

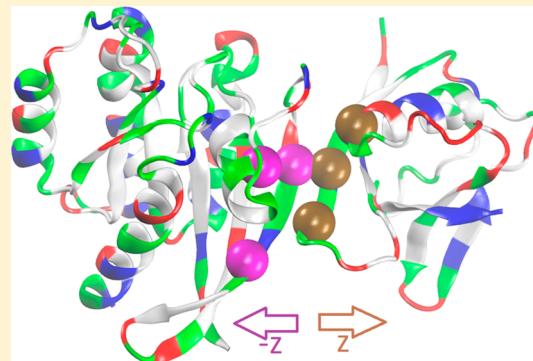
Computing Protein–Protein Association Affinity with Hybrid Steered Molecular Dynamics

Roberto A. Rodriguez,[†] Lili Yu,[†] and Liao Y. Chen*

Department of Physics, University of Texas at San Antonio, One UTSA Circle, San Antonio, Texas 78249, United States

 Supporting Information

ABSTRACT: Computing protein–protein association affinities is one of the fundamental challenges in computational biophysics/biochemistry. The overwhelming amount of statistics in the phase space of very high dimensions cannot be sufficiently sampled even with today's high-performance computing power. In this article, we extend a potential of mean force (PMF)-based approach, the hybrid steered molecular dynamics (hSMD) approach we developed for ligand–protein binding, to protein–protein association problems. For a protein complex consisting of two protomers, P1 and P2, we choose m (≥ 3) segments of P1 whose m centers of mass are to be steered in a chosen direction and n (≥ 3) segments of P2 whose n centers of mass are to be steered in the opposite direction. The coordinates of these $m + n$ centers constitute a phase space of $3(m + n)$ dimensions ($3(m + n)D$). All other degrees of freedom of the proteins, ligands, solvents, and solutes are freely subject to the stochastic dynamics of the all-atom model system. Conducting SMD along a line in this phase space, we obtain the $3(m + n)D$ PMF difference between two chosen states: one single state in the associated state ensemble and one single state in the dissociated state ensemble. This PMF difference is the first of four contributors to the protein–protein association energy. The second contributor is the $3(m + n - 1)D$ partial partition in the associated state accounting for the rotations and fluctuations of the $(m + n - 1)$ centers while fixing one of the $m + n$ centers of the P1–P2 complex. The two other contributors are the $3(m - 1)D$ partial partition of P1 and the $3(n - 1)D$ partial partition of P2 accounting for the rotations and fluctuations of their $m - 1$ or $n - 1$ centers while fixing one of the m/n centers of P1/P2 in the dissociated state. Each of these three partial partitions can be factored exactly into a 6D partial partition in multiplication with a remaining factor accounting for the small fluctuations while fixing three of the centers of P1, P2, or the P1–P2 complex, respectively. These small fluctuations can be well-approximated as Gaussian, and every 6D partition can be reduced in an exact manner to three problems of 1D sampling, counting the rotations and fluctuations around one of the centers as being fixed. We implement this hSMD approach to the Ras–RalGDS complex, choosing three centers on RalGDS and three on Ras ($m = n = 3$). At a computing cost of about 71.6 wall-clock hours using 400 computing cores in parallel, we obtained the association energy, -9.2 ± 1.9 kcal/mol on the basis of CHARMM 36 parameters, which well agrees with the experimental data, -8.4 ± 0.2 kcal/mol.



INTRODUCTION

Accurately computing the free-energy of binding proteins to proteins or ligands is a task of essential importance in biochemical and biophysical studies that still represents a considerable challenge to us even with today's high-performance computing power.^{1–22} An effective approach in the current literature is to use the relationship^{2,5,13,23} between the potential of mean force (PMF)^{24–28} and the binding affinity. These PMF-based and other equilibrium sampling approaches have a crucial reliance upon delicate choices of biasing/constraining potentials during the simulation processes and careful removal of the artifacts caused by these artificial potentials. The nonequilibrium steered molecular dynamics (SMD)^{29–48} approach, brute force in a certain sense, can be very efficient in sampling forced transition paths from the bound state to the dissociated state, but it has not been used reliably for free-energy calculations with quantitative accuracy⁴⁶

except for the recent development of the hybrid steered molecular dynamics (hSMD) approach for ligand–protein binding problems.⁴⁸

In this further development of the hSMD, we derive a formulation for the protein–protein association affinity of a protein complex consisting of two protomers, P1 and P2, associated by noncovalent interactions between P1 and P2. This hSMD approach is based on the relationship between the PMF and the binding affinity in the established literature. The widely used SMD involves pulling one center of mass of one selection of the ligand's atoms using a spring of finite, carefully chosen stiffness. In contrast, this hSMD approach involves pulling m ($m = 3, 4, \dots$) centers of mass of m selected segments of P1 (using m springs of infinite stiffness to disallow any

Received: April 11, 2015

fluctuations of the pulling centers along the way) and likewise (but in the opposite direction) pulling n ($n = 3, 4, \dots$) centers of mass of n selected segments of P2 to produce a $3(m + n)$ -dimensional ($3(m + n)D$) PMF curve leading from the associated state to the dissociated state of the P1–P2 complex. This PMF difference between the associated state and the dissociated state gives a large (but not dominant) part of the absolute association free energy. Three additional parts represent the rotations and fluctuations of the protomers individually in the dissociated state and of the complex in the associated state. The second part is from the $3(m + n - 1)D$ partial partition in the associated state accounting for the rotation and fluctuations of the $(m + n - 1)$ centers while fixing one of the $m + n$ centers of the P1–P2 complex. The two other parts are from the $3(m - 1)D$ partial partition of P1 and the $3(n - 1)D$ partial partition of P2 accounting for the rotations and fluctuations of their $m - 1$ or $n - 1$ centers while fixing one of the m/n centers of P1/P2 in the dissociated state. Each of these three partial partitions will be shown to factor exactly into a $6D$ partial partition in multiplication with a remaining factor. That remaining factor counts for the small fluctuations while fixing three of the centers of P1, P2, or the P1–P2 complex, respectively. These small fluctuations can be well-approximated as Gaussian, and every $6D$ partition can be reduced in an exact manner to three problems of $1D$ sampling, counting the rotations and fluctuations around one of the centers as being fixed.

We carry out applications of this hSMD approach to the Ras–RalGDS complex, whose association affinity was experimentally measured and whose structure was determined (PDB code: 1LFD). The computing time required was 71.6 wall-clock hours (all-atom model of 96 253 atoms), using 40 Intel Xeon E5-2680 v2 Ivy Bridge 2.8 GHz processors (400 cores) in parallel. The computed absolute free energy of association agrees well with the experimental data.

METHODS

Association Affinity/Energy from the $3(m + n)D$ PMF. Following the established literature,^{2,5,15} the association affinity of a dimer protein (consisting of protomers P1 and P2) is related to the $6D$ PMF as follows

$$\frac{1}{k_D/c_0} = \frac{c_0 \int_{\text{site}} d^3x_i^{(P2)} \exp[-W[\mathbf{r}_{10}^{(P1)}, \mathbf{r}_i^{(P2)}]/k_B T]}{\exp[-W[\mathbf{r}_{1\infty}^{(P1)}, \mathbf{r}_{1\infty}^{(P2)}]/k_B T]_{\text{bulk}}} \quad (1)$$

where c_0 is the standard concentration. For clarity and for convenience of unit conversion, we use two different but equivalent forms, $c_0 = 1 \text{ M}$ on the left-hand side and $c_0 = 6.02 \times 10^{-4}/\text{\AA}^3$ on the right-hand side of the equation. k_B is the Boltzmann constant, and T is the absolute temperature. The $3D$ integrations ($d^3x_i^{(P2)} \equiv dr_i^{(P2)}$) are over the x , y , and z coordinates of the position $\mathbf{r}_i^{(P2)}$ of protomer P2 that can be chosen as the center of mass of one segment of or the whole protomer P2. Likewise, the position $\mathbf{r}_i^{(P1)}$ of protomer P1 can be chosen as the center of mass of one segment of or the whole protomer P1. In general, $\mathbf{r}_i^{(P1)}$ can be fixed at an arbitrarily chosen $\mathbf{r}_{10}^{(P1)}$ that will be chosen as a point near the center of mass of our model system. The integral has the units of \AA^3 that renders the right-hand side dimensionless, as it should be. $W[\mathbf{r}_i^{(P1)}, \mathbf{r}_i^{(P2)}]$ is the $6D$ PMF that is a function of the positions of the protomers P1 and P2. The subscripts site and bulk indicate, respectively, that $\mathbf{r}_i^{(P2)}$ is near the PMF minimum and

that $\mathbf{r}_i^{(P2)} = \mathbf{r}_{1\infty}^{(P2)}$ is in the bulk region where the two protomers P1 and P2 are far away from one another.

Since the size of protomer P1 (or P2) is not small and the shape is not simple, the position of one segment center $\mathbf{r}_i^{(P1)}$ (or $\mathbf{r}_i^{(P2)}$) will not be sufficient/efficient to represent its location and situation. Instead, the protein complex can be better described with the positions $(\mathbf{r}_1^{(P1)}, \mathbf{r}_2^{(P1)}, \dots, \mathbf{r}_m^{(P1)}, \mathbf{r}_1^{(P2)}, \mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_n^{(P2)})$ of $m + n$ centers of mass of its $m + n$ chosen segments. Figure 1 shows

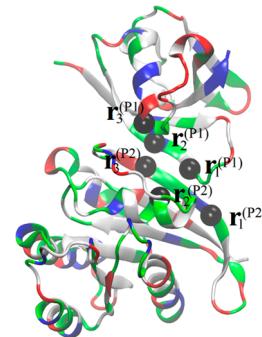


Figure 1. Three alpha carbons of protomer P1 (RalGDS) and three alpha carbons of P2 (Ras) are shown as black balls marked with their position vectors. The proteins are shown as ribbons colored by residue types. The coordinates are taken from the crystallographic structure (PDB code: 1LFD, chains A and B). All graphics in this article were rendered with VMD.⁴⁹

an example of $m = n = 3$ where the positions $(\mathbf{r}_1^{(P1)}, \mathbf{r}_2^{(P1)}, \mathbf{r}_3^{(P1)}, \mathbf{r}_1^{(P2)}, \mathbf{r}_2^{(P2)}, \mathbf{r}_3^{(P2)})$ of the six alpha carbons (three of P1 and three of P2) are chosen to quantify the location and situation of the complex. In the associated state, the $m + n - 1$ positions fluctuate without being in any way biased/constrained during the equilibrium MD simulation while $\mathbf{r}_i^{(P1)}$ is fixed at $\mathbf{r}_{10}^{(P1)}$. They are steered during the SMD runs from the bound state to the dissociated state for constructing the $3(m + n)D$ PMF $W[\mathbf{r}_1^{(P1)}, \mathbf{r}_2^{(P1)}, \dots, \mathbf{r}_m^{(P1)}, \mathbf{r}_1^{(P2)}, \mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_n^{(P2)}]$ as a function of these positions. In the dissociated state, one of the m centers on protomer P1 and one on P2 will be fixed at $\mathbf{r}_i^{(P1)} = \mathbf{r}_{1\infty}^{(P1)}$ and $\mathbf{r}_i^{(P2)} = \mathbf{r}_{1\infty}^{(P2)}$, while all others $(\mathbf{r}_2^{(P1)}, \dots, \mathbf{r}_m^{(P1)}, \mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_n^{(P2)})$ rotate and fluctuate according the stochastic dynamics of the system without any other bias/constraint.

