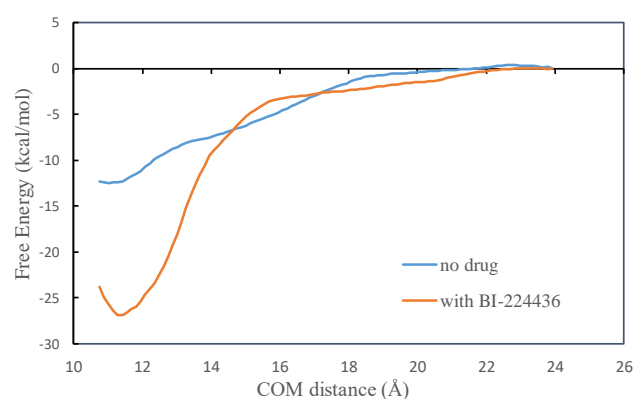


Figure 7. The computed PMF of CCD-CTD association with and without the drug BI224436.

It is known that ALLINI is required for the aberrant IN multimerization^{6, 35}. While the low-resolution crystal structure of the IN aberrant tetramer containing an ALLINI provides the structural basis for the multimer¹¹, it does not by itself contain information on the strengths of the association with or without ALLINI.

Computationally, we have previously used a FEP thermodynamic cycle to compute the relative association free energy to rationalize how different ALLINIs exhibit different potencies in stabilizing the multimer⁵. Here, we use PMF calculation to directly quantify the role of ALLINI in inducing the IN aggregation. Fig. 7 shows the computed PMF of CCD-CTD association with and without the ALLINI BI224436. The PMF of the association without the drug shows a shallow attractive basin that resembles that of a nonspecific protein-protein association. In contrast, the PMF computed with the bound BI224436 features a deep minimum characteristic of specific tight binding.

The free energies of association estimated from the PMF in Fig. 7 with and without the drug are -26.9 ± 0.96 kcal/mol and -12.4 ± 0.7 kcal/mol, respectively. The large difference in the association free energy can be understood from the structures of the complexes at their free energy minima in the PMF: see Fig. 8. The CTD-CCD binding interface in the complex containing the drug shows a significantly greater intermolecular atomic packing than the one without the drug. In addition, in the complex without the drug the CTD-CCD interfacial space is filled with several water molecules. These water molecules,

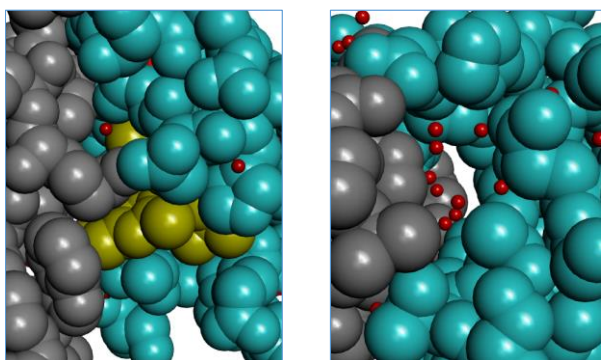


Figure 8. Representative MD snapshots of the complex of CTD (gray) and CCD (cyan) in the presence (left) and absence (right) of BI224436 (yellow). The snapshots are taken from the lowest free energy bins in the PMF (Fig. 7). Water molecules are shown as small red dots.

which are trapped in a narrow space between nonpolar interfacial residues (e.g. W235, Y226, L102, A128 and W132), form on average significantly fewer number of hydrogen bonds than the number of hydrogen bonds (≈ 3.5) per molecule formed in bulk water. Therefore, these interfacial water molecules in the complex without the drug likely weakened the association free energy. These simulation results provide a physical explanation for the

crucial role of ALLINI binding in promoting the IN aggregation.

The calculated PMF also explains why certain amino acid substitutions exhibited remarkable resistance to the drug-induced IN multimerization

Experimentally a number of engineered amino acid substitutions on the CTD, such as Y226D¹¹, W235A¹¹ and K266A⁷, are found to abrogate the ALLINI-induced aberrant IN multimerization. While the low-resolution crystal structure of the IN aberrant tetramer shows that these residues are at the protein-

