Supplementary Information to

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 USA

In the first two sections (I and II) of this SI, we present the step-by-step derivations of the partial partition functions in the associated and the dissociated states that are factors in the formula of the protein-protein association affinity.

In the next two sections (III and IV) of this SI, we present the fluctuations of the pulling centers in the associated and the dissociated states along with the work curves along the pulling paths that were used to compute the Ras-RalGDS affinity.

In the last section (V) of this SI, we present the study of the Barnase-Barstar complex.

I. Computation of the partial partitions in the dissociated state.

In the dissociated state, the stochastic dynamics of the two protomers are independent of one another. The computation of $Z_{m-1\infty}^{(\mathrm{P1})}$ will be no different from that of $Z_{n-1\infty}^{(\mathrm{P2})}$. Therefore, in this subsection, we drop the superscript (P1) or (P2) until when we give the final results of $Z_{m-1\infty}^{(\mathrm{P1})}$ and $Z_{n-1\infty}^{(\mathrm{P2})}$. Note that we consider m,n≥3. We need to evaluate the integral of Eq. (11) around the position $(\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, \mathbf{r}_{3\infty}, \mathbf{r}_{4\infty}, ..., \mathbf{r}_{m\infty})$ when P1 and P2 are far away from one another. Inserting an intermediate step, one can factor the 3(m-1)-D partial partition in Eq. (11) into two partial partitions of lower dimensions,

$$
\begin{aligned}
Z_{m-1\infty} &= \int_{\text{bulk}} \prod_{i=2}^{m} d^3x_i \, \exp\Big[-\Big(W[\mathbf{r}_{1\infty}, \mathbf{r}_2, ..., \mathbf{r}_m] - W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, ..., \mathbf{r}_{m\infty}]\Big)\Big/k_BT\Big] \\
&= \frac{\int_{\text{bulk}} \prod_{i=2}^{m} d^3x_i \, \exp\Big[-W[\mathbf{r}_{1\infty}, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4, ..., \mathbf{r}_m]/k_BT\Big]}{\int_{\text{bulk}} \prod_{i=4}^{m} d^3x_i \, \exp\Big[-W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, \mathbf{r}_{3\infty}, \mathbf{r}_4, ..., \mathbf{r}_m]/k_BT\Big]} \times \\
&\quad \frac{\int_{\text{bulk}} \prod_{i=4}^{m} d^3x_i \, \exp\Big[-W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, \mathbf{r}_{3\infty}, \mathbf{r}_4, ..., \mathbf{r}_m]/k_BT\Big]}{\exp\Big[-W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, \mathbf{r}_{3\infty}, \mathbf{r}_{4\infty}, ..., \mathbf{r}_{m\infty}]/k_BT\Big]} \\
&= Z_{3-1\infty} Z_{m-3\infty}.
\end{aligned}
\tag{S1}
$$

The 3(m-3)-D partial partition

$$
Z_{m-3\infty} = \frac{\int_{\text{bulk}} \prod_{i=4}^{m} d^3x_i \, \exp\Big[-W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, \mathbf{r}_{3\infty}, \mathbf{r}_4, ..., \mathbf{r}_m]/k_BT\Big]}{\exp\Big[-W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, \mathbf{r}_{3\infty}, \mathbf{r}_{4\infty}, ..., \mathbf{r}_{m\infty}]/k_BT\Big]} \tag{S2}
$$

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 the stochastic dynamics. The fluctuations of the other (m-3) centers in Eq. (S2) can be readily sampled to give

$$
Z_{m-3\infty} = (2\pi)^{3(m-3)/2} Det^{1/2}(\Sigma_{m-3\infty}) \exp\big[\Delta_{m-3\infty}/k_BT\big]. \tag{S3}
$$

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

$$
\Delta_{m-3\infty}/k_BT = \tfrac{1}{2}\Big(\langle \mathbf{r}_4 \rangle - \mathbf{r}_{4\infty}, \langle \mathbf{r}_5 \rangle - \mathbf{r}_{5\infty}, \cdots, \langle \mathbf{r}_m \rangle - \mathbf{r}_{m\infty}\Big) \Sigma_{m-3\infty}^{-1} \Big(\langle \mathbf{r}_4 \rangle - \mathbf{r}_{4\infty}, \langle \mathbf{r}_5 \rangle - \mathbf{r}_{5\infty}, \cdots, \langle \mathbf{r}_m \rangle - \mathbf{r}_{m\infty}\Big)^T. \tag{S4}
$$

Det represents the determinant. $\Sigma_{m-3\infty}$ is the 3(m-3)×3(m-3) matrix of the fluctuations/deviations of the pulling center coordinates $\delta x_4 = x_4 - \langle x_4 \rangle$ *etc.* in the bound state ensemble,

$$\Sigma_{m-3\infty} = \left\langle \left(\left\langle \mathbf{r}_4 \right\rangle - \mathbf{r}_{4\infty}, \left\langle \mathbf{r}_5 \right\rangle - \mathbf{r}_{5\infty}, \cdots, \left\langle \mathbf{r}_m \right\rangle - \mathbf{r}_{m\infty} \right)^T \left(\left\langle \mathbf{r}_4 \right\rangle - \mathbf{r}_{4\infty}, \left\langle \mathbf{r}_5 \right\rangle - \mathbf{r}_{5\infty}, \cdots, \left\langle \mathbf{r}_m \right\rangle - \mathbf{r}_{m\infty} \right) \right\rangle$$

$$= \begin{pmatrix} \left\langle \delta x_4^{\,2} \right\rangle & \left\langle \delta x_4 \delta y_4 \right\rangle & \left\langle \delta x_4 \delta z_4 \right\rangle & \cdots & \left\langle \delta x_4 \delta x_m \right\rangle & \left\langle \delta x_4 \delta y_m \right\rangle & \left\langle \delta x_4 \delta z_m \right\rangle \\ \left\langle \delta y_4 \delta x_4 \right\rangle & \left\langle \delta y_4^{\,2} \right\rangle & \left\langle \delta y_4 \delta z_4 \right\rangle & \cdots & \left\langle \delta y_4 \delta x_m \right\rangle & \left\langle \delta y_4 \delta y_m \right\rangle & \left\langle \delta y_4 \delta z_m \right\rangle \\ \left\langle \delta z_4 \delta x_4 \right\rangle & \left\langle \delta z_4 \delta y_4 \right\rangle & \left\langle \delta z_4^{\,2} \right\rangle & \cdots & \left\langle \delta z_4 \delta x_m \right\rangle & \left\langle \delta z_4 \delta y_m \right\rangle & \left\langle \delta z_4 \delta z_m \right\rangle \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \left\langle \delta x_m \delta x_4 \right\rangle & \left\langle \delta x_m \delta y_4 \right\rangle & \left\langle \delta x_m \delta z_4 \right\rangle & \cdots & \left\langle \delta x_m^{\,2} \right\rangle & \left\langle \delta x_m \delta y_m \right\rangle & \left\langle \delta x_m \delta z_m \right\rangle \\ \left\langle \delta y_m \delta x_4 \right\rangle & \left\langle \delta y_m \delta y_4 \right\rangle & \left\langle \delta y_m \delta z_4 \right\rangle & \cdots & \left\langle \delta y_m \delta x_m \right\rangle & \left\langle \delta y_m^{\,2} \right\rangle & \left\langle \delta y_m \delta z_m \right\rangle \\ \left\langle \delta z_m \delta x_4 \right\rangle & \left\langle \delta z_m \delta y_4 \right\rangle & \left\langle \delta z_m \delta z_4 \right\rangle & \cdots & \left\langle \delta z_m \delta x_m \right\rangle & \left\langle \delta z_m \delta y_m \right\rangle & \left\langle \delta z_m^{\,2} \right\rangle \end{pmatrix}. \quad \text{(S5)}$$

$\Sigma_{m-3\infty}^{-1}$ is the inverse matrix of $\Sigma_{m-3\infty}$ which can be accurately evaluated by running equilibrium MD in the dissociated state of a protomer while three of m centers are fixed at $(\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, \mathbf{r}_{3\infty})$. Putting the protomer label P1 and P2 back explicitly, the 3(m-3)-D partial partition of P1 and the 3(n-3)-D partial partition of P2 are

$$\begin{aligned} Z_{m-3\infty}^{(\mathrm{P1})} &= (2\pi)^{3(m-3)/2} Det^{1/2}(\Sigma_{m-3\infty}^{(\mathrm{P1})}) \exp\left[\Delta_{m-3\infty}^{(\mathrm{P1})} / k_B T \right], \\ Z_{n-3\infty}^{(\mathrm{P2})} &= (2\pi)^{3(\mathrm{n}-3)/2} Det^{1/2}(\Sigma_{n-3\infty}^{(\mathrm{P2})}) \exp\left[\Delta_{n-3\infty}^{(\mathrm{P2})} / k_B T \right]. \end{aligned} \quad \text{(S6)}$$