Note that in the relationship between the $6D$ and $3(m + n)D$ PMFs

$$\begin{aligned} & \exp[-W[\mathbf{r}_i^{(P1)}, \mathbf{r}_i^{(P2)}]/k_B T] \\ &= C \int \prod_{i=2}^m d^3x_i^{(P1)} \prod_{i=2}^n d^3x_i^{(P2)} \\ & \quad \times \exp[-W[\mathbf{r}_1^{(P1)}, \mathbf{r}_2^{(P1)}, \dots, \mathbf{r}_m^{(P1)}, \mathbf{r}_1^{(P2)}, \mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_n^{(P2)}]/k_B T] \end{aligned} \quad (2)$$

the $3(m + n - 2)D$ integration over the $(m + n - 2)$ positions $(\mathbf{r}_2^{(P1)}, \dots, \mathbf{r}_m^{(P1)}, \mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_n^{(P2)})$ is effectively in a defined neighborhood of $(\mathbf{r}_1^{(P1)}, \mathbf{r}_1^{(P2)})$ because the protomers, each as one whole molecule, dictate that the m (or n) centers cannot be stretched much farther away from one another than its molecular size. When $\mathbf{r}_i^{(P2)}$ is near the binding site, so will be $(\mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_n^{(P2)})$. When the complex is in the dissociated state, $\mathbf{r}_{1\infty}^{(P2)}$ of P2 needs to be so far away from the other protomer P1 that integration over $(\mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_n^{(P2)})$ will be all in the region far from P1. C is the normalization constant that will be canceled out in the following expressions.

Making use of eq 2 twice in eq 1 (for the binding site and for the bulk), one has the following expression for the binding affinity

$$\frac{c_0}{k_D} = \left(c_0 \int_{\text{site}} \prod_{i=2}^m d^3 x_i^{(P1)} \prod_{i=1}^n d^3 x_i^{(P2)} \exp[-W[\mathbf{r}_{10}^{(P1)}, \mathbf{r}_2^{(P1)}, \dots, \mathbf{r}_m^{(P1)}; \mathbf{r}_1^{(P2)}, \mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_n^{(P2)}]/k_B T] \right) / \left(\int_{\text{bulk}} \prod_{i=2}^m d^3 x_i^{(P1)} \prod_{i=2}^n d^3 x_i^{(P2)} \exp[-W[\mathbf{r}_{10}^{(P1)}, \mathbf{r}_2^{(P1)}, \dots, \mathbf{r}_m^{(P1)}; \mathbf{r}_{10}^{(P2)}, \mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_n^{(P2)}]/k_B T] \right) \quad (3)$$

Now inserting the Boltzmann factor at a single state $(\mathbf{r}_{10}^{(P1)}, \mathbf{r}_{20}^{(P1)}, \dots, \mathbf{r}_{m0}^{(P1)}, \mathbf{r}_{10}^{(P2)}, \mathbf{r}_{20}^{(P2)}, \dots, \mathbf{r}_{n0}^{(P2)})$ chosen from the bound state ensemble and the Boltzmann factor at the corresponding dissociated state $(\mathbf{r}_{10}^{(P1)}, \mathbf{r}_{20}^{(P1)}, \dots, \mathbf{r}_{m0}^{(P1)}, \mathbf{r}_{10}^{(P2)}, \mathbf{r}_{20}^{(P2)}, \dots, \mathbf{r}_{n0}^{(P2)})$, the binding affinity can be expressed as three contributing factors: The partial partition function at the binding site Z_{m-1+n0} of the protein complex, the PMF difference between two chosen states $(\mathbf{r}_{10}^{(P1)}, \mathbf{r}_{20}^{(P1)}, \dots, \mathbf{r}_{m0}^{(P1)}; \mathbf{r}_{10}^{(P2)}, \mathbf{r}_{20}^{(P2)}, \dots, \mathbf{r}_{n0}^{(P2)})$ and $(\mathbf{r}_{10}^{(P1)}, \mathbf{r}_{20}^{(P1)}, \dots, \mathbf{r}_{m0}^{(P1)}, \mathbf{r}_{10}^{(P2)}, \mathbf{r}_{20}^{(P2)}, \dots, \mathbf{r}_{n0}^{(P2)})$, and the partial partition function in the dissociated state $Z_{m-1+n-1\infty}$. Mathematically

$$\frac{c_0}{k_D} = \left(\left(c_0 \int_{\text{site}} \prod_{i=2}^m d^3 x_i^{(P1)} \prod_{i=1}^n d^3 x_i^{(P2)} \exp[-W[\mathbf{r}_{10}^{(P1)}, \mathbf{r}_2^{(P1)}, \dots, \mathbf{r}_m^{(P1)}; \mathbf{r}_1^{(P2)}, \mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_n^{(P2)}]/k_B T] \right) / \left(\exp[-W[\mathbf{r}_{10}^{(P1)}, \mathbf{r}_{20}^{(P1)}, \dots, \mathbf{r}_{m0}^{(P1)}; \mathbf{r}_{10}^{(P2)}, \mathbf{r}_{20}^{(P2)}, \dots, \mathbf{r}_{n0}^{(P2)}]/k_B T] \right) \right) \times \left(\left(\exp[-W[\mathbf{r}_{10}^{(P1)}, \mathbf{r}_{20}^{(P1)}, \dots, \mathbf{r}_{m0}^{(P1)}; \mathbf{r}_{10}^{(P2)}, \mathbf{r}_{20}^{(P2)}, \dots, \mathbf{r}_{n0}^{(P2)}]/k_B T] \right) / \left(\exp[-W[\mathbf{r}_{1\infty}^{(P1)}, \mathbf{r}_{2\infty}^{(P1)}, \dots, \mathbf{r}_{m\infty}^{(P1)}; \mathbf{r}_{1\infty}^{(P2)}, \mathbf{r}_{2\infty}^{(P2)}, \dots, \mathbf{r}_{n\infty}^{(P2)}]/k_B T] \right) \right) \times \left(\left(\exp[-W[\mathbf{r}_{1\infty}^{(P1)}, \mathbf{r}_{2\infty}^{(P1)}, \dots, \mathbf{r}_{m\infty}^{(P1)}; \mathbf{r}_{1\infty}^{(P2)}, \mathbf{r}_{2\infty}^{(P2)}, \dots, \mathbf{r}_{n\infty}^{(P2)}]/k_B T] \right) / \left(\int_{\text{bulk}} \prod_{i=2}^m d^3 x_i^{(P1)} \prod_{i=2}^n d^3 x_i^{(P2)} \exp[-W[\mathbf{r}_{1\infty}^{(P1)}, \mathbf{r}_2^{(P1)}, \dots, \mathbf{r}_m^{(P1)}; \mathbf{r}_{1\infty}^{(P2)}, \mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_n^{(P2)}]/k_B T] \right) \right) = \frac{c_0 Z_{m-1+n0}}{Z_{m-1+n-1\infty}} \exp\left[\frac{-\Delta W_{0,\infty}}{k_B T}\right] \quad (4)$$

Throughout the text, the subscript 0 refers to the associated state and the subscript ∞ refers to the dissociated state. The notation Z_{m-1+n0} refers to the fact that in the associated state one of the m pulling centers on P1 is held fixed. Thus, the integration is over $m - 1 + n$ variables. Likewise, the notation $Z_{m-1+n-1\infty}$ refers to the fact that, in the dissociated state, P1 and

P2 each has one of the pulling centers fixed. Thus, the integration is over $m + n - 2$ variables. Here, $(\mathbf{r}_{10}^{(P1)}, \mathbf{r}_{20}^{(P1)}, \dots, \mathbf{r}_{m0}^{(P1)}; \mathbf{r}_1^{(P2)}, \mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_{n0}^{(P2)})$ can be connected to $(\mathbf{r}_{10}^{(P1)}, \mathbf{r}_{20}^{(P1)}, \dots, \mathbf{r}_{m0}^{(P1)}; \mathbf{r}_{10}^{(P2)}, \mathbf{r}_{20}^{(P2)}, \dots, \mathbf{r}_{n0}^{(P2)})$ via many curves in the $3(m + n)$ D space, but the PMF is a function of state; thus, computation of the PMF difference between the two states can be achieved along a single curve passing through them both. The $3(m + n)$ D PMF difference

$$\Delta W_{0,\infty} = W[\mathbf{r}_{10}^{(P1)}, \dots, \mathbf{r}_{m0}^{(P1)}; \mathbf{r}_{10}^{(P2)}, \dots, \mathbf{r}_{n0}^{(P2)}] - W[\mathbf{r}_{1\infty}^{(P1)}, \dots, \mathbf{r}_{m\infty}^{(P1)}; \mathbf{r}_{1\infty}^{(P2)}, \dots, \mathbf{r}_{n\infty}^{(P2)}] \quad (5)$$

is between one chosen bound state and its corresponding dissociated state. This PMF difference can be computed by means of the SMD simulations described in the latter part of this section. Note that the one chosen position of the ligand in the bound state

$$\begin{aligned} & (\mathbf{r}_{10}^{(P1)}, \dots, \mathbf{r}_{m0}^{(P1)}; \mathbf{r}_{10}^{(P2)}, \dots, \mathbf{r}_{n0}^{(P2)}) \\ & = (x_{10}^{(P1)}, y_{10}^{(P1)}, z_{10}^{(P1)}, \dots, x_{m0}^{(P1)}, y_{m0}^{(P1)}, z_{m0}^{(P1)}; x_{10}^{(P2)}, y_{10}^{(P2)}, z_{10}^{(P2)}, \dots, \\ & \quad x_{n0}^{(P2)}, y_{n0}^{(P2)}, z_{n0}^{(P2)}) \end{aligned} \quad (6)$$

is the starting point for SMD runs. It is taken from the bound state ensemble of the system. It does not have to be the minimum of the PMF but can be any one state in its close neighborhood. Note that we take the collection of coordinate vectors, e.g., eq 6, as a single-row $1 \times 3(m + n)$ matrix. The one state chosen from the dissociated state ensemble

$$\begin{aligned} & (\mathbf{r}_{1\infty}^{(P1)}, \dots, \mathbf{r}_{m\infty}^{(P1)}; \mathbf{r}_{1\infty}^{(P2)}, \dots, \mathbf{r}_{n\infty}^{(P2)}) \\ & = (x_{1\infty}^{(P1)}, y_{1\infty}^{(P1)}, z_{1\infty}^{(P1)}, \dots, x_{m\infty}^{(P1)}, y_{m\infty}^{(P1)}, z_{m\infty}^{(P1)}; x_{1\infty}^{(P2)}, y_{1\infty}^{(P2)}, z_{1\infty}^{(P2)}, \dots, \\ & \quad x_{n\infty}^{(P2)}, y_{n\infty}^{(P2)}, z_{n\infty}^{(P2)}) \end{aligned} \quad (7)$$

is related to the SMD starting point by a large enough displacement in the $3(m + n)$ D space

$$\begin{aligned} & (\mathbf{r}_{1\infty}^{(P1)}, \dots, \mathbf{r}_{m\infty}^{(P1)}; \mathbf{r}_{1\infty}^{(P2)}, \dots, \mathbf{r}_{n\infty}^{(P2)}) \\ & = (\mathbf{r}_{10}^{(P1)} - \mathbf{v}_d t, \dots, \mathbf{r}_{m0}^{(P1)} - \mathbf{v}_d t; \mathbf{r}_{10}^{(P2)} + \mathbf{v}_d t, \dots, \mathbf{r}_{n0}^{(P2)} + \mathbf{v}_d t) \end{aligned} \quad (8)$$

Here, \mathbf{v}_d is the constant velocity of the SMD pulling and t is the time it takes to steer/pull the P1–P2 complex apart from the associated state to the dissociated state.