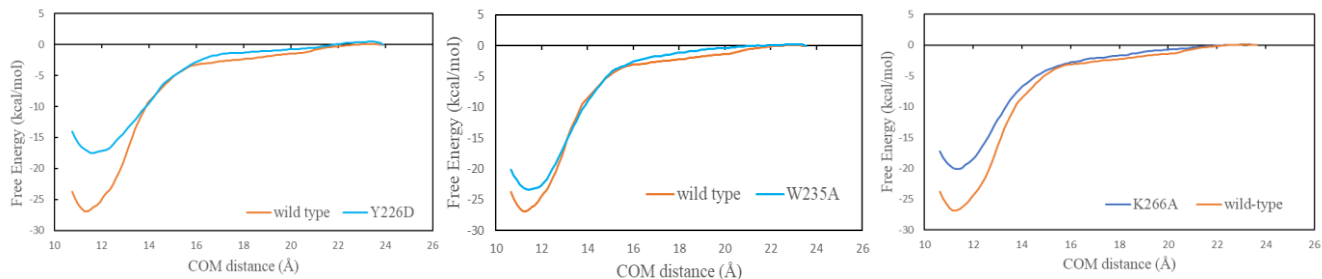


Figure 9. Free energy profile of the CTD-CCD binding in which the CTD contains the following single mutations: Y226D (left), W235A (center) and K266A (right).

protein interface, the experimental structure itself does not give information on the strength of the multimerization from the mutant proteins. Here we run PMF calculations on these mutant proteins to directly probe their effects on the association free energy. Fig. 9 shows the calculated PMF for the three mutant proteins. The calculated association free energies for the three mutant proteins, Y226D, W235A and K266A are -17.5 ± 1.0 kcal/mol, -23.4 ± 2.4 kcal/mol and -20.2 ± 2.7 kcal/mol, respectively. Thus, all three mutants bind significantly more weakly to the CCD than does the wild type CTD ($\Delta G_{assoc} = -26.9 \pm 0.96$ kcal/mol), in good agreement with the results of experimental multimerization assays^{7, 11}. These results provide further support for the validity of the free energy model adopted in this work.

Discussion

We have applied direct MD simulations and free energy calculations to explore the energy landscape of the protein-protein interaction responsible for the drug-induced aberrant multimerization of HIV-1 integrase. The PMF results captured the critical role of the ALLINI in promoting the CTD-CCD intersubunit interactions that are the cornerstone of the IN aggregation (Fig. 7 and Fig. 8). By showing that all three key substitutions Y226D, W235A and K266A results in significantly less attractive association than the wild type protein (Fig. 9), the calculations also provide physical explanations for why these amino acid substitutions block the drug-induced IN multimerization. These results demonstrate the power of free energy modeling in elucidating the effects of mutations in protein-protein interactions. However, potential of mean force calculations in explicit solvent as those performed here remain

computationally very demanding. In this regard we note a recent study by Siebenmorgen & Zacharias who have used a PMF-based binding free energy method in an implicit solvent to more accurately re-rank the docking poses of 20 different protein-protein complexes¹⁵.

In addition to providing a physical basis for the IN multimerization, the work sheds light on the

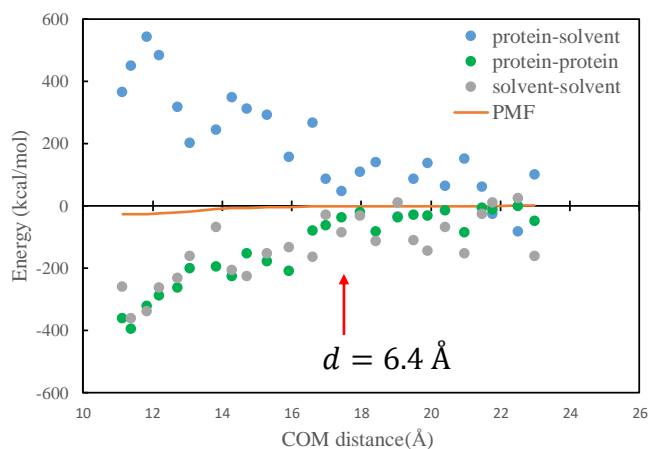


Figure 10. The PMF or free energy (orange) plotted together with the various interaction energies: protein-protein (green), solute-solvent (blue) and solvent-solvent (gray).

thermodynamic foundation of protein-protein association. We find that the protein-protein association between the CTD and CCD-ALLINI is driven by enthalpy change, and opposed by entropy: see Table 3 and Fig. 6. (Note that in the following we focus on the solvent and solute conformational entropy changes only, which are the values given in Table 3; the external entropy change that includes translational and rotational entropies will not be discussed.) To examine the origin of the favorable entropy we compare the free energy profile for the CTD and CCD-ALLINI association with the direct interaction energies that make up the total enthalpy of dimerization: the protein-protein, protein-solvent, and solvent-solvent interaction (Fig. 10). Although all the interaction energies are much noisier compared with the well-behaved free energy profile, it can be seen that as the two proteins approach each other, the association is favored by the protein-protein interaction energy. Conversely, the association is strongly opposed by the loss of protein-solvent interactions that are replaced by protein-protein interactions. Solvent-solvent interactions favor dimerization because water molecules expelled from the inter-protein space form more hydrogen bonds in the bulk. Note that the solvent-solvent interaction and the protein-protein direct interaction have similar orders of magnitudes.