The 6-D partial partition

$$Z_{3-1\infty} = \frac{\int_{\mathrm{bulk}} \prod_{i=2}^{3} d^3 x_i \, \exp\left[-W[\mathbf{r}_{1\infty}, \mathbf{r}_2, \mathbf{r}_3] / k_B T \right]}{\exp\left[-W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, \mathbf{r}_{3\infty}] / k_B T \right]} \quad \text{(S7)}$$

contains the fluctuations of the second and the third centers of a protomer in the dissociated state. Over there, the protomer's environment is spherically symmetrical around the position of the first pulling center $\mathbf{r}_{1\infty}$ that is fixed while the second and the third pulling centers $\mathbf{r}_2$ and $\mathbf{r}_3$ are free to sample all space available. The PMF $W[\mathbf{r}_{1\infty}, \mathbf{r}_2, \mathbf{r}_3]$ is only dependent upon three of the six degrees of freedom contained in $(\mathbf{r}_2, \mathbf{r}_3)$. Therefore, the computation of a protomer's partial partition in the dissociated state is a 3D

sampling problem. If the equilibrium sampling in 3D cannot be achieved with sufficient accuracy, one can insert an intermediate step of sampling fluctuations of the third center $\mathbf{r}_3$ while fixing the first and the second centers at $(\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty})$. Namely,

$$\begin{aligned} Z_{3-1\infty} &= \int_{\text{bulk}} \prod_{i=2}^{3} d^3x_i \, \exp\left[-\left(W[\mathbf{r}_{1\infty}, \mathbf{r}_2, \mathbf{r}_3] - W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, \mathbf{r}_{3\infty}]\right)/k_BT\right] \\ &= \frac{\int_{\text{bulk}} \prod_{i=2}^{3} d^3x_i \, \exp\left[-W[\mathbf{r}_{1\infty}, \mathbf{r}_2, \mathbf{r}_3]/k_BT\right]}{\int_{\text{bulk}} d^3x_3 \, \exp\left[-W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, \mathbf{r}_3]/k_BT\right]} \frac{\int_{\text{bulk}} d^3x_3 \, \exp\left[-W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, \mathbf{r}_3]/k_BT\right]}{\exp\left[-W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, \mathbf{r}_{3\infty}]/k_BT\right]} \\ &= Z_{2-1\infty} Z_{3-2\infty}, \end{aligned} \tag{S8}$$

$$Z_{2-1\infty} = \frac{\int_{\text{bulk}} d^3x_2 \, \exp\left[-W[\mathbf{r}_{1\infty}, \mathbf{r}_2]/k_BT\right]}{\exp\left[-W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}]/k_BT\right]},$$

$$Z_{3-2\infty} = \frac{\int_{\text{bulk}} d^3x_3 \, \exp\left[-W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, \mathbf{r}_3]/k_BT\right]}{\exp\left[-W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, \mathbf{r}_{3\infty}]/k_BT\right]},$$

where the 6D partial partition of a protomer in the dissociated state factors into two 3D partial partitions, $Z_{2-1\infty}$ that can be computed as a 1D sampling problem and $Z_{3-2\infty}$ that can be reduced to two 1D sampling problems.

The 3D partial partition $Z_{2-1\infty}$ contains the fluctuations of a protomer when its first center is fixed,

$$Z_{2-1\infty} = \int_{\text{bulk}} d^3x_2 \, \exp\left[-\left(W[\mathbf{r}_{1\infty}, \mathbf{r}_2] - W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}]\right)/k_BT\right], \tag{S9}$$

that is, an integration over the second steering center when the first steering center is fixed in a position $\mathbf{r}_{1\infty}$. All one needs to do now is to evaluate the integral of Eq. (S9) around the position $(\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty})$. Over there, far away from the other protomer, the environment is spherically symmetrical around the position of the first pulling center $\mathbf{r}_{1\infty}$ that is fixed while the second pulling center $\mathbf{r}_2$ is free to sample all space available. Therefore, the 3D integral becomes the following 1D integral,

$$
\begin{aligned}
Z_{2-1\infty} &= 4\pi \int dr_{21} \exp\left[-\left(W_\infty[r_{21}] - W_\infty[r_{21\infty}]\right)/k_B T\right] r_{21}^2 \\
&= 4\pi r_{21\infty}^2 \big/ \rho_\infty\left(r_{21\infty}\right)
\end{aligned}
\tag{S10}
$$

where $r_{21} = \left|\mathbf{r}_2 - \mathbf{r}_{1\infty}\right|$ is the distance between the two pulling centers.

$$
\rho_\infty\left(r_{21}\right) = r_{21}^2 \exp\left[-W_\infty[r_{21}]/k_B T\right] \Big/ \int dr'_{21} r'^2_{21} \exp\left[-W_\infty[r'_{21}]/k_B T\right] \tag{S11}
$$

is the 1-D probability distribution density. $W_\infty[r_{21}]$ here, as a function of r_{21}, is the PMF for stretching the protomer between the two pulling centers. It can be evaluated by conducting SMD runs of 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 $(\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty})$. This constitutes a 1D sampling problem that can be implemented efficiently to produce accurate results.

The other 3D partial partition function $Z_{3-2\infty}$ of a protomer takes account of the fluctuations of the third center when the first and the second centers are fixed,

$$
\begin{aligned}
Z_{3-2\infty} &= \int_{\text{bulk}} d^3 x_3 \exp\left[-\left(W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, \mathbf{r}_3] - W[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, \mathbf{r}_{3\infty}]\right)\big/ k_B T\right] \\
&= 2\pi \int dr_{31} r_{31}^2 \int_0^\pi d\theta \sin\theta \exp\left[-\left(\mathrm{W}[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, r_{31}, \theta] - \mathrm{W}[\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty}, r_{31\infty}, \theta_\infty]\right)/k_B T\right].
\end{aligned}
\tag{S12}
$$

Here the angular integration around the axis connecting the first and the second centers $(\mathbf{r}_{1\infty}, \mathbf{r}_{2\infty})$ was carried out explicitly, reducing the 3D partial partition into a 2D sampling problem. $r_{31} = \left|\mathbf{r}_3 - \mathbf{r}_{1\infty}\right|$ is the distance between the third pulling center and the first pulling center that is fixed at $\mathbf{r}_{1\infty}$. θ is the angle between $\left(\mathbf{r}_3 - \mathbf{r}_{1\infty}\right)$ and $\left(\mathbf{r}_{2\infty} - \mathbf{r}_{1\infty}\right)$. θ_∞ is the angle between $\left(\mathbf{r}_{2\infty} - \mathbf{r}_{1\infty}\right)$ and $\left(\mathbf{r}_{3\infty} - \mathbf{r}_{1\infty}\right)$, of course. And this 2D sampling problem, if necessary, can be reduced to two 1D sampling problems as follows:

$$
\begin{aligned}
Z_{3-2\infty} &= 2\pi\int dr_{31}r_{31}^2\int_0^\pi d\theta\sin\theta\exp\Big[-\Big(\mathrm{W}[\mathbf{r}_{1\infty},\mathbf{r}_{2\infty},r_{31},\theta]-\mathrm{W}[\mathbf{r}_{1\infty},\mathbf{r}_{2\infty},r_{31\infty},\theta_\infty]\Big)/k_BT\Big] \\
&= \frac{2\pi\int dr_{31}r_{31}^2\int_0^\pi d\theta\sin\theta\exp\Big[-\mathrm{W}[\mathbf{r}_{1\infty},\mathbf{r}_{2\infty},r_{31},\theta]/k_BT\Big]}{\int dr_{31}r_{31}^2\exp\Big[-\mathrm{W}[\mathbf{r}_{1\infty},\mathbf{r}_{2\infty},r_{31},\theta_\infty]/k_BT\Big]}\frac{\int dr_{31}r_{31}^2\exp\Big[-\mathrm{W}[\mathbf{r}_{1\infty},\mathbf{r}_{2\infty},r_{31},\theta_\infty]/k_BT\Big]}{\exp\Big[-\mathrm{W}[\mathbf{r}_{1\infty},\mathbf{r}_{2\infty},r_{31\infty},\theta_\infty]/k_BT\Big]} \\
&= \frac{2\pi\int_0^\pi d\theta\sin\theta\exp\Big[-\mathrm{W}[\mathbf{r}_{1\infty},\mathbf{r}_{2\infty},\theta]/k_BT\Big]}{\exp\Big[-\mathrm{W}[\mathbf{r}_{1\infty},\mathbf{r}_{2\infty},\theta_\infty]/k_BT\Big]}\frac{\int dr_{31}r_{31}^2\exp\Big[-\mathrm{W}[\mathbf{r}_{1\infty},\mathbf{r}_{2\infty},r_{31},\theta_\infty]/k_BT\Big]}{\exp\Big[-\mathrm{W}[\mathbf{r}_{1\infty},\mathbf{r}_{2\infty},r_{31\infty},\theta_\infty]/k_BT\Big]} \\
&= 2\pi\int dr_{31}\exp\Big[-\big(W_\infty[r_{31}]-W_\infty[r_{31\infty}]\big)/k_BT\Big]r_{31}^2\int_0^\pi d\theta\sin\theta\exp\Big[-\big(\mathrm{W}_\infty[\theta]-\mathrm{W}_\infty[\theta_\infty]\big)/k_BT\Big] \\
&= \frac{2\pi r_{31\infty}^2\sin\theta_\infty}{\rho_\infty\big(r_{31\infty}\big)\rho_\infty\big(\theta_\infty\big)}
\end{aligned}
\tag{S13}
$$

where $\rho_\infty(\theta)$ is the angular distribution probability density for the third center when the first and the second centers are fixed at $(\mathbf{r}_{1\infty},\mathbf{r}_{2\infty})$. $\rho_\infty(r_{31})$ is the radial distribution probability density.

$$
\begin{aligned}
\rho_\infty(\theta) &= \sin\theta\exp\big[-\mathrm{W}_\infty[\theta]/k_BT\big]\Big/\int_0^\pi d\theta'\sin\theta'\exp\big[-\mathrm{W}_\infty[\theta']/k_BT\big], \\
\rho_\infty(r_{31}) &= r_{31}^2\exp\big[-W_\infty[r_{31}]/k_BT\big]\Big/\int dr'_{31}r'^2_{31}\exp\big[-W_\infty[r'_{31}]/k_BT\big].
\end{aligned}
\tag{S14}
$$

$W_\infty[r_{31}]$ is a function of r_{31} when, again, the first and the second centers are fixed at $(\mathbf{r}_{1\infty},\mathbf{r}_{2\infty})$. It can be evaluated by conducting SMD runs of steering the third pulling center $\mathbf{r}_3$ to and from the first pulling center that is fixed at $\mathbf{r}_{1\infty}$ along the axis passing through $(\mathbf{r}_{1\infty},\mathbf{r}_{3\infty})$ when the second pulling center is fixed as well. $W_\infty[\theta]$ is the angular PMF. These two 1-D PMFs can also be sampled in equilibrium MD runs.