As noted above, the partial partition function Z_{m-1+n0} of the bound state has the integration over $m + n - 1$ centers and thus has units of $\text{\AA}^{3(m+n-1)}$

$$\begin{aligned} Z_{m-1+n0} &= \int_{\text{site}} \prod_{i=2}^m d^3 x_i^{(P1)} \prod_{i=1}^n d^3 x_i^{(P2)} \\ &\exp\left[-\left(W[\mathbf{r}_{10}^{(P1)}, \mathbf{r}_2^{(P1)}, \dots, \mathbf{r}_m^{(P1)}; \mathbf{r}_1^{(P2)}, \mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_n^{(P2)}] - W[\mathbf{r}_{10}^{(P1)}, \mathbf{r}_{20}^{(P1)}, \dots, \mathbf{r}_{m0}^{(P1)}; \mathbf{r}_{10}^{(P2)}, \mathbf{r}_{20}^{(P2)}, \dots, \mathbf{r}_{n0}^{(P2)}]\right)/k_B T\right] \end{aligned} \quad (9)$$

The partial partition function $Z_{m-1+n-1\infty}$ of the dissociated state has the integration over $m + n - 2$ centers and thus has units of $\text{\AA}^{3(m+n-2)}$

$$Z_{m-1+n-1\infty} = \int_{\text{bulk}} \prod_{i=2}^m d^3x_i^{(P1)} \prod_{i=2}^n d^3x_i^{(P2)} \\ \exp \left[- \left(W[\mathbf{r}_{1\infty}^{(P1)}, \mathbf{r}_2^{(P1)}, \dots, \mathbf{r}_m^{(P1)}; \mathbf{r}_{1\infty}^{(P2)}, \mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_n^{(P2)}] - \right) / k_B T \right] \quad (10)$$

It should be noted that, in the dissociated state, the two protomers are not interacting with each other. Therefore, the partial partition function can be factored into two separate partial partition functions of the independent protomers, $Z_{m-1+n-1\infty} = Z_{m-1\infty}^{(P1)} Z_{n-1\infty}^{(P2)}$

$$Z_{m-1\infty}^{(P1)} = \int_{\text{bulk}} \prod_{i=2}^m d^3x_i^{(P1)} \exp[-(W[\mathbf{r}_{1\infty}^{(P1)}, \mathbf{r}_2^{(P1)}, \dots, \mathbf{r}_m^{(P1)}] \\ - W[\mathbf{r}_{1\infty}^{(P1)}, \mathbf{r}_{2\infty}^{(P1)}, \dots, \mathbf{r}_{m\infty}^{(P1)}]) / k_B T], \\ Z_{n-1\infty}^{(P2)} = \int_{\text{bulk}} \prod_{i=2}^n d^3x_i^{(P2)} \exp[-(W[\mathbf{r}_{1\infty}^{(P2)}, \mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_n^{(P2)}] \\ - W[\mathbf{r}_{1\infty}^{(P2)}, \mathbf{r}_{2\infty}^{(P2)}, \dots, \mathbf{r}_{n\infty}^{(P2)}]) / k_B T] \quad (11)$$

$Z_{m-1\infty}^{(P1)}$ refers to the partition function of the m segment centers of P1 in the dissociated state, taking into account explicitly the fact that one of the centers is fixed. $Z_{n-1\infty}^{(P2)}$ is the corresponding partition function for P2 in the dissociated state. Again, the use of $c_0 = 6.02 \times 10^{-4}/\text{\AA}^3$ on the right-hand side of eq 3 renders it a pure number, as desired, and the dissociation constant will conveniently be in units of $M = \text{mol/L}$. In summary, we have the following formulas for the association affinity and the absolute free energy of association

$$\frac{c_0}{k_D} = \frac{c_0 Z_{m-1+n0}}{Z_{m-1\infty}^{(P1)} Z_{n-1\infty}^{(P2)}} \exp \left[- \frac{\Delta W_{0,\infty}}{k_B T} \right], \\ \Delta G_{\text{hSMD}} = k_B T \ln \left[\frac{Z_{m-1\infty}^{(P1)} Z_{n-1\infty}^{(P2)}}{Z_{m-1+n0} c_0} \right] + \Delta W_{0,\infty} \quad (12)$$

In order to determine the association affinity/energy, one needs to compute four factors: (a) The two independent partial partition functions in the dissociated state (eq 11) that are $3(m-1)\text{D}$ and $3(n-1)\text{D}$, respectively. (b) The $3(m+n-1)\text{D}$ partial partition function in the associated state (eq 9). (c) The PMF difference (eq 5) between one chosen bound state and its corresponding dissociated state that can be computed by running SMD simulations of pulling the two protomers forward and backward along a $3(m+n)\text{D}$ line connecting the associated and dissociated states. Note that the PMF is a function of state (a point in the $3(m+n)\text{D}$ space) and that the PMF difference is independent of the paths connecting the two end states.

Partial Partitions in the Dissociated State. As detailed in the [Supporting Information](#), Section I, each of the two dissociated state partitions can be factored into a 6D partition and a $3(m-3)\text{D}$ ($3(n-3)\text{D}$) partition

$$Z_{m-1\infty}^{(P1)} = Z_{3-1\infty}^{(P1)} Z_{m-3\infty}^{(P1)}, \quad Z_{n-1\infty}^{(P2)} = Z_{3-1\infty}^{(P2)} Z_{n-3\infty}^{(P2)} \quad (13)$$

$Z_{3-1\infty}$ refers to the partial partition function in the dissociated state of three centers with the first center fixed. $Z_{m-3\infty}$ and $Z_{n-3\infty}$ are the partial partition functions of the $m-3$ ($n-3$) centers with the first three centers fixed. Naturally, when $m=3$, $Z_{3-3\infty}^{(P1)}=1$, and when $n=3$, $Z_{3-3\infty}^{(P2)}=1$. When $m>3$ and/or $n>3$,

$Z_{m-3\infty}^{(P1)}$ and $Z_{n-3\infty}^{(P2)}$ can be well-approximated as Gaussian because, when three centers of a protomer are fixed in the dissociated state, the location and orientation of the promoter will not change much but fluctuate according to stochastic dynamics. Since the computations of the two partial partitions are identical, we work on protomer P1 and simply replace the superscript P1 with P2 and m with n in the formulas for protomer P2. Fixing the first three pulling centers of P1 at $(\mathbf{r}_{1\infty}^{(P1)}, \mathbf{r}_{2\infty}^{(P1)}, \mathbf{r}_{3\infty}^{(P1)})$, the small fluctuations of the other ($m-3$) centers can be readily sampled to give

$$Z_{m-3\infty}^{(P1)} = (2\pi)^{3(m-3)/2} \text{Det}^{1/2}(\sum_{m-3\infty}^{(P1)}) \exp[\Delta_{m-3\infty}^{(P1)}/k_B T] \quad (14)$$

Here, the dimensionless quantity $\Delta_{m-3\infty}^{(P1)}/k_B T$ gives a measure of how far $(\mathbf{r}_{4\infty}^{(P1)}, \mathbf{r}_{5\infty}^{(P1)}, \dots, \mathbf{r}_{m\infty}^{(P1)})$, the final state of SMD, is from the PMF minimum ($\langle \mathbf{r}_4^{(P1)} \rangle, \langle \mathbf{r}_5^{(P1)} \rangle, \dots, \langle \mathbf{r}_m^{(P1)} \rangle$) of P1 in the dissociated state

$$\Delta_{m-3\infty}^{(P1)}/k_B T = \frac{1}{2} (\langle \mathbf{r}_4^{(P1)} \rangle - \mathbf{r}_{4\infty}^{(P1)}, \langle \mathbf{r}_5^{(P1)} \rangle - \mathbf{r}_{5\infty}^{(P1)}, \dots, \langle \mathbf{r}_m^{(P1)} \rangle - \mathbf{r}_{m\infty}^{(P1)}) (\sum_{m-3\infty}^{(P1)})^{-1} \\ \times (\langle \mathbf{r}_4^{(P1)} \rangle - \mathbf{r}_{4\infty}^{(P1)}, \langle \mathbf{r}_5^{(P1)} \rangle - \mathbf{r}_{5\infty}^{(P1)}, \dots, \langle \mathbf{r}_m^{(P1)} \rangle - \mathbf{r}_{m\infty}^{(P1)})^T \quad (15)$$

Det represents the determinant. $\sum_{m-3\infty}^{(P1)}$ is the $3(m-3) \times 3(m-3)$ matrix of the fluctuations/deviations of the pulling center coordinates $\delta \mathbf{x}_4^{(P1)} = \mathbf{x}_4^{(P1)} - \langle \mathbf{x}_{4\infty}^{(P1)} \rangle$, etc.

$$\sum_{m-3\infty}^{(P1)} = \left\langle \begin{array}{l} (\langle \mathbf{r}_4^{(P1)} \rangle - \mathbf{r}_{4\infty}^{(P1)}, \langle \mathbf{r}_5^{(P1)} \rangle - \mathbf{r}_{5\infty}^{(P1)}, \dots, \langle \mathbf{r}_m^{(P1)} \rangle - \mathbf{r}_{m\infty}^{(P1)})^T \\ \times (\langle \mathbf{r}_4^{(P1)} \rangle - \mathbf{r}_{4\infty}^{(P1)}, \langle \mathbf{r}_5^{(P1)} \rangle - \mathbf{r}_{5\infty}^{(P1)}, \dots, \langle \mathbf{r}_m^{(P1)} \rangle - \mathbf{r}_{m\infty}^{(P1)}) \end{array} \right\rangle \quad (16)$$

The brackets $\langle \cdot \rangle$ stand for the conditional statistical average over the dissociated state ensemble with three of the m centers being fixed at $(\mathbf{r}_{1\infty}^{(P1)}, \mathbf{r}_{2\infty}^{(P1)}, \mathbf{r}_{3\infty}^{(P1)})$. T means transposing a single-row, $1 \times 3(m-3)$ matrix into a single-column, $3(m-3) \times 1$ matrix. Practically, $\sum_{m-3\infty}^{(P1)}$ can be accurately evaluated by running equilibrium MD in the dissociated state of a protomer while three of the m centers are fixed at $(\mathbf{r}_{1\infty}^{(P1)}, \mathbf{r}_{2\infty}^{(P1)}, \mathbf{r}_{3\infty}^{(P1)})$. Replacing P1 with P2 and m with n , we have explicitly the $3(n-3)\text{D}$ partial partition of P2 along with the $3(m-3)\text{D}$ partial partition of P1

$$Z_{m-3\infty}^{(P1)} = (2\pi)^{3(m-3)/2} \text{Det}^{1/2}(\sum_{m-3\infty}^{(P1)}) \exp[\Delta_{m-3\infty}^{(P1)}/k_B T], \\ Z_{n-3\infty}^{(P2)} = (2\pi)^{3(n-3)/2} \text{Det}^{1/2}(\sum_{n-3\infty}^{(P2)}) \exp[\Delta_{n-3\infty}^{(P2)}/k_B T] \quad (17)$$

which constitute half of the computations required for the dissociated-state partial partitions in eq 13.