We now discuss these results in the context of hydrophobic and hydrophilic protein-protein association, which can have different thermodynamic signatures. It has been known that the thermodynamic driving

force for hydrophobic association is length scale dependent³⁶⁻³⁷. Aggregation of small hydrophobic solutes is driven by favorable entropy change and opposed by unfavorable enthalpy change. This is because the water molecules near small hydrophobic solute can rearrange themselves to maintain the hydrogen bond network and, therefore, releasing such surface waters to the bulk upon solute association is more or less enthalpy neutral. In this case, the net enthalpy change of association results mainly from the loss of favorable solute-water interactions, which can outweigh the favorable solute-solute interaction. For example, Cui et al.³⁸ studied the thermodynamics of the rigid body binding between a hydrophobic α -helix with ubiquitin. The size and shape of the buried surface area in this complex suggest that it belongs to the small solute regime. Hence the binding is driven by entropy and opposed by enthalpy, due to the loss of protein-water interactions not compensated by the gain of protein-peptide interactions and solvent-solvent interactions from the desolvation of the solutes³⁸.

For large hydrophobic solutes (e.g. interface dimension $R > 10 \text{ \AA}$) association can be favored or disfavored by enthalpy, depending on the strength of the solute-water interaction. This is because unlike the water molecules around a small solute, water molecules near a large hydrophobic surface cannot orient themselves around the surface to maintain the hydrogen bond network. As a result, the loss of large hydrophobic surface patches upon binding releases energetically frustrated water molecules into the bulk resulting in a reduction of the enthalpy. For example, using highly hydrophobic plates with weak solute-water interaction parameters, Zangi and Berne³⁹ found that both entropy and enthalpy favor hydrophobic collapse. However, for graphene plates that have stronger plate-water interaction parameters, Choudhury and Pettitt⁴⁰ found that the collapse of the plates in water is driven only by entropy and opposed by enthalpy. This is because the loss of plates-water interaction upon collapse outweighs the enthalpy gain from the release of water molecules in between the plates and from the increased direct plate-plate attraction.

There have been few computational studies on characterizing the thermodynamic driving forces for hydrophilic protein-protein association⁴¹⁻⁴². Helms and coworkers have studied several hydrophilic protein-protein complexes including the barnase-barstar complex⁴¹, focusing on the calculation of the binding free energy and translation/rotation entropy changes. For the barnase-barstar complex, using the experimental data⁴³ one can estimate that the protein-protein binding is largely driven by favorable enthalpy with a relatively small favorable entropy contribution. Ben-Naim proposed a model suggesting that in hydrophilic protein-protein association, water molecules can “stitch” the two proteins by forming hydrogen-bond bridges between two hydrophilic groups on the surfaces of two binding partners, resulting in favorable enthalpy at the expense of entropy⁴⁴.

In the present study, as shown in Fig. 11 the protein-protein interfaces in the complex contain a mixture of mainly hydrophobic groups and also several hydrophilic residues, with the nonpolar patch roughly

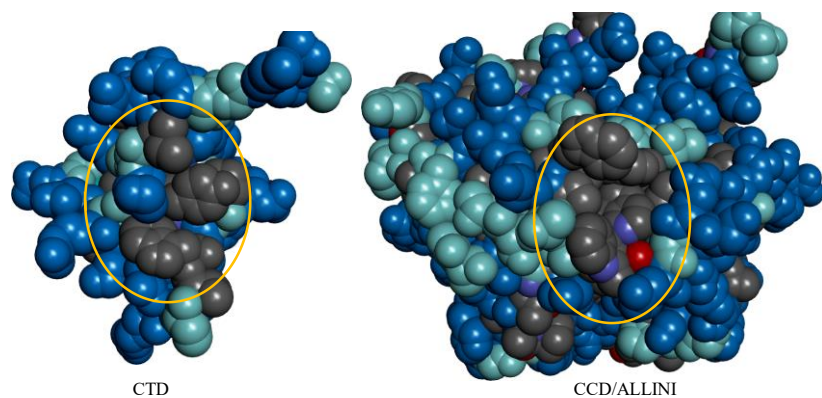


Figure 11. The VDW surfaces of the binding interfaces in the two proteins. Gray: hydrophobic. Blue: hydrophobic. Cyan: intermediate between hydrophilic and hydrophobic. The central contact residues involved in the binding are circled.