Putting the protomer labels P1 or P2 back explicitly, one has the 6D partial partition of the two protomers P1 and P2,

$$Z_{3-1\infty}^{(\mathrm{P1})}=\frac{8\pi^2\left[r_{21\infty}^{(\mathrm{P1})}\right]^2\left[r_{31\infty}^{(\mathrm{P1})}\right]^2\sin\theta_{\infty}^{(\mathrm{P1})}}{\rho_{\infty}\left(r_{21\infty}^{(\mathrm{P1})}\right)\rho_{\infty}\left(r_{31\infty}^{(\mathrm{P1})}\right)\rho_{\infty}\left(\theta_{\infty}^{(\mathrm{P1})}\right)},\quad Z_{3-1\infty}^{(\mathrm{P2})}=\frac{8\pi^2\left[r_{21\infty}^{(\mathrm{P2})}\right]^2\left[r_{31\infty}^{(\mathrm{P2})}\right]^2\sin\theta_{\infty}^{(\mathrm{P2})}}{\rho_{\infty}\left(r_{21\infty}^{(\mathrm{P2})}\right)\rho_{\infty}\left(r_{31\infty}^{(\mathrm{P2})}\right)\rho_{\infty}\left(\theta_{\infty}^{(\mathrm{P2})}\right)}. \tag{S15}$$

To sum up this subsection, now we have the partial partitions in the dissociated state of the protomers,

$$Z_{m-1\infty}^{(\mathrm{P1})}=Z_{3-1\infty}^{(\mathrm{P1})}Z_{m-3\infty}^{(\mathrm{P1})},\quad Z_{n-1\infty}^{(\mathrm{P2})}=Z_{3-1\infty}^{(\mathrm{P2})}Z_{n-3\infty}^{(\mathrm{P2})}, \tag{S16}$$

with the factors given in Eq. (S6) (m>3 and/or n>3) and Eq. (S15) that can be computed by equilibrium MD sampling or 1D PMFs. Naturally, when m=3, $Z_{3-3\infty}^{(\mathrm{P1})}=1$ and, when n=3, $Z_{3-3\infty}^{(\mathrm{P2})}=1$.

II. Computation of the partial partition 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 $\left(\mathbf{r}_1^{(\mathrm{P1})},\mathbf{r}_2^{(\mathrm{P1})},...,\mathbf{r}_m^{(\mathrm{P1})};\mathbf{r}_1^{(\mathrm{P2})},\mathbf{r}_2^{(\mathrm{P2})},...,\mathbf{r}_n^{(\mathrm{P2})}\right)$ that are denoted as $(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)$. The rest k=m+n-3 centers are denoted as $(\mathbf{r}_4,\mathbf{r}_5,...,\mathbf{r}_{m+n})$. With these notations, and inserting an intermediate partial partition into Eq. (9), one obtains the following factorization:

$$\begin{aligned}
Z_{m-1+n0}&=\int_{\mathrm{site}}\prod_{i=2}^{m+n}d^3x_i\,\exp\left[-\left(W[\mathbf{r}_{10},\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4,\mathbf{r}_5,...,\mathbf{r}_{m+n}]-W[\mathbf{r}_{10},\mathbf{r}_{20},\mathbf{r}_{30},\mathbf{r}_{40},\mathbf{r}_{50},...,\mathbf{r}_{m+n0}]\right)/k_BT\right]\\
&=\frac{\int_{\mathrm{site}}\prod_{i=2}^{m+n}d^3x_i\,\exp\left[-W[\mathbf{r}_{10},\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4,\mathbf{r}_5,...,\mathbf{r}_{m+n}]/k_BT\right]}{\int_{\mathrm{site}}\prod_{i=4}^{m+n}d^3x_i\,\exp\left[-W[\mathbf{r}_{10},\mathbf{r}_{20},\mathbf{r}_{30},\mathbf{r}_4,\mathbf{r}_5,...,\mathbf{r}_{m+n}]/k_BT\right]}\times\\
&\qquad\qquad\frac{\int_{\mathrm{site}}\prod_{i=4}^{m+n}d^3x_i\,\exp\left[-W[\mathbf{r}_{10},\mathbf{r}_{20},\mathbf{r}_{30},\mathbf{r}_4,\mathbf{r}_5,...,\mathbf{r}_{m+n}]/k_BT\right]}{\exp\left[-W[\mathbf{r}_{10},\mathbf{r}_{20},\mathbf{r}_{30},\mathbf{r}_{40},\mathbf{r}_{50},...,\mathbf{r}_{m+n0}]/k_BT\right]}\\
&=Z_{3-1,0}Z_{k0}\,.
\end{aligned} \tag{S17}$$

Here $Z_{3-1,0}$ represents fluctuations of two centers $(\mathbf{r}_2, \mathbf{r}_3)$ when one center is fixed at $\mathbf{r}_1 = \mathbf{r}_{10}$

$$Z_{3-1,0} = \frac{\int_{\text{site}} \prod_{i=2}^{3} d^3 x_i \, \exp\left[-W[\mathbf{r}_{10}, \mathbf{r}_2, \mathbf{r}_3]/k_B T\right]}{\exp\left[-W[\mathbf{r}_{10}, \mathbf{r}_{20}, \mathbf{r}_{30}]/k_B T\right]}. \tag{S18}$$

This 6D partial partition can be factored into three 1D sampling problems in the same manner as Eq. (S8). The resultant form is

$$\begin{aligned} Z_{3-1,0} &= 8\pi^2 \int dr_{21} r_{21}^2 \exp\left[-\frac{W_0[r_{21}] - W_0[r_{210}]}{k_B T}\right] \int d\theta \sin\theta \exp\left[-\frac{W_0[\theta] - W_0[\theta_0]}{k_B T}\right] \times \\ &\qquad \int dr_{31} r_{31}^2 \exp\left[-\frac{W_0[r_{31}] - W_0[r_{310}]}{k_B T}\right] \\ &= \frac{8\pi^2 r_{210}^2 r_{310}^2 \sin\theta_0}{\rho_0(r_{210}) \rho_0(r_{310}) \rho_0(\theta_0)}. \end{aligned} \tag{S19}$$

Noting that the probability distributions here in the associated state take forms similar to those of the dissociated state in Eq. (S11) and Eq. (S14).

$$\begin{aligned} \rho_0(\theta) &= \sin\theta \exp\left[-\mathrm{W}_0[\theta]/k_B T\right] \Big/ \int d\theta' \sin\theta' \exp\left[-\mathrm{W}_0[\theta']/k_B T\right], \\ \rho_0(r_{21}) &= r_{21}^2 \exp\left[-W_0[r_{21}]/k_B T\right] \Big/ \int dr'_{21} r'^2_{21} \exp\left[-W_0[r'_{21}]/k_B T\right], \\ \rho_0(r_{31}) &= r_{31}^2 \exp\left[-W_0[r_{31}]/k_B T\right] \Big/ \int dr'_{31} r'^2_{31} \exp\left[-W_0[r'_{31}]/k_B T\right]. \end{aligned} \tag{S20}$$

$W_0[r_{31}]$ as a function of $\mathrm{r}_{31} = |\mathbf{r}_1 - \mathbf{r}_3|$, 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 $\mathrm{r}_{21} = |\mathbf{r}_1 - \mathbf{r}_2|$, 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 $(\mathbf{r}_{10})$. $\mathrm{r}_{310} = |\mathbf{r}_{10} - \mathbf{r}_{30}|$ and $\mathrm{r}_{210} = |\mathbf{r}_{10} - \mathbf{r}_{20}|$, of course.