The other half required in eq 13, the 6D partial partitions, can be reduced into 3D because each protomer's environment in the dissociated state is spherically symmetrical around $\mathbf{r}_{1\infty}^{(P1)}$ or $\mathbf{r}_{1\infty}^{(P2)}$. As shown in the [Supporting Information](#), the 3D partition can be factored into three 1D sampling problems in an exact manner

$$Z_{3-1\infty}^{(P1)} = \frac{8\pi^2 [r_{21\infty}^{(P1)}]^2 [r_{31\infty}^{(P1)}]^2 \sin \theta_\infty^{(P1)}}{\rho_\infty(r_{21\infty}^{(P1)}) \rho_\infty(r_{31\infty}^{(P1)}) \rho_\infty(\theta_\infty^{(P1)})}, \\ Z_{3-1\infty}^{(P2)} = \frac{8\pi^2 [r_{21\infty}^{(P2)}]^2 [r_{31\infty}^{(P2)}]^2 \sin \theta_\infty^{(P2)}}{\rho_\infty(r_{21\infty}^{(P2)}) \rho_\infty(r_{31\infty}^{(P2)}) \rho_\infty(\theta_\infty^{(P2)})} \quad (18)$$

In the rest of this subsection, we imply the superscript (P1) for protomer P1 or (P2) for promoter P2 to save space without causing confusion because the two formulas in eq 18 have identical dimensions and structures. The factor of $8\pi^2 = 2\pi \times 4\pi$ accounts for the azimuthal symmetry around an axis ($\mathbf{r}_{2\infty} - \mathbf{r}_{1\infty}$) and the spherical symmetry around a point $\mathbf{r}_{1\infty}$. $r_{21} = |\mathbf{r}_2 - \mathbf{r}_{1\infty}|$ and $r_{31} = |\mathbf{r}_3 - \mathbf{r}_{1\infty}|$ are the distances between the two pulling centers. θ is the angle between the two vectors $\mathbf{r}_2 - \mathbf{r}_{1\infty}$ and $\mathbf{r}_3 - \mathbf{r}_{1\infty}$. $r_{21\infty} = |\mathbf{r}_{2\infty} - \mathbf{r}_{1\infty}|$ and $r_{31\infty} = |\mathbf{r}_{3\infty} - \mathbf{r}_{1\infty}|$ are the distances between the fixed centers. θ_∞ is the angle between $\mathbf{r}_{2\infty} - \mathbf{r}_{1\infty}$ and $\mathbf{r}_{3\infty} - \mathbf{r}_{1\infty}$. The three 1D probability distribution densities are

$$\begin{aligned}\rho_\infty(r_{21}) &= r_{21}^2 \exp[-W_\infty[r_{21}]/k_B T] / \int dr'_{21} r'^2 \exp[-W_\infty[r'_{21}]/k_B T], \\ \rho_\infty(r_{31}) &= r_{31}^2 \exp[-W_\infty[r_{31}]/k_B T] / \int dr'_{31} r'^2 \exp[-W_\infty[r'_{31}]/k_B T], \\ \rho_\infty(\theta) &= \sin \theta \exp[-W_\infty[\theta]/k_B T] / \int_0^\pi d\theta' \sin \theta' \exp[-W_\infty[\theta']/k_B T]\end{aligned}\quad (19)$$

Each of these three constitutes a 1D sampling problem that can be implemented efficiently to produce accurate results. $W_\infty[r_{21}]$ is the 1D PMF for stretching the protomer between the two pulling centers, steering the second pulling center \mathbf{r}_2 to and from the first pulling center that is fixed at $\mathbf{r}_{1\infty}$ along the axis passing through $(\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty})$. $W_\infty[r_{31}]$ is a function of r_{31} when the first two centers are fixed at $(\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty})$ and the third pulling center \mathbf{r}_3 is pulled to and from the first pulling center that is fixed at $\mathbf{r}_{1\infty}$ along the axis passing through $(\mathbf{r}_{1\infty}, \mathbf{r}_{3\infty})$. $W_\infty[\theta]$ is the angular PMF for the third center when the first two centers are fixed at $(\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty})$ but the third center \mathbf{r}_3 is freely subject to the stochastic dynamics of the system.

Partial Partitions in the Associated State. In this state, the stochastic dynamics of protomer P1 are tightly coupled to that of P2. It is not necessary to distinguish their coordinates into two separate groups. Instead, one can pick any three out of the $(m+n)$ centers $(\mathbf{r}_1^{(P1)}, \mathbf{r}_2^{(P1)}, \dots, \mathbf{r}_m^{(P1)}, \mathbf{r}_1^{(P2)}, \mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_n^{(P2)})$ that are denoted $(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$. The rest $k = m+n-3$ centers are denoted $(\mathbf{r}_4, \mathbf{r}_5, \dots, \mathbf{r}_{m+n})$. With these notations and the detailed derivation in Supporting Information, Section II, we have

$$Z_{m-1+n0} = Z_{3-1,0} Z_{k0} \quad (20)$$

Here, the 6D partial partition $Z_{3-1,0}$ represents fluctuations of two centers $(\mathbf{r}_2, \mathbf{r}_3)$ when one center is fixed at $\mathbf{r}_1 = \mathbf{r}_{10}$ that can be factored into three 1D sampling problems in the same manner as eq 18

$$Z_{3-1,0} = \frac{8\pi^2 r_{210}^2 r_{310}^2 \sin \theta_0}{\rho_0(r_{210}) \rho_0(r_{310}) \rho_0(\theta_0)} \quad (21)$$

Noting that the probability distributions here in the associated state take forms similar to those of the dissociated state in eq 19

$$\begin{aligned}\rho_0(r_{21}) &= r_{21}^2 \exp[-W_0[r_{21}]/k_B T] / \int dr_{21} r_{21}^2 \exp[-W_0[r_{21}]/k_B T], \\ \rho_0(r_{31}) &= r_{31}^2 \exp[-W_0[r_{31}]/k_B T] / \int dr_{31} r_{31}^2 \exp[-W_0[r_{31}]/k_B T], \\ \rho_0(\theta) &= \sin \theta \exp[-W_0[\theta]/k_B T] / \int_0^\pi d\theta \sin \theta \exp[-W_0[\theta]/k_B T]\end{aligned}\quad (22)$$

$W_0[r_{31}]$ as a function of $\mathbf{r}_{31} = |\mathbf{r}_1 - \mathbf{r}_3|$, the distance between the first and the third center, is the 1D PMF along the $(\mathbf{r}_{10}, \mathbf{r}_{30})$ direction when two centers are fixed at $(\mathbf{r}_{10}, \mathbf{r}_{20})$. $W_0[r_{21}]$ as a function of $\mathbf{r}_{21} = |\mathbf{r}_1 - \mathbf{r}_2|$, the distance between the first and the second center, is the 1D PMF along the $(\mathbf{r}_{10}, \mathbf{r}_{20})$ direction when only one center is fixed at (\mathbf{r}_{10}) . $\mathbf{r}_{310} = |\mathbf{r}_{10} - \mathbf{r}_{30}|$ and $\mathbf{r}_{210} = |\mathbf{r}_{10} - \mathbf{r}_{20}|$ are, respectively, the distance between the third and first pulling centers and the same between the second and first pulling centers.

Z_{k0} represents fluctuations of k centers $(\mathbf{r}_4, \mathbf{r}_5, \dots, \mathbf{r}_{m+n})$ when three center are fixed at $(\mathbf{r}_{10}, \mathbf{r}_{20}, \mathbf{r}_{30})$. When the protein association is tight, one can approximate the fluctuations as Gaussian in the neighborhood of the PMF minimum. The coordinates of the minimum of a Gaussian distribution are equal to the average coordinates, of course, $(\langle \mathbf{r}_4 \rangle, \langle \mathbf{r}_5 \rangle, \dots, \langle \mathbf{r}_{m+n} \rangle)$.

$$Z_{k0} = (2\pi)^{3k/2} \text{Det}^{1/2}(\Sigma_{k0}) \exp[\Delta_{k0}/k_B T] \quad (23)$$

Here, the dimensionless quantity $\Delta_{k0}/k_B T$ gives a measure of how far $(\mathbf{r}_{40}, \mathbf{r}_{50}, \dots, \mathbf{r}_{m+n0})$, the initial state chosen for SMD, is from the PMF minimum $(\langle \mathbf{r}_4 \rangle, \langle \mathbf{r}_5 \rangle, \dots, \langle \mathbf{r}_{m+n} \rangle)$

$$\begin{aligned}\Delta_{k0}/k_B T &= \frac{1}{2} (\langle \mathbf{r}_4 \rangle - \mathbf{r}_{40}, \langle \mathbf{r}_5 \rangle - \mathbf{r}_{50}, \dots, \langle \mathbf{r}_{m+n} \rangle - \mathbf{r}_{m+n0}) \\ &\quad \Sigma_{k0}^{-1} (\langle \mathbf{r}_4 \rangle - \mathbf{r}_{40}, \langle \mathbf{r}_5 \rangle - \mathbf{r}_{50}, \dots, \langle \mathbf{r}_{m+n} \rangle - \mathbf{r}_{m+n0})^T\end{aligned}\quad (24)$$

Σ_{k0} is the $3k \times 3k$ matrix of the fluctuations/deviations of the pulling center coordinates $\delta \mathbf{x}_4 = \mathbf{x}_4 - \langle \mathbf{x}_4 \rangle$, etc.

$$\begin{aligned}\Sigma_{k0} &= \langle (\langle \mathbf{r}_4 \rangle - \mathbf{r}_{40}, \langle \mathbf{r}_5 \rangle - \mathbf{r}_{50}, \dots, \langle \mathbf{r}_{m+n} \rangle - \mathbf{r}_{m+n0}) \rangle \\ &\quad \langle (\mathbf{r}_4) - \mathbf{r}_{40}, \langle \mathbf{r}_5 \rangle - \mathbf{r}_{50}, \dots, \langle \mathbf{r}_{m+n} \rangle - \mathbf{r}_{m+n0} \rangle \rangle\end{aligned}\quad (25)$$

Σ_{k0}^{-1} is the inverse matrix of Σ_{k0} which can be accurately evaluated by running equilibrium MD in the associated state of the protein–protein complex while three of $m+n$ centers are fixed at $(\mathbf{r}_{10}, \mathbf{r}_{20}, \mathbf{r}_{30})$.