surrounded by polar residues. Since both the solvent and solute degrees of freedom are unconstrained, our calculations reveal a more realistic thermodynamic picture of protein-protein association than did previous studies which either used idealized geometry (plates and ellipsoids) or have fixed intramolecular degrees of

freedom of the protein. Importantly, we found that when the protein conformational degrees of freedom are accounted for, the entropy of association becomes unfavorable (Table 3). Consistent with previous work, we also found that during the water expulsion phase, the entropy change is weakly favorable, as the large increase in the solvent entropy due to inter-protein water release is largely cancelled by the reduction in the solute entropy. As the two proteins approach the contact minimum, the entropy drops further as the

conformational motions are more limited. The configuration entropy loss at the final stage of the complexation of ~ 20 kcal/mol is in the same order as those estimated by Chang et al. for the binding of AKAP protein HT31 with the D/D domain of RII α -regulatory subunit of protein kinase A⁴⁵.

In the current study, we find evidence that is consistent with the bridging water effect proposed by Ben-Naim: see Fig. 12. The thermodynamic properties for the association of large hydrophobic and hydrophilic solutes surveyed here are summarized in Table 4. It should be noted that while the signs of the entropy and enthalpy may be generalized for idealized systems, for realistic proteins the situation is more complex and likely to be system dependent because of the presence of many variables such as the interface composition, curvature, flexibility, and solvent conditions (e.g. ionic concentration and pH).

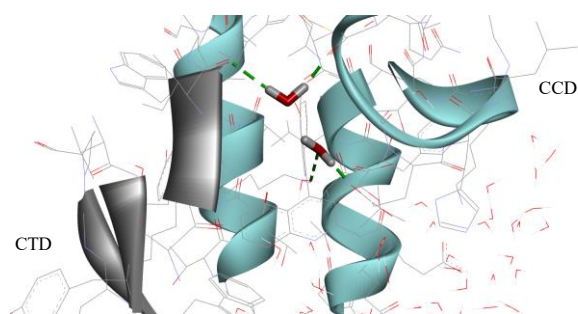


Figure 12. A MD snapshot showing two bridging waters forming hydrogen bonds with both proteins.

Table 4 Summary of the thermodynamic driving forces in the association of large solutes surveyed here

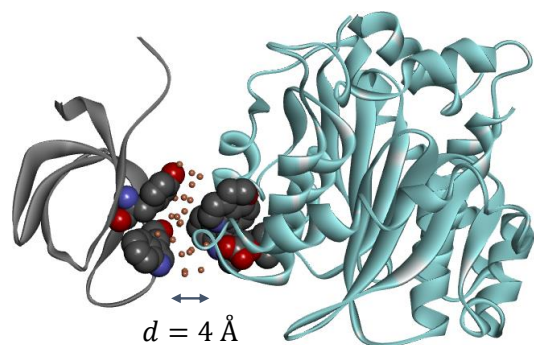
Super plates ³⁹	hydrophobic	Normal plates ⁴⁰	hydrophobic	Hydrophilic proteins ⁴³	Proteins that associate using both nonpolar and polar groups (this work)
ΔS favorable		ΔS favorable		ΔS small	ΔS unfavorable
ΔH favorable		ΔH unfavorable		ΔH favorable	ΔH favorable

In this work, both the direct simulations and PMF calculations show that the effective range of the attraction between the two binding proteins CTD and ALLINI-bound CCD is very short, at interface distance $d \approx 4$ Å or less than two layers of water molecules, even though the two binding partners CTD

and CCD carry 6+e and 3-e net charges, respectively. Apparently, the short-ranged effective interaction is the combined result of several factors, including the spatial distribution of the charges and the electrostatic screening from the solvent. To gain more insights into the origin of this short-ranged effective interaction we analyze the calculated protein-protein PMF in terms of the interaction energy components in Fig. 10. Although all the interaction energies are much noisier compared with the well-behaved free energy profile, some qualitative trends can be gleaned from such an analysis. First, the short-range protein-protein PMF is the result of near complete cancellation of several much larger forces (note that entropy contributions not included in Fig. 10). Second, the interaction energies have somewhat longer range compared with the free energy. For example, the protein-solvent interaction energy starts to increase at about $d = 6.4 \text{ \AA}$, while the free energy starts its descent at $d = 4.1 \text{ \AA}$.