Z_{k0} represents fluctuations of k centers $(\mathbf{r}_4, \mathbf{r}_5, ..., \mathbf{r}_{m+n})$ when three center are fixed at $(\mathbf{r}_{10}, \mathbf{r}_{20}, \mathbf{r}_{30})$,

$$Z_{k0} = \frac{\int_{\text{site}} \prod_{i=4}^{m+n} d^3x_i \, \exp\left[-W[\mathbf{r}_{10}, \mathbf{r}_{20}, \mathbf{r}_{30}, \mathbf{r}_4, \mathbf{r}_5, ..., \mathbf{r}_{m+n}]/k_BT\right]}{\exp\left[-W[\mathbf{r}_{10}, \mathbf{r}_{20}, \mathbf{r}_{30}, \mathbf{r}_{40}, \mathbf{r}_{50}, ..., \mathbf{r}_{m+n0}]/k_BT\right]}. \quad \text{(S21)}$$

When the protein association is tight, one can approximate the integral of Eq. (S21) 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, $(\langle \mathbf{r}_4 \rangle, \langle \mathbf{r}_5 \rangle, \cdots, \langle \mathbf{r}_{m+n} \rangle)$. Carrying out the Gaussian integral, one has

$$Z_{k0} = (2\pi)^{3k/2} Det^{1/2}(\Sigma_{ko}) \exp\left[\Delta_{ko} / k_BT\right]. \quad \text{(S22)}$$

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

$$\Delta_{ko}/k_BT = \tfrac{1}{2}\left(\langle \mathbf{r}_4 \rangle - \mathbf{r}_{40}, \langle \mathbf{r}_5 \rangle - \mathbf{r}_{50}, \cdots, \langle \mathbf{r}_{m+n} \rangle - \mathbf{r}_{m+n0}\right) \Sigma_{ko}^{-1} \left(\langle \mathbf{r}_4 \rangle - \mathbf{r}_{40}, \langle \mathbf{r}_5 \rangle - \mathbf{r}_{50}, \cdots, \langle \mathbf{r}_{m+n} \rangle - \mathbf{r}_{m+n0}\right)^T. \quad \text{(S23)}$$

Det represents the determinant. Σ_{ko} is the 3k×3k matrix of the fluctuations/deviations of the pulling center coordinates $\delta x_4 = x_4 - \langle x_4 \rangle$ *etc.*

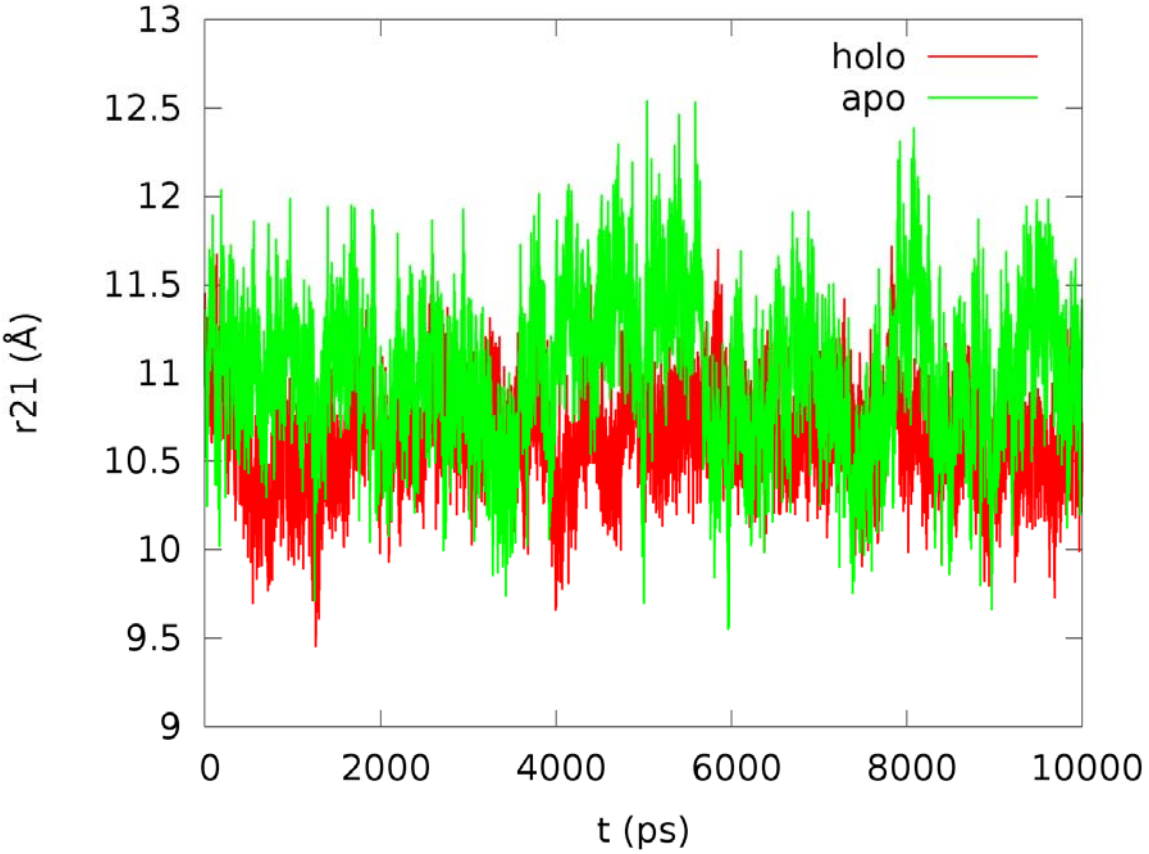
$$\Sigma_{ko} = \left\langle \left(\langle \mathbf{r}_4 \rangle - \mathbf{r}_{40}, \langle \mathbf{r}_5 \rangle - \mathbf{r}_{50}, \cdots, \langle \mathbf{r}_{m+n} \rangle - \mathbf{r}_{m+n0}\right)^T \left(\langle \mathbf{r}_4 \rangle - \mathbf{r}_{40}, \langle \mathbf{r}_5 \rangle - \mathbf{r}_{50}, \cdots, \langle \mathbf{r}_{m+n} \rangle - \mathbf{r}_{m+n0}\right) \right\rangle$$

$$= \begin{pmatrix} \langle \delta x_4^2 \rangle & \langle \delta x_4 \delta y_4 \rangle & \langle \delta x_4 \delta z_4 \rangle & \cdots & \langle \delta x_4 \delta x_{m+n} \rangle & \langle \delta x_4 \delta y_{m+n} \rangle & \langle \delta x_4 \delta z_{m+n} \rangle \\ \langle \delta y_4 \delta x_4 \rangle & \langle \delta y_4^2 \rangle & \langle \delta y_4 \delta z_4 \rangle & \cdots & \langle \delta y_4 \delta x_{m+n} \rangle & \langle \delta y_4 \delta y_{m+n} \rangle & \langle \delta y_4 \delta z_{m+n} \rangle \\ \langle \delta z_4 \delta x_4 \rangle & \langle \delta z_4 \delta y_4 \rangle & \langle \delta z_4^2 \rangle & \cdots & \langle \delta z_4 \delta x_{m+n} \rangle & \langle \delta z_4 \delta y_{m+n} \rangle & \langle \delta z_4 \delta z_{m+n} \rangle \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \langle \delta x_{m+n} \delta x_4 \rangle & \langle \delta x_{m+n} \delta y_4 \rangle & \langle \delta x_{m+n} \delta z_4 \rangle & \cdots & \langle \delta x_{m+n}^2 \rangle & \langle \delta x_{m+n} \delta y_{m+n} \rangle & \langle \delta x_{m+n} \delta z_{m+n} \rangle \\ \langle \delta y_{m+n} \delta x_4 \rangle & \langle \delta y_{m+n} \delta y_4 \rangle & \langle \delta y_{m+n} \delta z_4 \rangle & \cdots & \langle \delta y_{m+n} \delta x_{m+n} \rangle & \langle \delta y_{m+n}^2 \rangle & \langle \delta y_{m+n} \delta z_{m+n} \rangle \\ \langle \delta z_{m+n} \delta x_4 \rangle & \langle \delta z_{m+n} \delta y_4 \rangle & \langle \delta z_{m+n} \delta z_4 \rangle & \cdots & \langle \delta z_{m+n} \delta x_{m+n} \rangle & \langle \delta z_{m+n} \delta y_{m+n} \rangle & \langle \delta z_{m+n}^2 \rangle \end{pmatrix} \quad \text{(S24)}$$

Σ_{ko}^{-1} is the inverse matrix of Σ_{ko} 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 $(\mathbf{r}_{10}, \mathbf{r}_{20}, \mathbf{r}_{30})$.

III. Fluctuations of the pulling centers.

Shown in Figs. S1 to S3 are the fluctuations of the three pulling centers on Ras and the three pulling centers on RalGDS. These data were used to compute the partial partitions.