PMF from SMD Simulations. In an SMD²⁹ simulation of the current literature, one steers/pulls one center of mass of one selection of atoms using a spring with a carefully chosen stiffness (spring constant). The use of a spring of finite stiffness introduces additional fluctuation and dissipation in the added degrees of freedom.⁴⁴ In this article, we choose $m+n$ segments (mutually exclusive $m+n$ selections of atoms) of protomers P1 and P2 for steering/pulling with $m+n$ infinitely stiff springs ($m, n = 1, 2, 3, \dots$). Namely, the $m+n$ centers of mass of the chosen $m+n$ segments will be controlled as functions of time t

$$\mathbf{r}_i = \mathbf{r}_{iA} \pm \mathbf{v}_d t, \quad i = 1 \dots m+n \quad (26)$$

while all other degrees of freedom of the system are freely subject to stochastic dynamics. Here, $\mathbf{r}_i = (x_i, y_i, z_i)$ is the center of mass coordinates of the i th segments, \mathbf{v}_d is the pulling velocity, and \mathbf{r}_{iA} are coordinates of the centers of mass of the

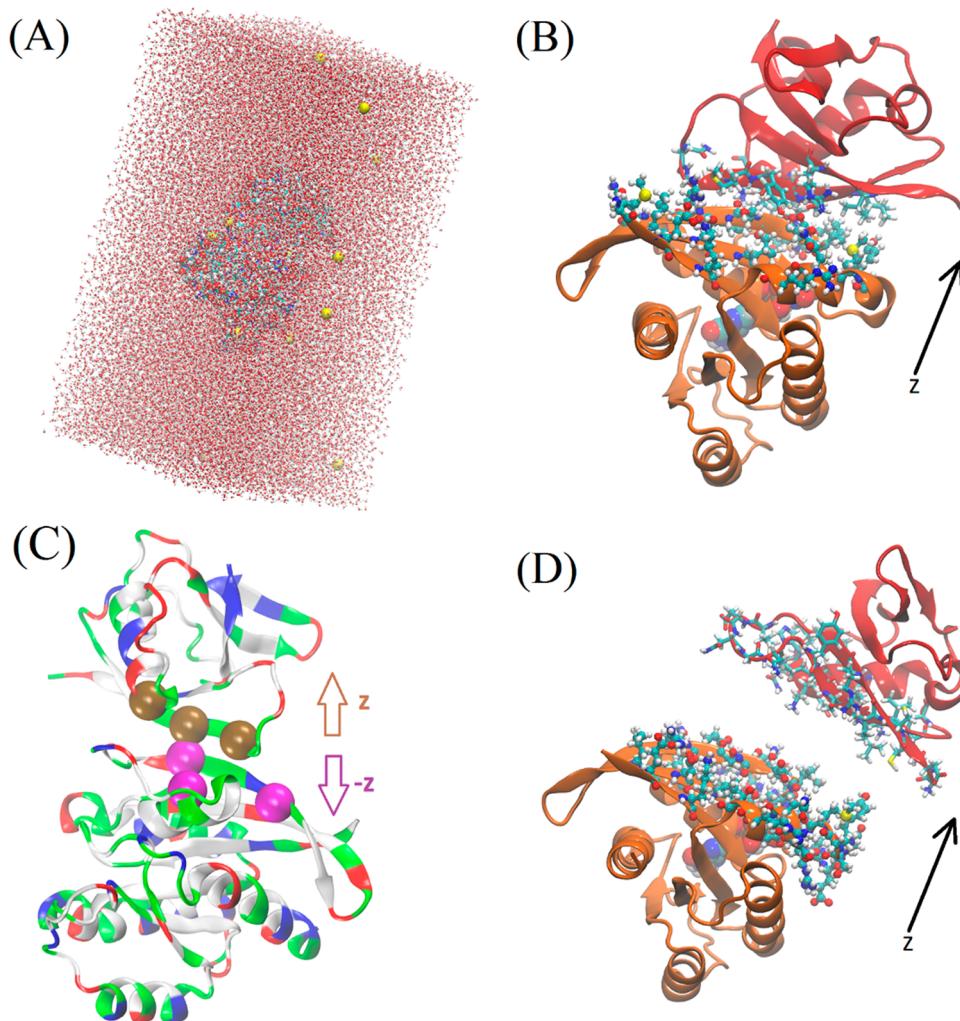


Figure 2. All-atom model system of the Ras–RalGDS complex. (A) The simulation box of $80 \times 80 \times 147 \text{ \AA}^3$ in dimensions. The proteins reside near the center of the water box. Also visible are some of the counterions. (B) The protein complex in its associated state. (C) The Ras–RalGDS complex (ribbons, colored by residue type) with six steered centers marked: Three alpha carbons of Asn29, Lys32, and Lys52 are shown as gold spheres to be steered along the z -direction. Three alpha carbons of Ile221, Asp238, and Lys242 are shown as purple spheres to be steered along the negative z -direction. (D) Ras and RalGDS in the dissociated state. Shown in (B) and (D) are Ras (colored yellow) and RalGDS (colored red) in ribbons, GTP in large spheres (colored by atom names), and the interfacial residues of Ras (in ball and stick, colored by atom) and of RalGDS (in licorice, colored by atom).

steered segments at the end state A. The + and - signs are for the forward and reverse pulling paths, respectively, for protomer P1, and the opposites of these signs are for P2. $\{\mathbf{r}_i\}$ denotes $(\mathbf{r}_1^{(P1)}, \mathbf{r}_2^{(P1)}, \dots, \mathbf{r}_m^{(P1)}; \mathbf{r}_1^{(P2)}, \mathbf{r}_2^{(P2)}, \dots, \mathbf{r}_n^{(P2)})$, etc. We adopt the multisectioinal scheme of ref 50. The path from the bound state to the dissociated state is divided into a number of sections. Within a given section whose end states are marked as A and B, respectively, multiple forward and reverse pulling paths are sampled along which the work done to the system is recorded. The Gibbs free-energy difference (namely, the PMF or the reversible work) is computed via the Brownian dynamics fluctuation-dissipation theorem (BD-FDT)⁵¹ as follows

$$W[\{\mathbf{r}_i\}] - W[\{\mathbf{r}_{iA}\}] = -k_B T \ln \left(\frac{\langle \exp[-W_{\{\mathbf{r}_{iA}\}} \rightarrow \{\mathbf{r}_i\}]/2k_B T] \rangle_F}{\langle \exp[-W_{\{\mathbf{r}_i\}} \rightarrow \{\mathbf{r}_{iA}\}]/2k_B T] \rangle_R} \right) \quad (27)$$

Here, the brackets with subscript F/R represent the statistical average over the forward/reverse paths. $W_{\{\mathbf{r}_{iA}\}} \rightarrow \{\mathbf{r}_i\}$ is the work

done to the system along a forward path when the proteins are steered from A to \mathbf{r} . $W_{\{\mathbf{r}_i\}} \rightarrow \{\mathbf{r}_{iA}\} = W_{\{\mathbf{r}_{iB}\}} \rightarrow \{\mathbf{r}_{iA}\} - W_{\{\mathbf{r}_{iB}\}} \rightarrow \{\mathbf{r}_i\}$ is the work for the part of a reverse path when the protomers are pulled from \mathbf{r} to A. $\{\mathbf{r}_{iA}\}$, $\{\mathbf{r}_i\}$, and $\{\mathbf{r}_{iB}\}$ are the coordinates of the centers of mass of the steered segments at the end state A, the general state \mathbf{r} , and the end state B of the system, respectively. At each end of a section, A/B, the system is equilibrated for a time long enough to reach conditioned equilibrium while the steered centers are fixed at $\{\mathbf{r}_{iA}\}/\{\mathbf{r}_{iB}\}$. In this way, running SMD section by section, we map the PMF $W[\{\mathbf{r}_i\}]$ as a function of the steered centers along a chosen path from the associated state to the dissociated state.

Simulation Parameters. In all of the equilibrium MD and nonequilibrium SMD runs, we used the CHARMM36^{52,53} force field for all intra- and intermolecular interactions. We implemented Langevin stochastic dynamics with NAMD⁵⁴ to simulate the systems at constant temperature of 298 K and constant pressure of 1 bar. Full electrostatics was implemented by means of particle mesh Ewald (PME) at $128 \times 128 \times 256$. The time step was 1 fs for short-range and 2 fs for long-range

Table 1. Summary of Dissociation Constant Calculation

quantity	description	equation/calculation method
k_D	dissociation constant	$\frac{1}{k_D} = \frac{Z_{m-1+n-1\infty}}{Z_{m-1+n-1\infty}} \exp \left[\frac{-\Delta W_{0,\infty}}{k_B T} \right]$
$Z_{m-1+n-1\infty}$	partial partition function of proteins in the dissociated state (one center fixed on each protein)	$Z_{m-1+n-1\infty} = Z_{m-1\infty}^{(P1)} Z_{n-1\infty}^{(P2)}$
Z_{m-1+n0}	partition function in the associated state (one center fixed on the protein complex)	$Z_{m-1+n0} = Z_{3-1,0} k_0$
$\Delta W_{0,\infty}$	PMF difference between a chosen state in the associated state ensemble and the corresponding dissociated state.	calculated through SMD, eq 27

Table 2. Partial Partition of the Associated State

Z_{m-1+n0}	partial partition function in the associated state (one center fixed on the complex)	$Z_{m-1+n0} = Z_{3-1,0} k_0$
$Z_{3-1,0}$	partial partition for the fluctuations of the complex in the associated state when one center is fixed	$Z_{3-1,0} = \frac{8\pi^2 r_{210}^2 r_{310}^2 \sin \theta_0}{\rho_0(r_{210}) \rho_0(r_{310}) \rho_0(\theta_0)}$
r_{ij0}	distance between \mathbf{r}_{i0} and \mathbf{r}_{j0} (positions of the i th and j th centers in the one state chosen from the associated state ensemble)	$r_{ij0} = \mathbf{r}_{i0} - \mathbf{r}_{j0} $
θ_0	angle between $\mathbf{r}_{20} - \mathbf{r}_{10}$ and $\mathbf{r}_{30} - \mathbf{r}_{10}$	$\arccos \left[\frac{(\mathbf{r}_{20} - \mathbf{r}_{10}) \cdot (\mathbf{r}_{30} - \mathbf{r}_{10})}{ \mathbf{r}_{20} - \mathbf{r}_{10} \mathbf{r}_{30} - \mathbf{r}_{10} } \right]$
$\rho_0(\cdot)$	1D probability distribution densities	Calculated through MD sampling, see eq 22
Z_{k0}	partial partition for fluctuations of the $k = m + n - 3$ centers in the associated state with the three of the $m + n$ centers fixed	Gaussian approximation, see eq 23