Next, we compare the effective range of the protein-protein binding observed in this work with those reported in the literature^{13, 41-42, 46}. Helms and coworkers have computed the PMF for several hydrophilic protein-protein complexes including the barnase-barstar complex⁴¹ where they report a large effective range of the binding funnel of $\approx 14 \text{ \AA}$. The PMF of the barnase-barstar complex, which is characterized by highly polar/charged interface in explicit solvent, has also been reported by other groups. For example, Hoefling and Gottschalk⁴⁶ reported the width of the binding funnel to be $\sim 15 \text{ \AA}$, similar to the result by Helms and coworkers⁴¹. On the other hand, in a study by Gumbart et al¹³ in which the protein conformational degrees of freedom are constrained, the effective range of attraction in barnase-barstar is $\sim 6 \text{ \AA}$. Overall, the effective range of the association funnels for hydrophilic protein-protein binding is considerably larger than the $\sim 4 \text{ \AA}$ for the HIV-1 integrase aggregation studied here. The difference in the width of the binding funnel could be attributed to the more hydrophobic nature of the system studied in the present work, as the CTD and drug-bound CCD interface involve several nonpolar residues such as A128, T226, W235 and I268, in addition to the charged K266.

Lastly, some studies of hydrophobic protein-protein association^{37, 47-49} have focused on determining the effects of dewetting or cavitation⁵⁰, i.e. the cooperative liquid-to-vapor phase transition in the interdomain region which was first observed in idealized plates when the inter-plate region is still large enough to hold more than two layers of water.³⁷ While a dewetting transition may occur below a critical separation between idealized nanoscale hydrophobic plates or ellipsoids⁵¹⁻⁵², it has rarely been observed in realistic hydrophobic protein association⁵³. For example, in the hydrophobic collapse of protein



domains of 2,3-dihydroxybiphenyl dioxygenase (BphC), dewetting is observed only after turning off the protein-water electrostatic interactions in the simulation⁴⁸. The only system in which the cooperative drying phase transition has been observed is in the melittin tetramer⁴⁷, which was

Figure 13. The orange dots represent interfacial water sites whose occupancy is at least 1.5 times the bulk water occupancy.

separated to create a nanoscale tube-like channel. Interestingly, Ricci and McCammon⁵⁴ recently observed

that during the association of a 15-residue hydrophobic MDM2 peptide to a highly hydrophobic cleft of the p53 protein, the water density in the interfacial space starts to drop from the bulk value when the interfacial separation is less than 7.6 \AA , which is believed to be the onset of dewetting transition. However, it appears that cavitation in the inter-protein space did not occur even in this highly nonpolar system. In the present study, we did not observe a dewetting transition in the protein-protein association of the CTD and CCD-ALLINI system. As shown in Fig. 12, within the binding funnel, the intervening region remains well hydrated with water molecules whose occupancy is even higher than that in the bulk water. The small number of hydrophilic groups in the interface region (Fig. 11) which interact strongly with interfacial water molecules is likely cancelling any tendency to dewet⁵³.

Conclusion

Elucidating the energetics and thermodynamic aspects of protein-protein association is important to our understanding of this fundamental molecular recognition process. Computational characterizations of the thermodynamic signatures of protein-protein association for realistic systems are still scarce. Here using explicit solvent MD and umbrella sampling, we have provided a more complete picture of the protein-protein association responsible for the drug-induced HIV-1 integrase multimerization. The calculated free energy profile has correctly captured the critical role of the inhibitor in the integrase multimerization, and accounted for the effects of various amino acid substitutions on the protein association. Remarkably, the results reveal that the effective range of the protein-protein attractive funnel is less than two layers of water molecules, which is significantly shorter than those reported for the more hydrophilic protein-protein complexes, despite the fact that the two binding partners studied here carry opposite net charges. Lastly, we find that while the majority of the residues forming the contact interface are hydrophobic, the overall association is favored by enthalpy and opposed by entropy. Our analysis suggests that while the solvent release from the inter-protein region into the bulk solution results in an increase in entropy, this is more than offset by the loss of the solute configurational entropy due to the complexation.

Author Contributions

N. Deng, R. Levy and E. Gallicchio designed the research. C. Tse, L. Wickstrom and N. Deng performed the research. M. Kvaratskhelia contributed research data. N. Deng, R. Levy, E. Gallicchio and L. Wickstrom analyzed the results. N. Deng, R. Levy and E. Gallicchio wrote the paper.

Acknowledgments

The calculations were run on the XSEDE allocation resource TG-MCB100145 and a shared computing cluster at Temple University supported by NIH S10 OD020095. N. Deng thanks Prof. Wei Yang at Florida State University and Mr. Eric Chen from Lehman College, CUNY for helpful discussions. This study was supported by NIH grant 5U54AI150472-09 and by a Bridge fund from Pace University to N. Deng.