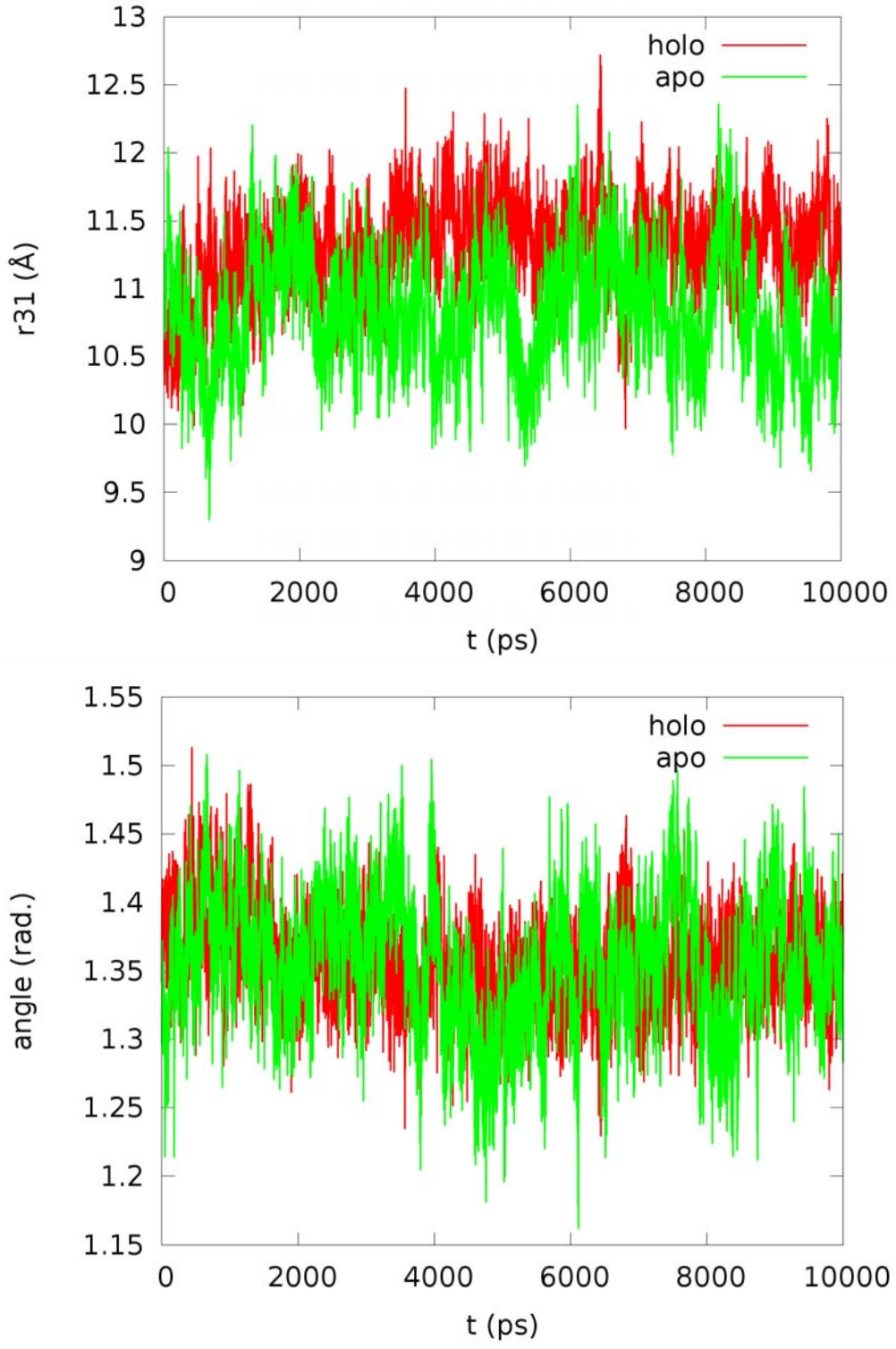


Fig. S1. The fluctuations of the three pulling centers on Ras in the dissociated state vs the same in the associated state.

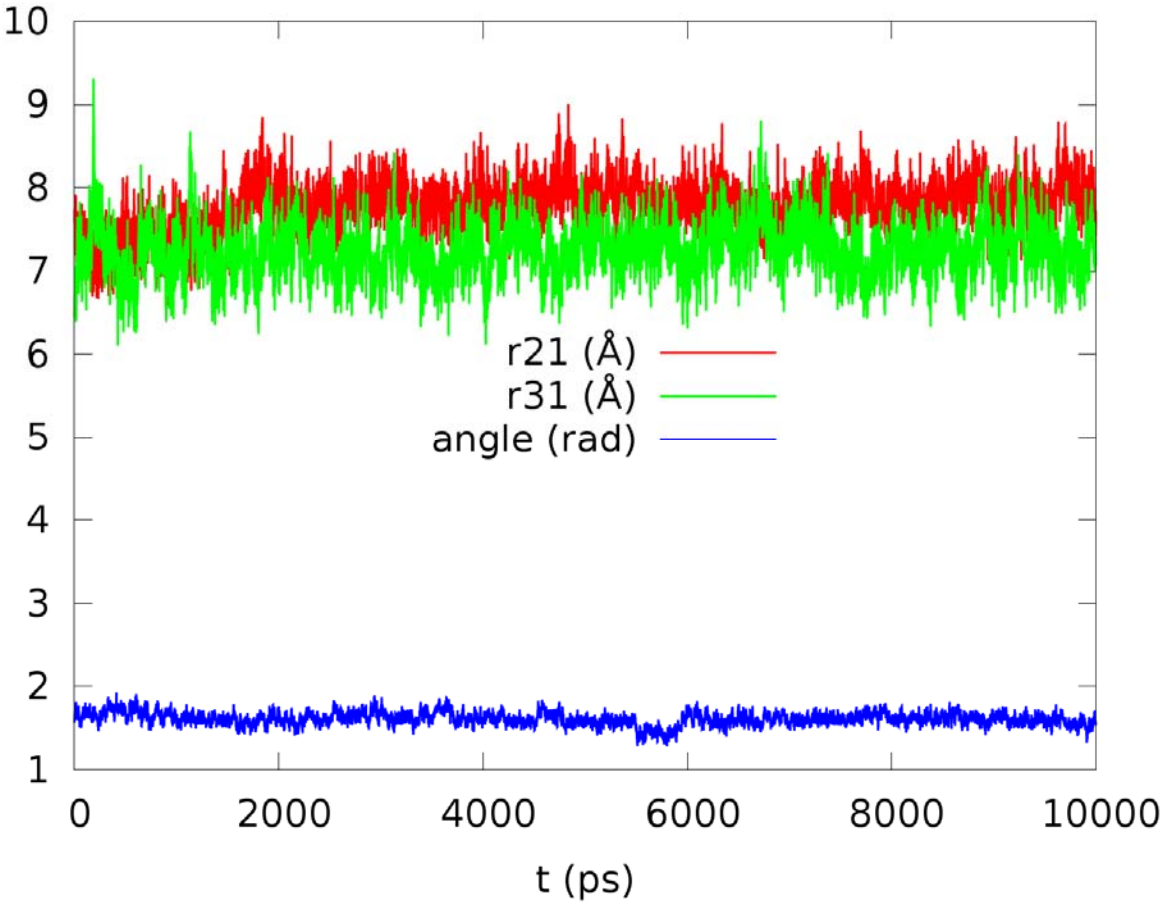
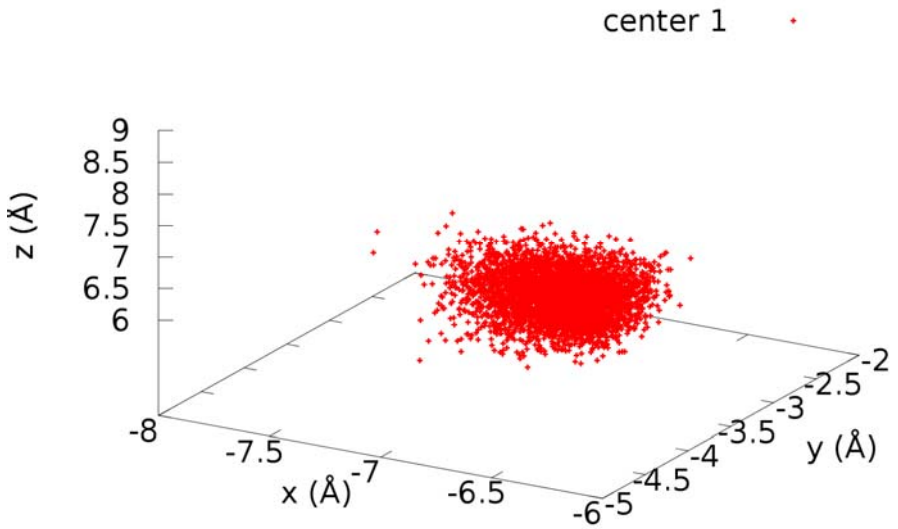
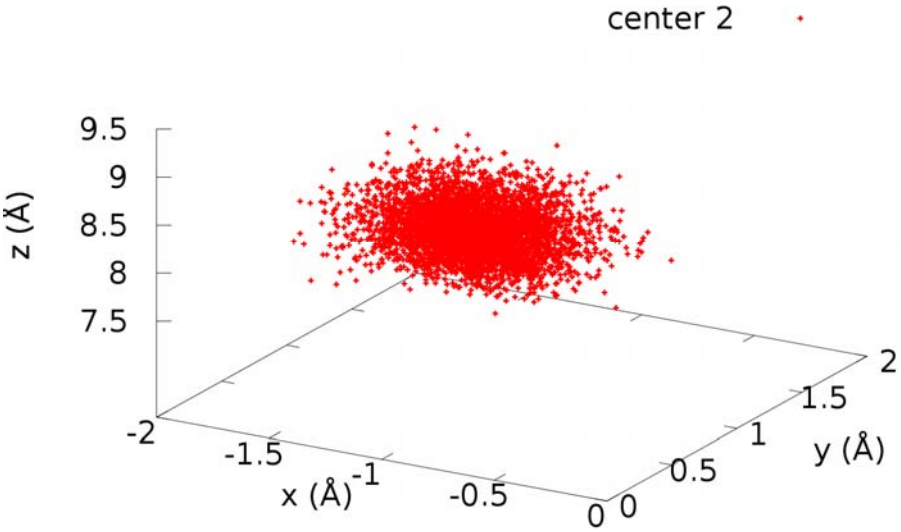


Fig. S2. The fluctuations of the three pulling centers on RalGDS in the dissociated state.


center 1
z (Å)
9
8.5
8
7.5
7
6.5
6
-8
-7.5
-7
-6.5
-6
x (Å)
-2
-2.5
-3
-3.5
-4
-4.5
-5
y (Å)


center 2
z (Å)
9.5
9
8.5
8
7.5
-2
-1.5
-1
-0.5
0
x (Å)
2
1.5
1
0.5
0
y (Å)