Table 3. Partial Partition Function of One Protein (P1) in the Dissociated State^a

$Z_{m-1+n-1\infty}$	partial partition function of proteins in the dissociated state (one center fixed on each protein)	$Z_{m-1+n-1\infty} = Z_{3-1\infty}^{(P1)} Z_{m-3\infty}^{(P1)} Z_{3-1\infty}^{(P2)} Z_{n-3\infty}^{(P2)}$
$Z_{3-1\infty}^{(P1)}$	partial partition of P1 in the dissociated state when one of the centers fixed	$Z_{3-1\infty}^{(P1)} = \frac{8\pi^2 [r_{21\infty}^{(P1)}]^2 [r_{31\infty}^{(P1)}]^2 \sin \theta_{\infty}^{(P1)}}{\rho_{\infty}(r_{21\infty}^{(P1)}) \rho_{\infty}(r_{31\infty}^{(P1)}) \rho_{\infty}(\theta_{\infty}^{(P1)})}$
$r_{ij\infty}$	distance between $\mathbf{r}_{i\infty}$ and $\mathbf{r}_{j\infty}$ (positions of the i th and the j th centers in the dissociated state corresponding to the one state chosen from the associated state ensemble)	$r_{ij\infty} = \mathbf{r}_{i\infty} - \mathbf{r}_{j\infty} $
θ_{∞}	angle between $\mathbf{r}_{2\infty} - \mathbf{r}_{1\infty}$ and $\mathbf{r}_{3\infty} - \mathbf{r}_{1\infty}$	$\arccos \left[\frac{(\mathbf{r}_{2\infty} - \mathbf{r}_{1\infty}) \cdot (\mathbf{r}_{3\infty} - \mathbf{r}_{1\infty})}{ \mathbf{r}_{2\infty} - \mathbf{r}_{1\infty} \mathbf{r}_{3\infty} - \mathbf{r}_{1\infty} } \right]$
$\rho_{\infty}(\cdot)$	1D probability distribution densities	calculated through MD sampling, see eq 19
$Z_{m-3\infty}^{(P1)}$	partial partition of P1 in the dissociated state when three of the m centers fixed	Gaussian approximation, see eq 17

^aIdentical formulas for protein P2 are obtained by substitution of P1 with P2.

Table 4. Computed Results of Ras–RalGDS Complex

the 6D partial partition of ralGDS in the dissociated state when one center is fixed	$Z_{3-1\infty}^{(P1)}$	$7.05 \times 10^5 \text{ \AA}^6$
the 6D partial partition of ras in the dissociated state when one center is fixed	$Z_{3-1\infty}^{(P2)}$	$1.34 \times 10^6 \text{ \AA}^6$
the 6D partial partition of Ras–RalGDS complex in the associated state when one center is fixed	$Z_{3-1,0}$	$7.63 \times 10^6 \text{ \AA}^6$
the 9 × 9 matrix for fluctuations in the associated state defined in eq 25	$\text{Det}(\sum_{3\infty})$	$7.036 \times 10^{-12} \text{ \AA}^{18}$
the deviations of the initial state from the equilibrium average state as measured in eq 24 for $k = 3$	$\Delta_{3\infty}$	3.79 kcal/mol
PMF difference between the one chosen associated state and the corresponding dissociated state	$\Delta W_{0,\infty}$	-18.2 kcal/mol
the absolute free energy of ras-ralGDS association, computed with hSMD, in aqueous solution of 50 mM Tris/HCl, 5 mM MgCl ₂ , and 100 mM NaCl	ΔG_{hSMD}	-9.2 ± 1.9 kcal/mol
the experimental data of the ras-ralGDS association free energy measured by ITC ⁵⁶ in aqueous solution of 50 mM Tris/HCl, 5 mM MgCl ₂ , and 100 mM NaCl	ΔG_{ITC}	-8.4 ± 0.2 kcal/mol
the Ras–RalGDS association free energy, computed by MM/GBSA, ¹ in aqueous solution of 100 mM NaCl	ΔG_{GBSA}	-19.5 ± 5.9 kcal/mol

interactions. The PME was updated every 4 fs. The damping constant was 5/ps. Explicit solvent was represented with the TIP3P model. The pulling was along the $\pm z$ -axis at a speed of $v_d = (0, 0, \pm 1.25 \text{ \AA/ns})$ so that the two partners are separated at a speed of 2.5 Å/ns in all SMD runs except one test case, when the separation speed was chosen as 1.25 Å/ns to illustrate the accuracy of the hSMD method. In all sections, four forward and four reverse pulling paths were sampled.

The all-atom model system of Ras–RalGDS complex, shown in Figure 2, was formed from the crystallographic structure (PDB code: 1LFD)⁵⁵ by taking its A and B chains, rotating it to

the orientation of RalGDS on top Ras along the z -axis, replacing GNP with GTP, putting the complex in the center of a water box of $80 \times 80 \times 160 \text{ \AA}^3$, and neutralized with 12 Na⁺ ions. During the 50 ns equilibrium MD run, the system settles down (after 3 ns) to fluctuate slightly around the dimensions of $80 \times 80 \times 147 \text{ \AA}^3$. To correspond exactly with the *in vitro* experimental conditions,⁵⁶ 50 mM Tris/HCl, 5 mM MgCl₂, and 100 mM NaCl were added to the model system.

RESULTS

Analytical Formula for Protein–Protein Association

Affinity. The derivation of the previous section is our main result, which is summarized in Tables 1–3.

For the Ras–RalGDS and barnase–barstar complexes (Supporting Information, Section V), we choose $m = n = 3$, three pulling centers on each of the two partner proteins, and we use Gaussian approximations for the fluctuations in the associated state when three centers are fixed, eq 23. Then, we have, from eq 12, the free energy of association as follows

$$\Delta G_{\text{hSMD}} = \Delta W_{0,\infty} - \Delta_{3o} + k_B T$$

$$\ln \left[\frac{Z_{3-1\infty}^{(P1)} Z_{3-1\infty}^{(P2)}}{c_0 (2\pi)^{9/2} \text{Det}^{1/2} (\sum_{3o}) Z_{3-1,0}} \right] \quad (28)$$

where three 6D partial partitions are all factored out into 1D partial partitions as shown Tables 2 and 3. $c_0 = 6.02 \times 10^{-4} / \text{\AA}^3$. Δ_{3o} and \sum_{3o} are given in eqs 24 and 25, respectively, with $k = 3$.

Numerical Results. For the Ras–RalGDS system, we choose three alpha carbons on each protein as the steering/pulling centers: Asn29, Lys32, and Lys52 on RalGDS (noted as P1, their position vectors noted as $(\mathbf{r}_1^{(P1)}, \mathbf{r}_2^{(P1)}, \mathbf{r}_3^{(P1)})$) along with Asp238, Ile221, and Lys242 on Ras (notated as P2, position vectors as $(\mathbf{r}_1^{(P2)}, \mathbf{r}_2^{(P2)}, \mathbf{r}_3^{(P2)})$). See Figures 1 and 2C for illustrations. The numerical results are tabulated in Table 4.

Note that each of the three 6D partial partitions involved in eq 28 was reduced three 1D sampling problems for which the numerical data are presented in Supporting Information Figures S1 and S2. The 9×9 matrix \sum_{3o} for fluctuations in the associated state defined in eq 25 and the deviations of the initial state from the equilibrium average state Δ_{3o} as measured in eq 24 for $k = 3$ were computed from the data shown in Supporting Information, Section III, Figure S3. The 18D PMF along the pulling path illustrated in Figure 3 was computed from the work curves shown in Supporting Information, Section IV, Figure S4 using eq 27. The PMF difference $\Delta W_{0,\infty}$ was taken between $z = 0$ and 12 \AA .

DISCUSSION

It is worth noting that the PMF difference (-18.2 kcal/mol) is very far off the computed free energy of association (-9.2 kcal/mol) in this case and for protein–protein associations in general because it is only the difference between two single states (two points in the 18D space): one state chosen from the associated state ensemble and the corresponding one state in the dissociated state ensemble. The fluctuations around these two single states are represented in the 15D partial partition (the associated conformation) and the 12D partial partition (the dissociated conformation), respectively. Shown in Figure 4 and in Supporting Information Section III are the behaviors of these fluctuations. They give equally important contributions (9.0 kcal/mol), as one would expect.

In particular, we use the crystal structures of Ras and RalGDS as the reference to compute the structural deviations of Ras and RalGDS in the associated or dissociated state that is shown in Figure 4. In the dissociated state, one center is fixed on each protein, Ras or RalGDS. In the associated state, only one center is fixed on Ras, whereas RalGDS follows stochastic dynamics in interactions with Ras. The backbone structure of the Ras–RalGDS complex does not deviate from the crystal structure by more than 1.7 \AA and not by more than 2.2 \AA when including

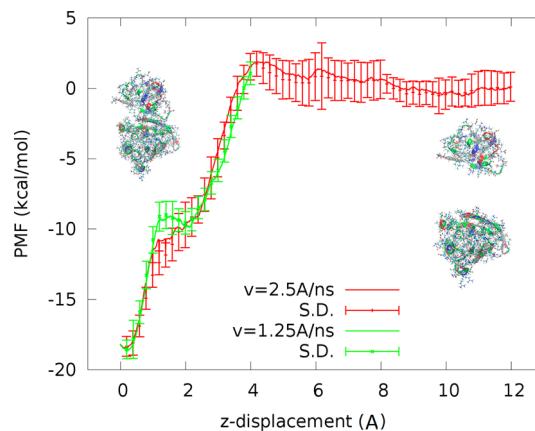


Figure 3. PMF along the pulling path from the associated state to the dissociated state. The horizontal axis indicates displacement of RalGDS relative to Ras when they are steered in the positive and negative z -directions, respectively. The second set of data were obtained at a slower pulling speed of 1.25 \AA/ns to demonstrate convergence of the numerical sampling. Note that equilibrations were performed only at $z = 0$ and 4 \AA but not in between them. Therefore, the convergence is expected only for the PMF difference between the two points, not between them.

the fluctuations of the side chains (Figure 4C), indicating the stability of Ras–RalGDS complex and the quality of the crystal structure data.⁵⁵ In the dissociated state, Ras does not deviate significantly from the crystal structure ($<1.2 \text{ \AA}$ for backbones), and its side chains do not fluctuate more than 1.9 \AA , on average. In contrast, RalGDS's deviations are as high as 2.3 \AA for backbones and 2.8 \AA for side chains. These fluctuations account for most of the association energy on top of the PMF difference.

What interactions are responsible for the Ras–RalGDS association? As shown in Figure 5, the Ras protein residues in contact with RalGDS are Lys231, Asp233, Pro234, Thr235, Ile236, Glu237, Asp238, Ser239, Tyr240, Arg241, Leu256, Tyr264, and Met267. They are in contact in the crystal structure, and they remain in close contact during the 50 ns equilibration process, indicating their relevance in the Ras–RalGDS recognition.