References

1. Lapelosa, M.; Gallicchio, E.; Arnold, G. F.; Arnold, E.; Levy, R. M., In silico vaccine design based on molecular simulations of rhinovirus chimeras presenting HIV-1 gp41 epitopes. *J Mol Biol* **2009**, *385* (2), 675-91.
2. Feng, L.; Larue, R. C.; Slaughter, A.; Kessl, J. J.; Kvaratskhelia, M., HIV-1 Integrase Multimerization as a Therapeutic Target. In *The Future of HIV-1 Therapeutics*, Torbett, B. E.; Goodsell, D. S.; Richman, D. D., Eds. Springer International Publishing: 2015; Vol. 389, pp 93-119.
3. Jin, L.; Wang, W.; Fang, G., Targeting Protein-Protein Interaction by Small Molecules. *Annual Review of Pharmacology and Toxicology* **2014**, *54* (1), 435-456.
4. Koneru, P. C.; Francis, A. C.; Deng, N.; Rebensburg, S. V.; Hoyte, A. C.; Lindenberger, J.; Adu-Ampratwum, D.; Larue, R. C.; Wempe, M. F.; Engelman, A. N.; Lyumkis, D.; Fuchs, J. R.; Levy, R. M.; Melikyan, G. B.; Kvaratskhelia, M., HIV-1 integrase tetramers are the antiviral target of pyridine-based allosteric integrase inhibitors. *Elife* **2019**, *8*.
5. Deng, N.; Hoyte, A.; Mansour, Y. E.; Mohamed, M. S.; Fuchs, J. R.; Engelman, A. N.; Kvaratskhelia, M.; Levy, R., Allosteric HIV-1 integrase inhibitors promote aberrant protein multimerization by directly mediating inter-subunit interactions: Structural and thermodynamic modeling studies: Mechanism of Allosteric HIV-1 Integrase Inhibitors. *Protein Science* **2016**, *25* (11), 1911-1917.
6. Sharma, A.; Slaughter, A.; Jena, N.; Feng, L.; Kessl, J. J.; Fadel, H. J.; Malani, N.; Male, F.; Wu, L.; Poeschla, E.; Bushman, F. D.; Fuchs, J. R.; Kvaratskhelia, M., A New Class of Multimerization Selective Inhibitors of HIV-1 Integrase. *PLOS Pathogens* **2014**, *10* (5), e1004171.
7. Shkriabai, N.; Dharmarajan, V.; Slaughter, A.; Kessl, J. J.; Larue, R. C.; Feng, L.; Fuchs, J. R.; Griffin, P. R.; Kvaratskhelia, M., A Critical Role of the C-terminal Segment for Allosteric Inhibitor-induced Aberrant Multimerization of HIV-1 Integrase. *Journal of Biological Chemistry* **2014**, *289* (38), 26430-26440.
8. Feng, L.; Sharma, A.; Slaughter, A.; Jena, N.; Koh, Y.; Shkriabai, N.; Larue, R. C.; Patel, P. A.; Mitsuya, H.; Kessl, J. J.; Engelman, A.; Fuchs, J. R.; Kvaratskhelia, M., The A128T Resistance Mutation Reveals Aberrant Protein Multimerization as the Primary Mechanism of Action of Allosteric HIV-1 Integrase Inhibitors. *Journal of Biological Chemistry* **2013**, *288* (22), 15813-15820.
9. Jurado, K. A.; Wang, H.; Slaughter, A.; Feng, L.; Kessl, J. J.; Koh, Y.; Wang, W.; Ballandras-Colas, A.; Patel, P. A.; Fuchs, J. R.; Kvaratskhelia, M.; Engelman, A., Allosteric integrase inhibitor potency is determined through the inhibition of HIV-1 particle maturation. *Proceedings of the National Academy of Sciences* **2013**, *110* (21), 8690-8695.
10. Engelman, A. N., Multifaceted HIV integrase functionalities and therapeutic strategies for their inhibition. *J Biol Chem* **2019**, *294* (41), 15137-15157.
11. Gupta, K.; Turkki, V.; Sherrill-Mix, S.; Hwang, Y.; Eilers, G.; Taylor, L.; McDanal, C.; Wang, P.; Temelkoff, D.; Nolte, R. T.; Velthuisen, E.; Jeffrey, J.; Van Duyne, G. D.; Bushman, F. D., Structural Basis for Inhibitor-Induced Aggregation of HIV Integrase. *PLOS Biology* **2016**, *14* (12), e1002584.