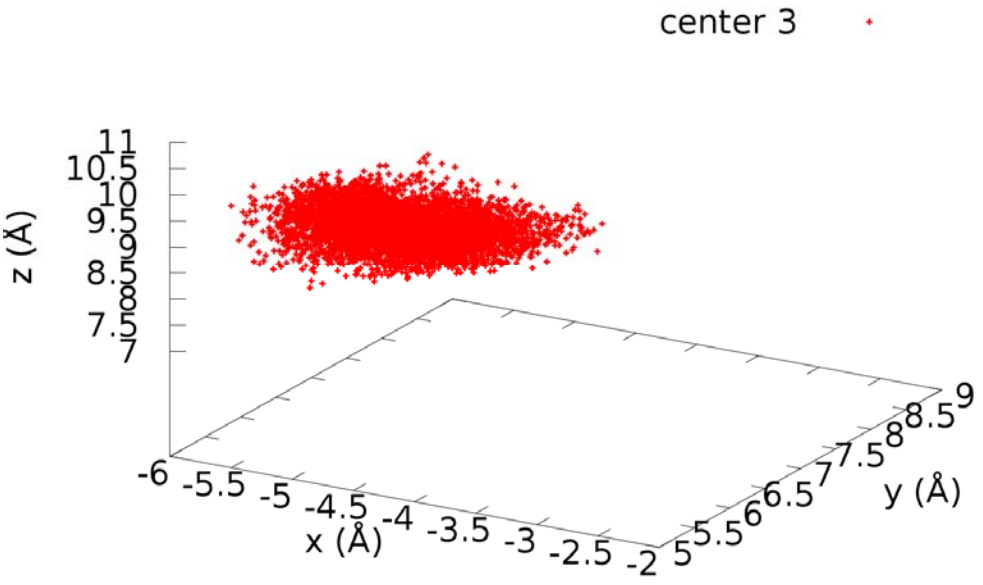


Fig. S3. The fluctuations of the three pulling centers on RalGDS in the associated state.

IV. Work along the pulling paths.

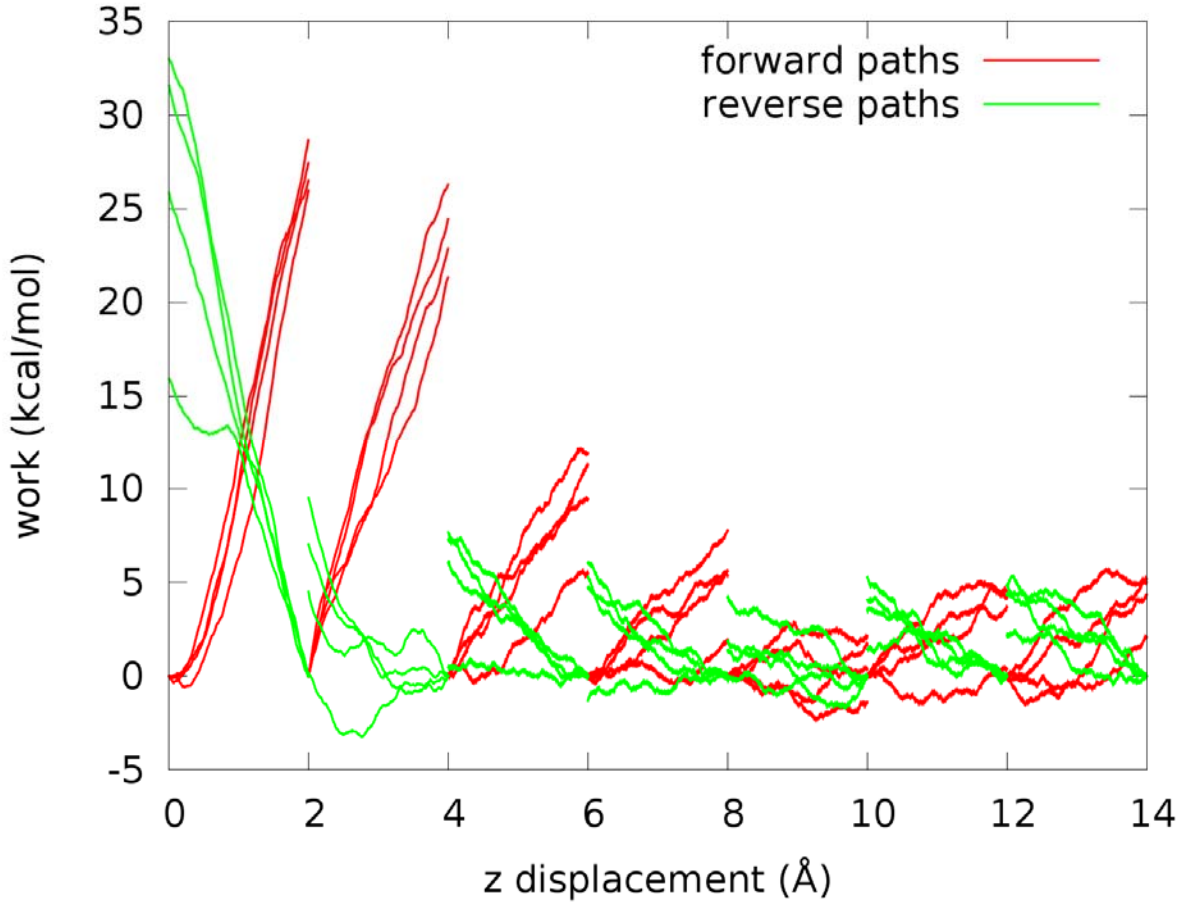


Fig. S4. Work curves along the pulling paths that were used to extract the PMF curve shown in Fig. 3.

V. Computing the association affinity of the barnase-barstar complex.

The all-atom model system of Barnase-Barstar complex[1-6], shown in Fig. S5, was formed from the crystallographic structure (PDB code: 1BRS)[5] by taking its C and F chains, rotating it to the orientation of Barstar on top of Barnase along the z-axis, putting the complex in the center of a water box of $80\mathring{A}\times 80\mathring{A}\times 160\mathring{A}$, neutralizing the system with four Na^+ ions, and salinating the system to 100 mM of NaCl. During the 50 ns equilibrium MD run, system settles down (after 3 ns) to fluctuate slightly around the dimensions of $80\mathring{A}\times 80\mathring{A}\times 147\mathring{A}$. It needs to be pointed out that, unlike what was done in the computational study of the wild-type complex,[6] residues 40 and 82 of this protein-protein complex were not mutated back to cysteine in this study.