Along the forced separation path, the PMF does not rise monotonously (Figure 3). The peculiarity around 2 to 4 \AA reflects the importance of the following interactions: the salt bridges between Glu237 of Ras and Arg20 of RalGDS and between Asp233 of Ras and Lys52 of RalGDS along with the vdW contacts between Ile236 of Ras and (Ile18, Leu35) of RalGDS. These interactions cause conformational distortions before their separation, as shown in Figure 6. We therefore confirm that they are the main contributors to the Ras–RalGDS binding.^{1,55}

It should be helpful to note that hSMD is intended to be a brute force method that does not require much sophistication of sampling schemes yet gives estimations of standard binding free energies with chemical accuracy. The PMF is a function of state, and consequently the PMF difference between two states can be obtained via multiple paths connecting the two end states of the paths sampled. This fundamental principle enables us to compute the PMF difference accurately by sampling a few forward and the same number of reverse pulling paths connecting one chosen state in the associated state ensemble and its corresponding state in the dissociated state ensemble.

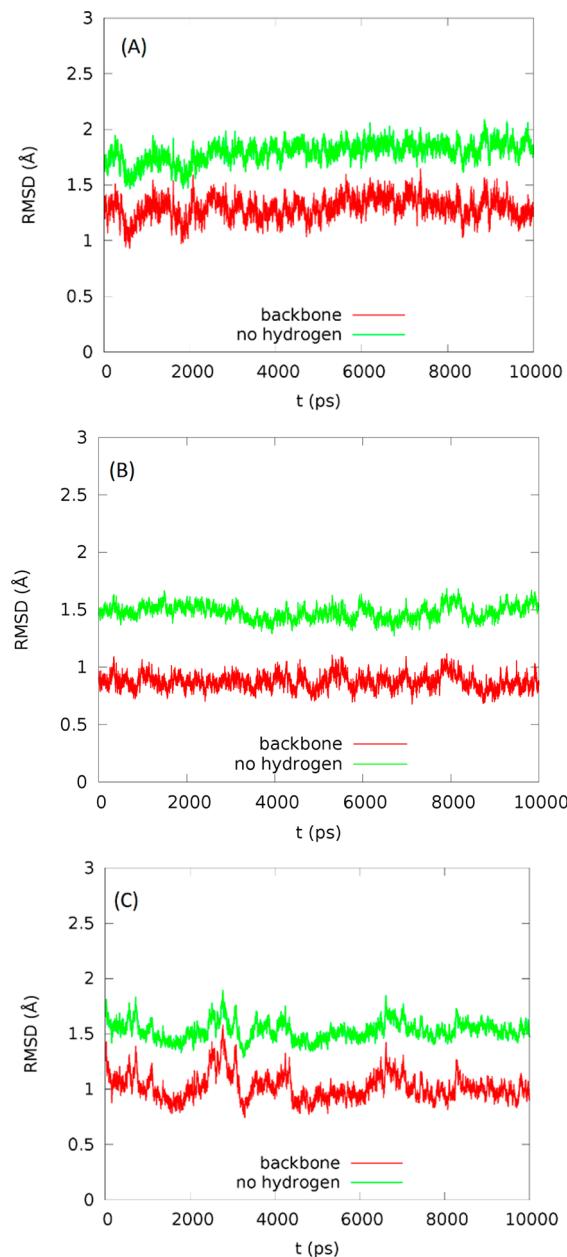


Figure 4. RMSD of the proteins from the crystallographic structure (PDB: 1LFD). (A) RalGDS in dissociated state. (B) Ras in dissociated state. (C) Ras–RalGDS complex in associated state.

Naturally, equilibrium sampling of the associated state ensemble and the same of the dissociated state ensemble must be carried out to account for all of the relevant fluctuations of the pulling centers. These $3(k+3)$ D sampling problems are resolved as $3k$ D Gaussian fluctuations of k centers when three of the pulling centers are fixed in multiplication with 6D sampling problems. Here, $k = m + n - 3$ for the associated P1–P2 complex, $k = m - 3$ for P1 and $k = n - 3$ for P2 in the dissociated state. Therefore, the choice of the pulling centers, the k centers in particular, needs to be the most stable parts of the proteins so that the fluctuations can be accurately accounted for in the Gaussian approximation. In the two protein–protein complexes studied in this work, we made the choice of three α -carbons on each protein without multiple trials. Given that $m = n = 3$ is the crudest implementation of hSMD, the agreement of our results with the experimental data

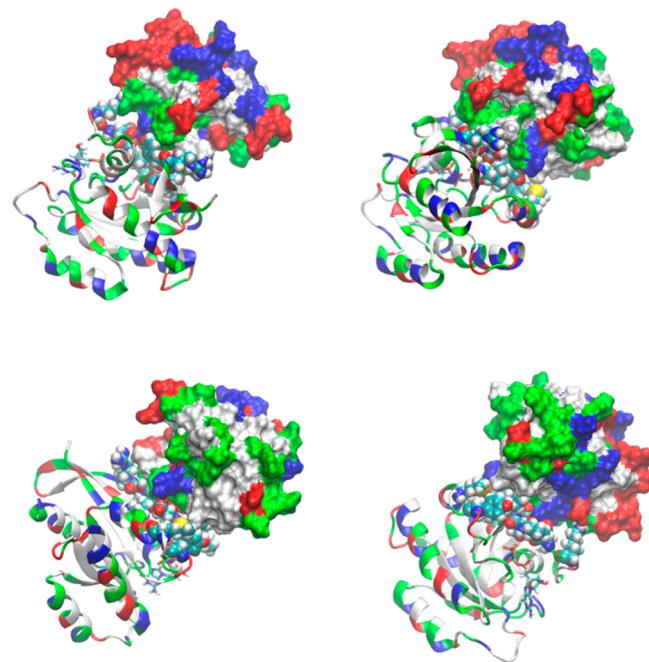


Figure 5. Proteins Ras (in ribbons, colored by residue type) and RalGDS (in surface, colored by residue type) in associated state viewed from four different angles. Also shown are the Ras residues (in spheres, colored by atom names) in contact with RalGDS. The coordinates were from the end of the 50 ns MD run of the associated state.

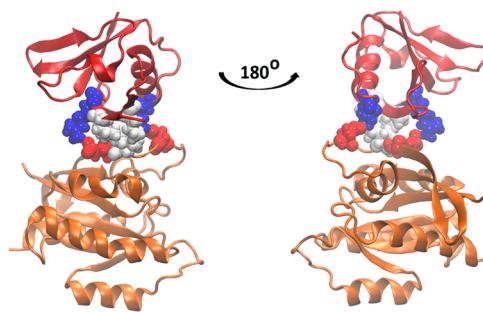


Figure 6. Residues pulled out of their ways along the SMD run. Asp233, Ile236, and Glu237 of Ras are in close contact with Lys52, (Ile18 and Leu35), and Arg20 of RalGDS, respectively. Ras (yellow) and RalGDS (red) are shown in ribbons, and the aforementioned residues are shown in spheres colored by residue type.

suggests that hSMD is not very sensitive to the choice of pulling centers. A finer choice of the same number of pulling centers will certainly lead to better accuracy without additional computing cost. Using $m > 3$ and/or $n > 3$ will involve more computing and is expected to give better accuracy in terms of statistical mechanics. All of this depends on the accuracy of the force field parameters we use to characterize the intra- and intermolecular interactions of the systems, of course.

CONCLUSIONS

We have developed a hybrid steered molecular dynamics approach for computing the absolute free energy of protein–protein association from the PMF along a dissociation path. Applying this hSMD approach with high-performance parallel processing, one can obtain, within a few wall-clock days, the association affinity of one protein–protein complex with

accuracy comparable with experimental measurements. Using this brute force approach, one does not have to delicately devise biasing and constraining potentials during the course of simulations. One simply steers/pulls the $m + n$ centers of mass of m (≥ 3) chosen segments of one protein and n (≥ 3) chosen segments of the partner protein by using $m + n$ infinitely stiff springs along one predetermined path of separation, disallowing any deviations from the path. This use of a single path correctly gives the $3(m + n)$ D PMF difference between one state in the ensemble of the associated states and its counterpart in the ensemble of the dissociated states, noting that the PMF is a function of state and thus the PMF difference between two states is independent of the paths connecting them. All other contributions, in addition to the PMF difference between the two end states of this one dissociation path, are rigorously accounted for in the $3(m + n - 1)$ D partial partition function of the associated state and two partial partitions of the dissociated states that are $3(m - 1)$ D and $3(n - 1)$ D, respectively. Each of the three partial partitions can be well-approximated as the product of a 6D partial partition and the Gaussian fluctuations when three of the centers are fixed, and the 6D partial partition can be factored into three 1D sampling problems. The total computing cost for our hSMD study of the Ras–RalGDS complex was approximately 3 wall-clock days using 400 computing cores (20 nodes) on the TACC supercomputer Maverick (no GPU usage) or, based on a prorate estimate, 12 wall-clock days using one node (20 cores and one nVidia K40) with GPU usage. We hereby assert that the hSMD approach is efficient for protein–protein binding affinities given that today’s high-performance computing power affords us a proportionally large number of computing cores needed for a given protein–protein system.

Studying the Ras–RalGDS complex, we have provided atomistic details in support of the binding mechanisms elucidated in the experimental and theoretical investigations in the current literature. Our equilibrium MD simulations confirm the experimentally determined binding conformations. During the long time dynamics, the protomers were found to fluctuate around the crystal structure coordinates with deviations less than 2 Å. In addition, our computed association energy agrees with the experimentally measured value within the margin of error. We expect that the hSMD method is usable to accurately predict association affinities of other protein–protein complexes.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jctc.5b00340](https://doi.org/10.1021/acs.jctc.5b00340).

Step-by-step derivation for the partial partitions (Sections I and II); data produced in MD and SMD runs used to compute the association energy of Ras–RalGDS (Figures S1–S4, Sections III and IV); and study of a second protein–protein complex (Section V) (PDF).

AUTHOR INFORMATION

Corresponding Author

*Phone: (210) 458-5457. Fax: (210) 458-4919. E-mail: Liao.Chen@utsa.edu.

Author Contributions

[†]R.A.R. and L.Y. contributed equally to this work.

Funding

The authors acknowledge support from the NIH (GM 084834) and the computing resources provided by the Texas Advanced Computing Center at University of Texas at Austin.

Notes

The authors declare no competing financial interest.