12. Ahmad, M.; Gu, W.; Geyer, T.; Helms, V., Adhesive water networks facilitate binding of protein interfaces. *Nat Commun* **2011**, *2*, 261.
13. Gumbart, J. C.; Roux, B.; Chipot, C., Efficient Determination of Protein–Protein Standard Binding Free Energies from First Principles. *Journal of Chemical Theory and Computation* **2013**, *9* (8), 3789-3798.
14. Pan, A. C.; Jacobson, D.; Yatsenko, K.; Sritharan, D.; Weinreich, T. M.; Shaw, D. E., Atomic-level characterization of protein-protein association. *Proc Natl Acad Sci US A* **2019**, *116* (10), 4244-4249.
15. Siebenmorgen, T.; Zacharias, M., Evaluation of Predicted Protein-Protein Complexes by Binding Free Energy Simulations. *J Chem Theory Comput* **2019**, *15* (3), 2071-2086.
16. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, *4* (3), 435-447.
17. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29* (7), 845-854.
18. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 1950-1958.
19. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004**, *25* (9), 1157-1174.
20. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *Journal of Computational Chemistry* **2000**, *21* (2), 132-146.
21. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79* (2), 926.
22. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995**, *103* (19), 8577.
23. Cruz, J.; Wickstrom, L.; Yang, D.; Gallicchio, E.; Deng, N., Combining Alchemical Transformation with a Physical Pathway to Accelerate Absolute Binding Free Energy Calculations of Charged Ligands to Enclosed Binding Sites. *J Chem Theory Comput* **2020**, *16* (4), 2803-2813.
24. Deng, N.; Cui, D.; Zhang, B. W.; Xia, J.; Cruz, J.; Levy, R., Comparing alchemical and physical pathway methods for computing the absolute binding free energy of charged ligands. *Phys Chem Chem Phys* **2018**, *20* (25), 17081-17092.
25. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **1992**, *13* (8), 1011-1021.
26. Fenwick, C.; Amad, M.; Bailey, M. D.; Bethell, R.; Bos, M.; Bonneau, P.; Cordingley, M.; Coulombe, R.; Duan, J.; Edwards, P.; Fader, L. D.; Faucher, A. M.; Garneau, M.; Jakalian, A.; Kawai, S.; Lamorte, L.; LaPlante, S.; Luo, L.; Mason, S.; Poupart, M. A.; Rioux, N.; Schroeder, P.; Simoneau, B.; Tremblay, S.; Tsantrizos, Y.; Witvrouw, M.; Yoakim, C., Preclinical Profile of BI 224436, a Novel HIV-1 Non-Catalytic-Site Integrase Inhibitor. *Antimicrob Agents and Chemotherapy* **2014**, *58* (6), 3233-3244.
27. Fader, L. D.; Malenfant, E.; Parisien, M.; Carson, R.; Bilodeau, F.; Landry, S.; Pesant, M.; Brochu, C.; Morin, S.; Chabot, C.; Halmos, T.; Bousquet, Y.; Bailey, M. D.; Kawai, S. H.; Coulombe, R.; LaPlante, S.; Jakalian, A.; Bhardwaj, P. K.; Wernic, D.; Schroeder, P.; Amad, M. a.; Edwards, P.; Garneau, M.; Duan, J.; Cordingley, M.; Bethell, R.; Mason, S. W.; Bös, M.; Bonneau, P.; Poupart, M.-A.; Faucher, A.-M.; Simoneau, B.; Fenwick, C.; Yoakim, C.; Tsantrizos, Y., Discovery of BI 224436, a Ncatalytic Site Integrase Inhibitor (NCINI) of HIV-1. *ACS Medicinal Chemistry Letters* **2014**, *5* (4), 422-427.