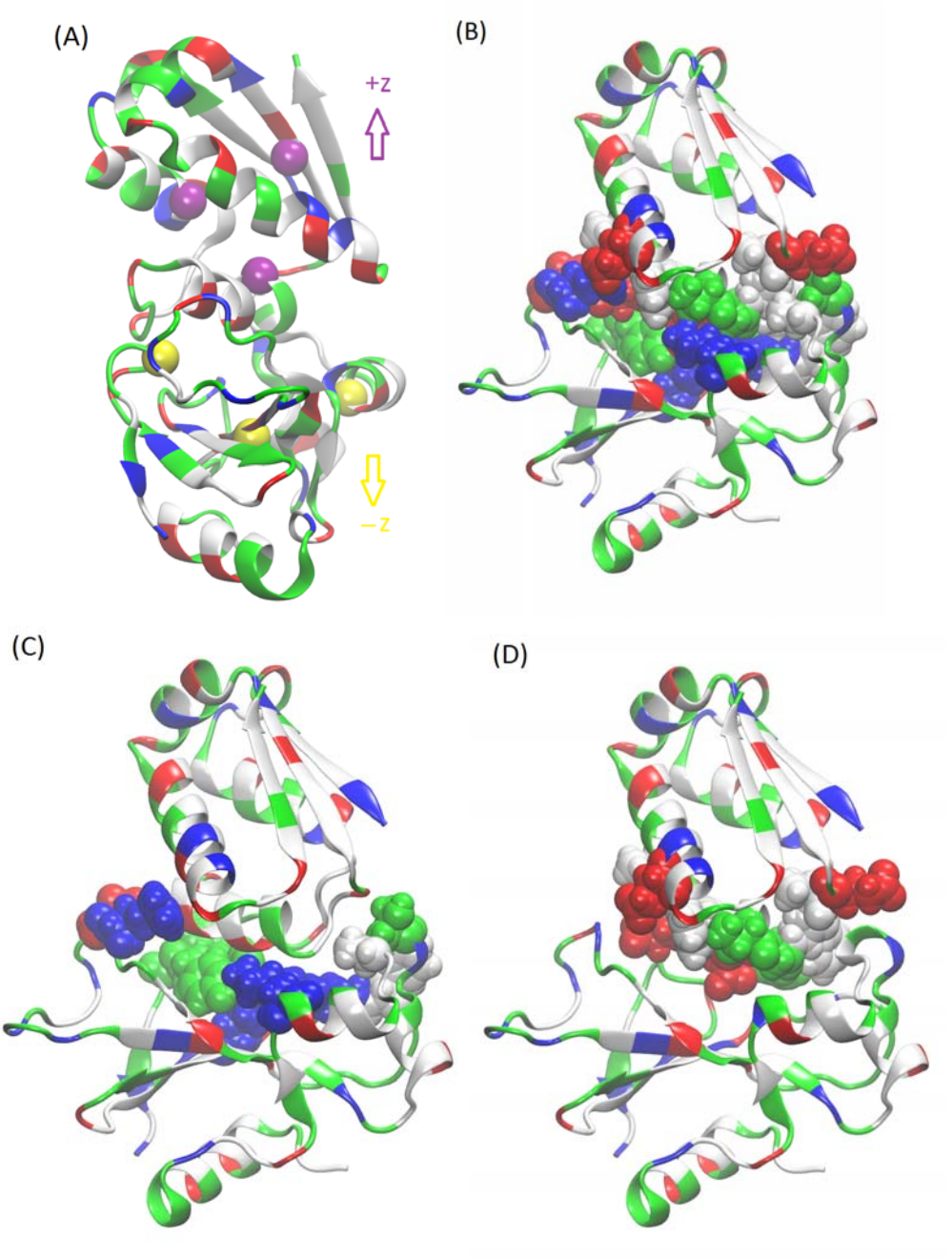


Fig. S5. Barnase-barstar complex shown in ribbons colored by residue types. Also shown in (A) are the six pulling centers: The alpha carbons of three barstar residues (Leu 20, Leu 41, and Leu 51) are shown as purple spheres. They are chosen as pulling centers to steered along the +z direction. The alpha carbons of three barnase residues (Ala 30, Ala 74, and Tyr 103) are shown as yellow spheres. These three are chosen as pulling centers to be steered in the –z direction. (B) The interfacial residues are shown as spheres colored by residue types: Barstar residues Tyr 29, Asn 33, Leu 34, Asp 35, Ala 36, Trp 38, Trp 44, Val 45, Glu 46, and Glu 76. Barnase residues Lys 27, Ala 37, Ser 38, Arg 59, Glu 60, Phe 82, Arg 83, Ser 85 His 102, Tyr 103, and Gln 104. (C) and (D) show, respectively, the barnase and barstar residues within a distance of 3 Å away from the partner proteins barstar and barnase. The representations in (C) and (D) are identical to (B).

For the barnase-barstar system, we choose three alpha carbons on each protein as the steering/pulling centers: Ala 30, Ala 74, and Tyr 103 on barnase (noted as P1, their position vectors noted as $\left(\mathbf{r}_1^{(P1)},\mathbf{r}_2^{(P1)},\mathbf{r}_3^{(P1)}\right)$) along with Leu 20, Leu 41 and Leu 51 on barstar (notated as P2, position vectors as $\left(\mathbf{r}_1^{(P2)},\mathbf{r}_2^{(P2)},\mathbf{r}_3^{(P2)}\right)$). The numerical results are tabulated in Table S1.

Table S1. Computed results of barnase-barstar complex.

The 6-D partial partition of barnase in the dissociated state when one center is fixed	$Z_{3-1\infty}^{(P1)}$	$5.92\times10^{5}\,\text{Å}^{6}$
The 6-D partial partition of barstar in the dissociated state when one center is fixed	$Z_{3-1\infty}^{(P2)}$	$3.30\times10^{5}\,\text{Å}^{6}$
The 6-D partial partition of barnase-barstar complex in the associated state when one center is fixed	$Z_{3-1,0}$	$3.17\times10^{5}\,\text{Å}^{6}$
The 9x9 matrix for fluctuations in the associated state defined in Eq. (25) [or Eq. (S24)]	$Det(\Sigma_{3o})$	$1.808\times10^{-10}\,\text{Å}^{18}$
The deviations of the initial state from the equilibrium average state as measured in Eq. (23) [or Eq. (S25)] for k=3	Δ_{3o}	4.66 kcal/mol
PMF difference between the one chosen associated state and the corresponding dissociated state	$\Delta W_{0,\infty}$	−26.3±1.5 kcal/mol
The absolute free energy of barnase-barstar association, computed with hSMD, Cys(40, 83)Ala double mutant, 100 mM NaCl	ΔG_{hSMD}	−16.8±1.5 kcal/mol
The experimental data of Cys(40, 83)Ala double mutant, 100 mM NaCl [2, 5] (The effect of 100 mM NaCl reduces the binding by 1.4 kcal/mol.[5])	$\Delta G_{\text{exp.}}$	−16.0 kcal/mol
The experimental data of wild type, no NaCl[5]	$\Delta G_{\text{exp.}}$	−18.9 kcal/mol
The barnase-barstar association free energy, computed by Gumbart, Roux, and Chipot[6], wild type	ΔG_{GBSA}	−21.0±1.4 kcal/mol

Note that each of the three 6-D partial partitions involved in Eq. (28) was reduced three 1-D sampling problems for which the numerical data were presented in Figs. S7 and S8. The 9x9 matrix Σ_{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 Fig. S9. The 18-D PMF along the puling path illustrated in Fig. S6 was computed from the work curves shown in Fig. S10 using Eq. (27). The PMF difference $\Delta W_{0,\infty}$ was taken between z=0 and z=9Å.