ABBREVIATIONS

1D, one-dimensional; m D, m -dimensional; GBSA, generalized Born surface area; GNP, phosphoaminophosphonic acid guanylate ester; GTP, guanosine-5'-triphosphate; hSMD, hybrid steered molecular dynamics; ITC, isothermal titration calorimetry; k_D , dissociation constant; MD, molecular dynamics; PME, particle mesh Ewald; PMF, potential of mean force; RalGDS, Ral guanine nucleotide dissociation stimulator; RMSD, root-mean-square deviation; SMD, steered molecular dynamics; vdW, van der Waals

REFERENCES

- (1) Gohlke, H.; Kiel, C.; Case, D. A. Insights into Protein–Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes. *J. Mol. Biol.* **2003**, *330*, 891–913.
- (2) Woo, H.-J.; Roux, B. Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6825–6830.
- (3) Ytreberg, F. M.; Swendsen, R. H.; Zuckerman, D. M. Comparison of free energy methods for molecular systems. *J. Chem. Phys.* **2006**, *125*, 184114.
- (4) Mobley, D. L.; Dill, K. A. Binding of Small-Molecule Ligands to Proteins: “What You See” Is Not Always “What You Get. *Structure* **2009**, *17*, 489–498.
- (5) Zhou, H.-X.; Gilson, M. K. Theory of Free Energy and Entropy in Noncovalent Binding. *Chem. Rev.* **2009**, *109*, 4092–4107.
- (6) General, I. J. A Note on the Standard State’s Binding Free Energy. *J. Chem. Theory Comput.* **2010**, *6*, 2520–2524.
- (7) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, *51*, 69–82.
- (8) Cai, L.; Zhou, H.-X. Theory and simulation on the kinetics of protein–ligand binding coupled to conformational change. *J. Chem. Phys.* **2011**, *134*, 105101.
- (9) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150–160.
- (10) Gallicchio, E.; Levy, R. M. Recent theoretical and computational advances for modeling protein–ligand binding affinities. In *Advances in Protein Chemistry and Structural Biology*; Christo, C., Ed.; Academic Press: Oxford, 2011; Vol. 85, pp 27–80.
- (11) General, I. J.; Dragomirova, R.; Meirovitch, H. Absolute Free Energy of Binding of Avidin/Biotin, Revisited. *J. Phys. Chem. B* **2012**, *116*, 6628–6636.
- (12) Wu, X.; Damjanovic, A.; Brooks, B. R. Efficient and Unbiased Sampling of Biomolecular Systems in the Canonical Ensemble: A Review of Self-Guided Langevin Dynamics. In *Advances in Chemical Physics*; John Wiley & Sons, Inc.: Hoboken, NJ, 2012; pp 255–326.
- (13) Ahmed, S. M.; Kruger, H. G.; Govender, T.; Maguire, G. E. M.; Sayed, Y.; Ibrahim, M. A. A.; Naicker, P.; Soliman, M. E. S. Comparison of the Molecular Dynamics and Calculated Binding Free Energies for Nine FDA-Approved HIV-1 PR Drugs Against Subtype B and C-SA HIV PR. *Chem. Biol. Drug Des.* **2013**, *81*, 208–218.
- (14) Buch, I.; Ferruz, N.; De Fabritiis, G. Computational Modeling of an Epidermal Growth Factor Receptor Single-Mutation Resistance to Cetuximab in Colorectal Cancer Treatment. *J. Chem. Inf. Model.* **2013**, *53*, 3123–3126.

- (15) Gumbart, J. C.; Roux, B.; Chipot, C. Efficient Determination of Protein–Protein Standard Binding Free Energies from First Principles. *J. Chem. Theory Comput.* **2013**, *9*, 3789–3798.
- (16) Gumbart, J. C.; Roux, B.; Chipot, C. Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? *J. Chem. Theory Comput.* **2013**, *9*, 794–802.
- (17) Sun, H.; Li, Y.; Li, D.; Hou, T. Insight into Crizotinib Resistance Mechanisms Caused by Three Mutations in ALK Tyrosine Kinase using Free Energy Calculation Approaches. *J. Chem. Inf. Model.* **2013**, *53*, 2376–2389.
- (18) Takahashi, R.; Gil, V. A.; Guallar, V. Monte Carlo Free Ligand Diffusion with Markov State Model Analysis and Absolute Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2014**, *10*, 282–288.
- (19) Wu, M.; Strid, Å.; Eriksson, L. A. Interactions and Stabilities of the UV RESISTANCE LOCUS8 (UVR8) Protein Dimer and Its Key Mutants. *J. Chem. Inf. Model.* **2013**, *53*, 1736–1746.
- (20) Hansen, N.; van Gunsteren, W. F. Practical Aspects of Free-Energy Calculations: A Review. *J. Chem. Theory Comput.* **2014**, *10*, 2632–2647.
- (21) Heinzemann, G.; Chen, P.-C.; Kuyucak, S. Computation of Standard Binding Free Energies of Polar and Charged Ligands to the Glutamate Receptor GluA2. *J. Phys. Chem. B* **2014**, *118*, 1813–1824.
- (22) Zeller, F.; Zacharias, M. Evaluation of Generalized Born Model Accuracy for Absolute Binding Free Energy Calculations. *J. Phys. Chem. B* **2014**, *118*, 7467–7474.
- (23) 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*, 909–918.
- (24) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300–313.
- (25) Chandler, D. Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *J. Chem. Phys.* **1978**, *68*, 2959–2970.
- (26) Pratt, L. R.; Hummer, G.; Garcíá, A. E. Ion pair potentials-of-mean-force in water. *Biophys. Chem.* **1994**, *51*, 147–165.
- (27) Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **1995**, *91*, 275–282.
- (28) Allen, T. W.; Andersen, O. S.; Roux, B. Molecular dynamics—potential of mean force calculations as a tool for understanding ion permeation and selectivity in narrow channels. *Biophys. Chem.* **2006**, *124*, 251–267.
- (29) Isralewitz, B.; Izrailev, S.; Schulten, K. Binding pathway of retinal to bacterio-opsin: A prediction by molecular dynamics simulations. *Biophys. J.* **1997**, *73*, 2972–2979.
- (30) Gullingsrud, J. R.; Braun, R.; Schulten, K. J. Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations. *Biophys. J.* **1999**, *76*, A200–A200.
- (31) Jensen, M. O.; Park, S.; Tajkhorshid, E.; Schulten, K. Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 6731–6736.
- (32) Tajkhorshid, E.; Aksimentiev, A.; Balabin, I.; Gao, M.; Isralewitz, B.; Phillips, J. C.; Zhu, F. Q.; Schulten, K. Large scale simulation of protein mechanics and function. *Adv. Protein Chem.* **2003**, *66*, 195–247.
- (33) Li, P. C.; Makarov, D. E. Simulation of the mechanical unfolding of ubiquitin: Probing different unfolding reaction coordinates by changing the pulling geometry. *J. Chem. Phys.* **2004**, *121*, 4826–4832.
- (34) Park, S.; Schulten, K. Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.* **2004**, *120*, 5946–5961.
- (35) Jensen, M. O.; Yin, Y.; Tajkhorshid, E.; Schulten, K. Sugar transport across lactose permease probed by steered molecular dynamics. *Biophys. J.* **2007**, *93*, 92–102.
- (36) Wang, Y.; Zhang, L. X. Steered molecular dynamics simulation of elastic behavior of adsorbed single polyethylene chains. *J. Polym. Sci., Part B: Polym. Phys.* **2007**, *45*, 2322–2332.
- (37) Minh, D. D. L.; McCammon, J. A. Springs and speeds in free energy reconstruction from irreversible single-molecule pulling experiments. *J. Phys. Chem. B* **2008**, *112*, 5892–5897.
- (38) Chen, L. Y. Exploring the free-energy landscapes of biological systems with steered molecular dynamics. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6176–6183.
- (39) Trinh, T.; Lastoskie, C.; Epstein, T.; Philbert, M. Steered Molecular Dynamics Simulation of Kinesin Detachment from the Microtubule Surface. *Biophys. J.* **2011**, *100*, 194–195.
- (40) Fukunishi, H.; Shimada, J.; Shiraiishi, K. Antigen-Antibody Interactions and Structural Flexibility of a Femtomolar-Affinity Antibody. *Biochemistry* **2012**, *51*, 2597–2605.
- (41) Moradi, M.; Tajkhorshid, E. Driven Metadynamics: Reconstructing Equilibrium Free Energies from Driven Adaptive-Bias Simulations. *J. Phys. Chem. Lett.* **2013**, *4*, 1882–1887.
- (42) Nicolini, P.; Frezzato, D.; Gellini, C.; Bizzarri, M.; Chelli, R. Toward quantitative estimates of binding affinities for protein-ligand systems involving large inhibitor compounds: A steered molecular dynamics simulation route. *J. Comput. Chem.* **2013**, *34*, 1561–1576.
- (43) Trinh, T. D.; Philbert, M. A.; Lastoskie, C. M. Steered Molecular Dynamics Simulation of Kinesin Detachment from the Microtubule Surface and the Effect of 1,3-Dinitrobenzene. *Biophys. J.* **2013**, *104*, 332a–333a.
- (44) Velez-Vega, C.; Gilson, M. K. Overcoming Dissipation in the Calculation of Standard Binding Free Energies by Ligand Extraction. *J. Comput. Chem.* **2013**, *34*, 2360–2371.
- (45) Yu, T.; Lee, O. S.; Schatz, G. C. Steered Molecular Dynamics Studies of the Potential of Mean Force for Peptide Amphiphile Self-Assembly into Cylindrical Nanofibers. *J. Phys. Chem. A* **2013**, *117*, 7453–7460.
- (46) Baştug, T.; Chen, P.-C.; Patra, S. M.; Kuyucak, S. Potential of mean force calculations of ligand binding to ion channels from Jarzynski's equality and umbrella sampling. *J. Chem. Phys.* **2008**, *128*, 155104.
- (47) Jeřábek, P.; Florián, J.; Stiborová, M.; Martínek, V. Flexible Docking-Based Molecular Dynamics/Steered Molecular Dynamics Calculations of Protein–Protein Contacts in a Complex of Cytochrome P450 1A2 with Cytochrome b5. *Biochemistry* **2014**, *53*, 6695–6705.
- (48) Chen, L. Y. Hybrid Steered Molecular Dynamics Approach to Computing Absolute Binding Free Energy of Ligand-Protein Complexes: A Brute Force Approach That Is Fast and Accurate. *J. Chem. Theory Comput.* **2015**, *11*, 1928–1938.
- (49) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (50) Chen, L. Y. Glycerol modulates water permeation through Escherichia coli aquaglyceroporin GlpF. *Biochim. Biophys. Acta, Biomembr.* **2013**, *1828*, 1786–1793.
- (51) Chen, L. Y. Nonequilibrium fluctuation-dissipation theorem of Brownian dynamics. *J. Chem. Phys.* **2008**, *129*, 144113–144114.
- (52) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (53) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671–690.
- (54) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

- (55) Huang, L.; Hofer, F.; Martin, G. S.; Kim, S.-H. Structural basis for the interaction of Ras with RaIGDS. *Nat. Struct. Biol.* **1998**, *5*, 422–426.
- (56) Rudolph, M. G.; Linnemann, T.; Grünwald, P.; Wittinghofer, A.; Vetter, I. R.; Herrmann, C. Thermodynamics of Ras/Effecter and Cdc42/Effecter Interactions Probed by Isothermal Titration Calorimetry. *J. Biol. Chem.* **2001**, *276*, 23914–23921.