28. Woo, H. J.; Roux, B., Calculation of absolute protein-ligand binding free energy from computer simulations. *Proceedings of the National Academy of Sciences* **2005**, *102* (19), 6825-6830.
29. Doudou, S.; Burton, N. A.; Henchman, R. H., Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force. *J Chem Theory Comput* **2009**, *5* (4), 909-18.
30. Deng, N.; Wickstrom, L.; Cieplak, P.; Lin, C.; Yang, D., Resolving the Ligand-Binding Specificity in c-MYC G-Quadruplex DNA: Absolute Binding Free Energy Calculations and SPR Experiment. *J Phys Chem B* **2017**, *121* (46), 10484-10497.
31. You, W.; Tang, Z.; Chang, C. A., Potential Mean Force from Umbrella Sampling Simulations: What Can We Learn and What Is Missed? *J Chem Theory Comput* **2019**, *15* (4), 2433-2443.
32. Schreiber, G.; Fersht, A. R., Energetics of protein-protein interactions: Analysis of the Barnase-Barstar interface by single mutations and double mutant cycles. *Journal of Molecular Biology* **1995**, *248* (2), 478-486.
33. Levy, R. M.; Gallicchio, E., COMPUTER SIMULATIONS WITH EXPLICIT SOLVENT: Recent Progress in the Thermodynamic Decomposition of Free Energies and in Modeling Electrostatic Effects. **1998**, *49* (1), 531-567.
34. Wan, S.; Stote, R. H.; Karplus, M., Calculation of the aqueous solvation energy and entropy, as well as free energy, of simple polar solutes. *The Journal of Chemical Physics* **2004**, *121* (19), 9539.
35. Gupta, K.; Brady, T.; Dyer, B. M.; Malani, N.; Hwang, Y.; Male, F.; Nolte, R. T.; Wang, L.; Velthuisen, E.; Jeffrey, J.; Van Duyne, G. D.; Bushman, F. D., Allosteric Inhibition of Human Immunodeficiency Virus Integrase: LATE BLOCK DURING VIRAL REPLICATION AND ABNORMAL MULTIMERIZATION INVOLVING SPECIFIC PROTEIN DOMAINS. *Journal of Biological Chemistry* **2014**, *289* (30), 20477-20488.
36. Chandler, D., Interfaces and the driving force of hydrophobic assembly. *Nature* **2005**, *437* (7059), 640-647.
37. Berne, B. J.; Weeks, J. D.; Zhou, R., Dewetting and hydrophobic interaction in physical and biological systems. *Annu Rev Phys Chem* **2009**, *60*, 85-103.
38. Cui, D.; Ou, S.; Patel, S., Free energetics of rigid body association of ubiquitin binding domains: A biochemical model for binding mediated by hydrophobic interaction: Ubiquitin: Model Hydrophobic Interactions. *Proteins: Structure, Function, and Bioinformatics* **2014**, *82* (7), 1453-1468.
39. Zangi, R.; Berne, B. J., Temperature dependence of dimerization and dewetting of large-scale hydrophobes: a molecular dynamics study. *J Phys Chem B* **2008**, *112* (29), 8634-44.
40. Choudhury, N.; Pettitt, B. M., Enthalpy-entropy contributions to the potential of mean force of nanoscopic hydrophobic solutes. *J Phys Chem B* **2006**, *110* (16), 8459-63.
41. Ulucan, O.; Jaitly, T.; Helms, V., Energetics of Hydrophilic Protein-Protein Association and the Role of Water. *J Chem Theory Comput* **2014**, *10* (8), 3512-24.
42. Wang, L.; Siu, S. W.; Gu, W.; Helms, V., Downhill binding energy surface of the barnase-barstar complex. *Biopolymers* **2010**, *93* (11), 977-85.
43. Frisch, C.; Schreiber, G.; Johnson, C. M.; Fersht, A. R., Thermodynamics of the interaction of barnase and barstar: changes in free energy versus changes in enthalpy on mutation. *J Mol Biol* **1997**, *267* (3), 696-706.
44. Ben-Naim, A., On the driving forces for protein-protein association. *J Chem Phys* **2006**, *125* (2), 24901.
45. Chang, C. E.; McLaughlin, W. A.; Baron, R.; Wang, W.; McCammon, J. A., Entropic contributions and the influence of the hydrophobic environment in promiscuous protein-protein association. *Proc Natl Acad Sci U S A* **2008**, *105* (21), 7456-61.
46. Hoefling, M.; Gottschalk, K. E., Barnase-Barstar: from first encounter to final complex. *J Struct Biol* **2010**, *171* (1), 52-63.

47. Liu, P.; Huang, X.; Zhou, R.; Berne, B. J., Observation of a dewetting transition in the collapse of the melittin tetramer. *Nature* **2005**, *437* (7055), 159-62.
48. Zhou, R.; Huang, X.; Margulis, C. J.; Berne, B. J., Hydrophobic collapse in multidomain protein folding. *Science* **2004**, *305* (5690), 1605-9.
49. Wallqvist, A.; Gallicchio, E.; Levy, R. M., A Model for Studying Drying at Hydrophobic Interfaces: Structural and Thermodynamic Properties†. *The Journal of Physical Chemistry B* **2001**, *105* (28), 6745-6753.
50. Lum, K.; Luzar, A., Pathway to surface-induced phase transition of a confined fluid. *Physical Review E* **1997**, *56* (6), R6283-R6286.
51. Huang, X.; Margulis, C. J.; Berne, B. J., Dewetting-induced collapse of hydrophobic particles. *Proc Natl Acad Sci U S A* **2003**, *100* (21), 11953-8.
52. Choudhury, N.; Pettitt, B. M., On the mechanism of hydrophobic association of nanoscopic solutes. *J Am Chem Soc* **2005**, *127* (10), 3556-67.
53. Giovambattista, N.; Lopez, C. F.; Rossky, P. J.; Debenedetti, P. G., Hydrophobicity of protein surfaces: Separating geometry from chemistry. *Proc Natl Acad Sci U S A* **2008**, *105* (7), 2274-9.
54. Ricci, C. G.; McCammon, J. A., Heterogeneous Solvation in Distinctive Protein-Protein Interfaces Revealed by Molecular Dynamics Simulations. *J Phys Chem B* **2018**, *122* (49), 11695-11701.