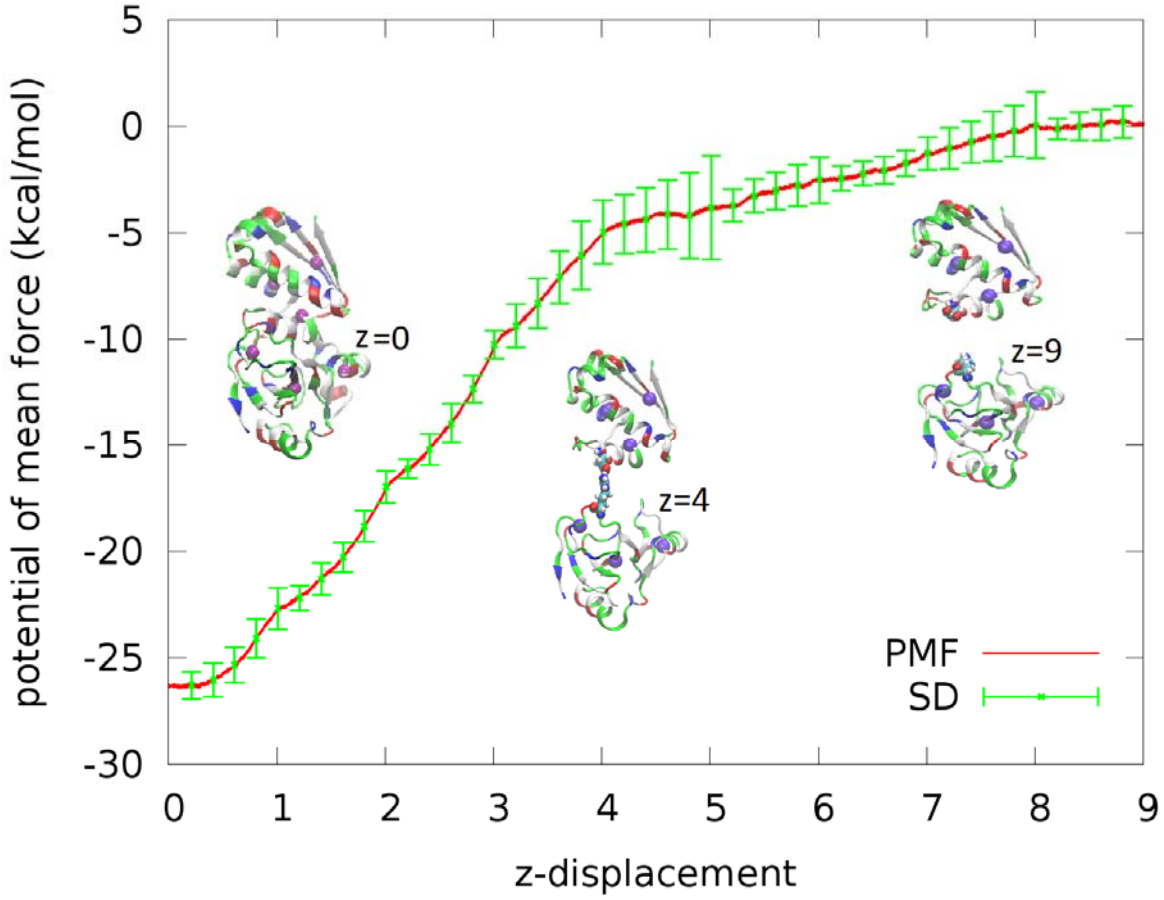
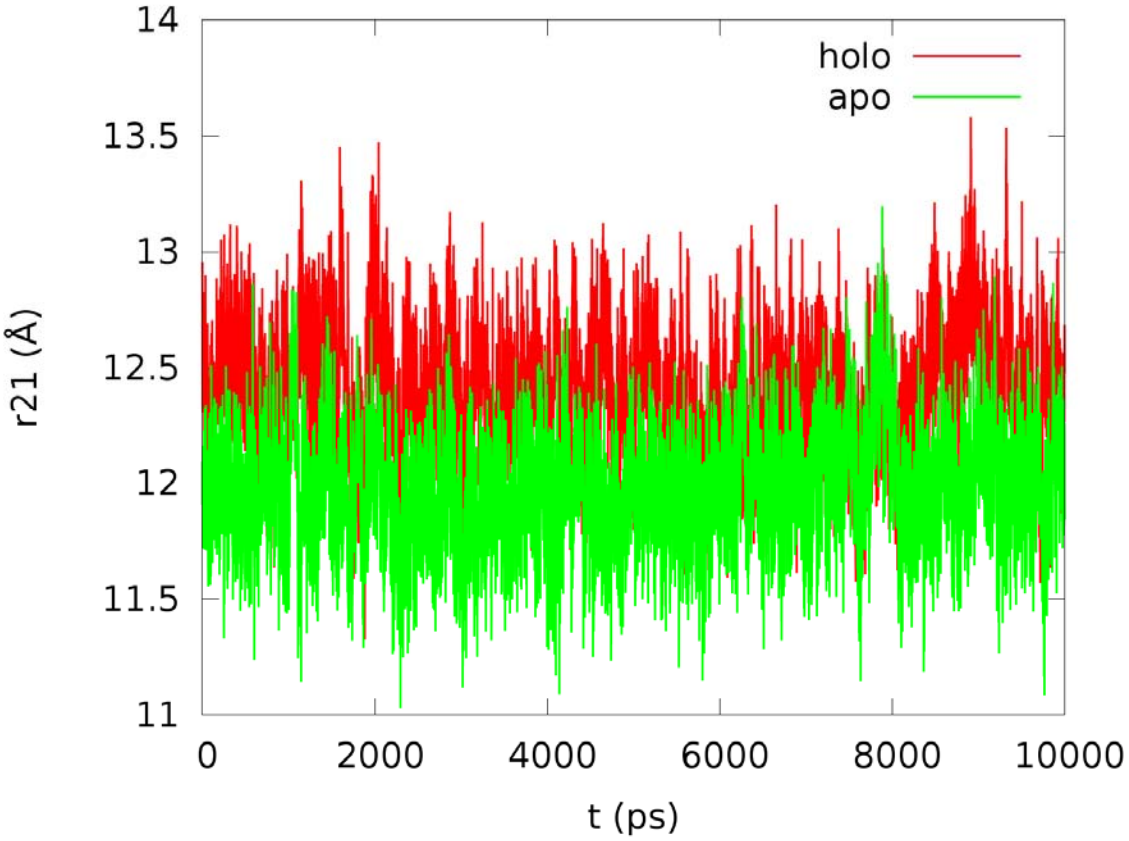
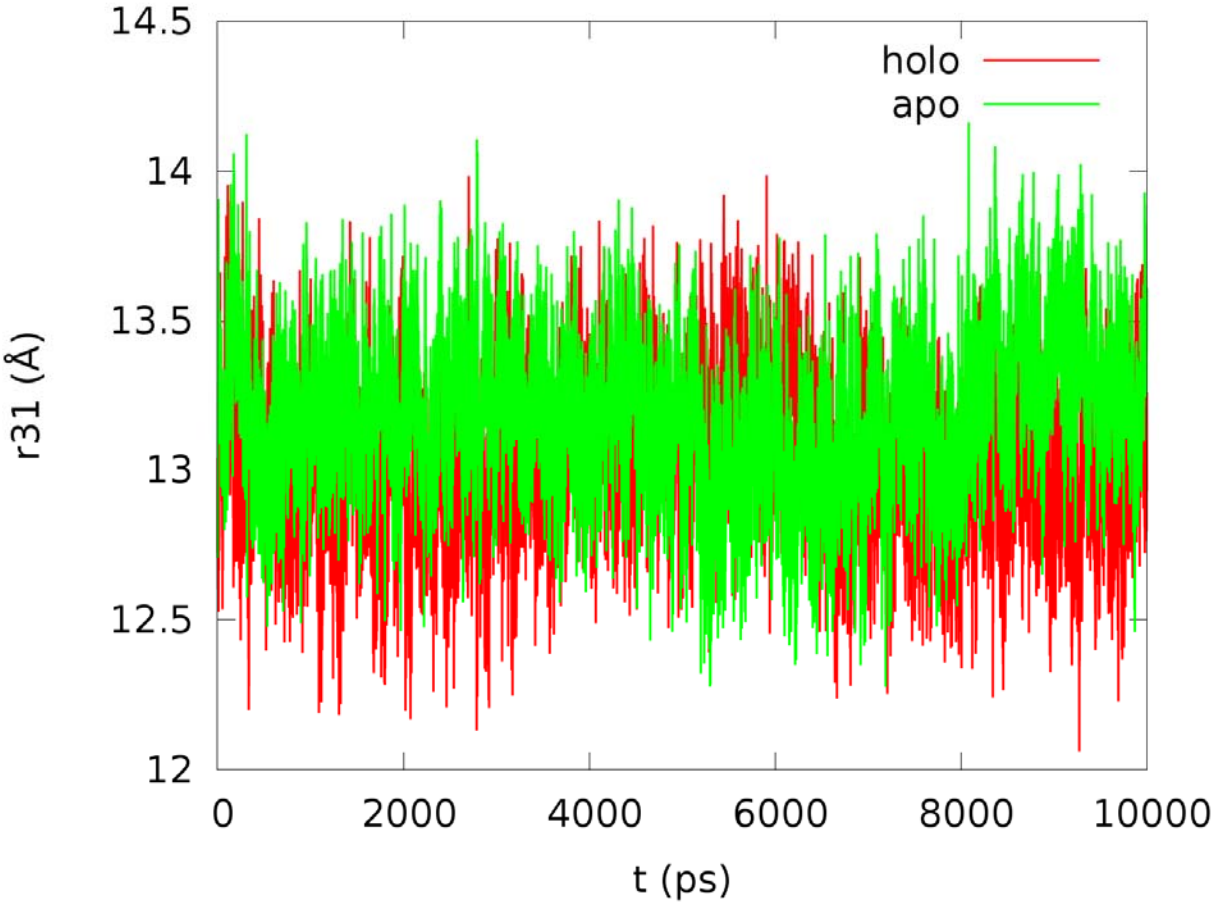


Fig. S6. PMF along the pulling path from the associated state to the dissociated state. The horizontal axis indicates displacement z of barstar relative to barnase when they are steered in the positive and the negative z-direction, respectively.


holo
apo
r21 (Å)
t (ps)
14
13.5
13
12.5
12
11.5
11
0
2000
4000
6000
8000
10000


14.5
14
13.5
13
12.5
12
r31 (Å)
0
2000
4000
6000
8000
10000
t (ps)
holo
apo

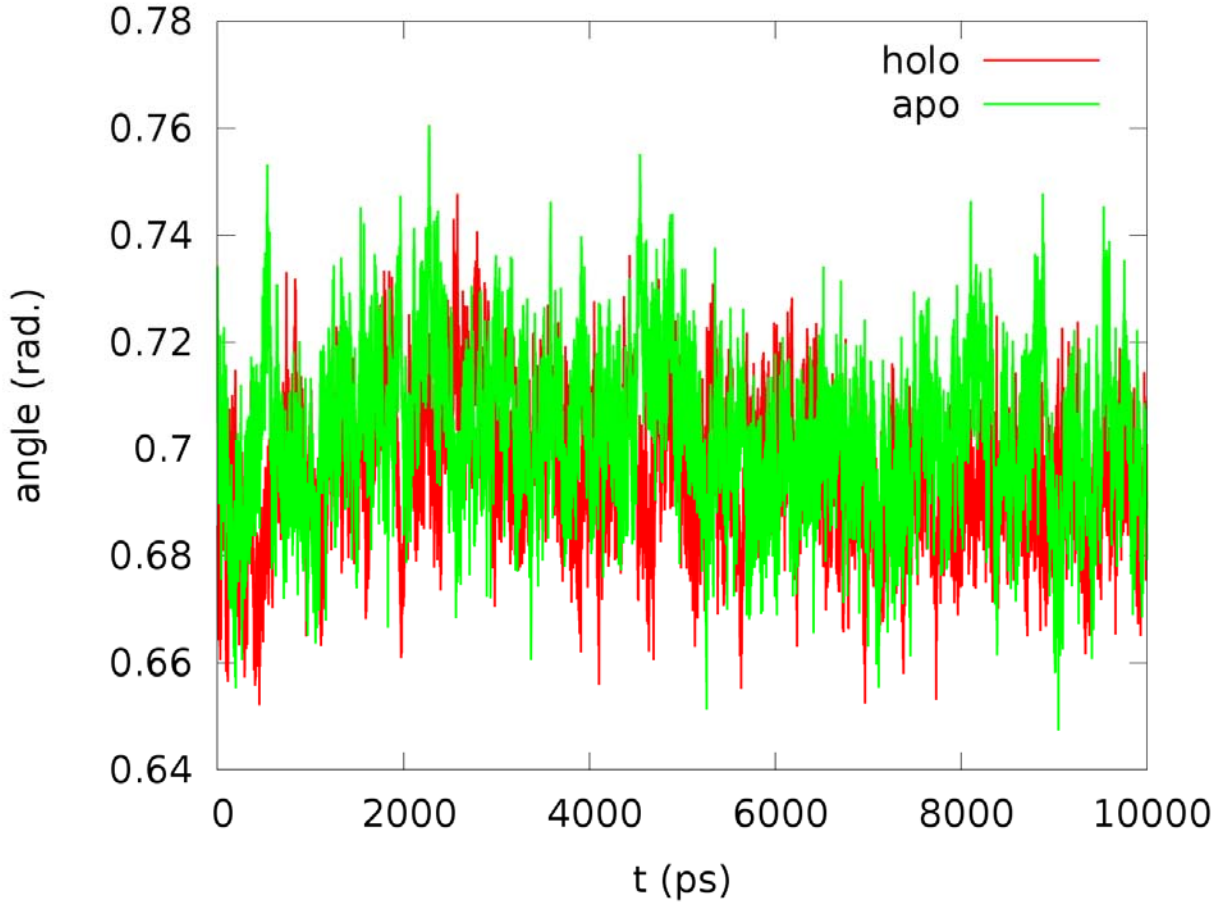


Fig. S7. The fluctuations of the three pulling centers on Barnase in the dissociated state vs the same in the associated state.

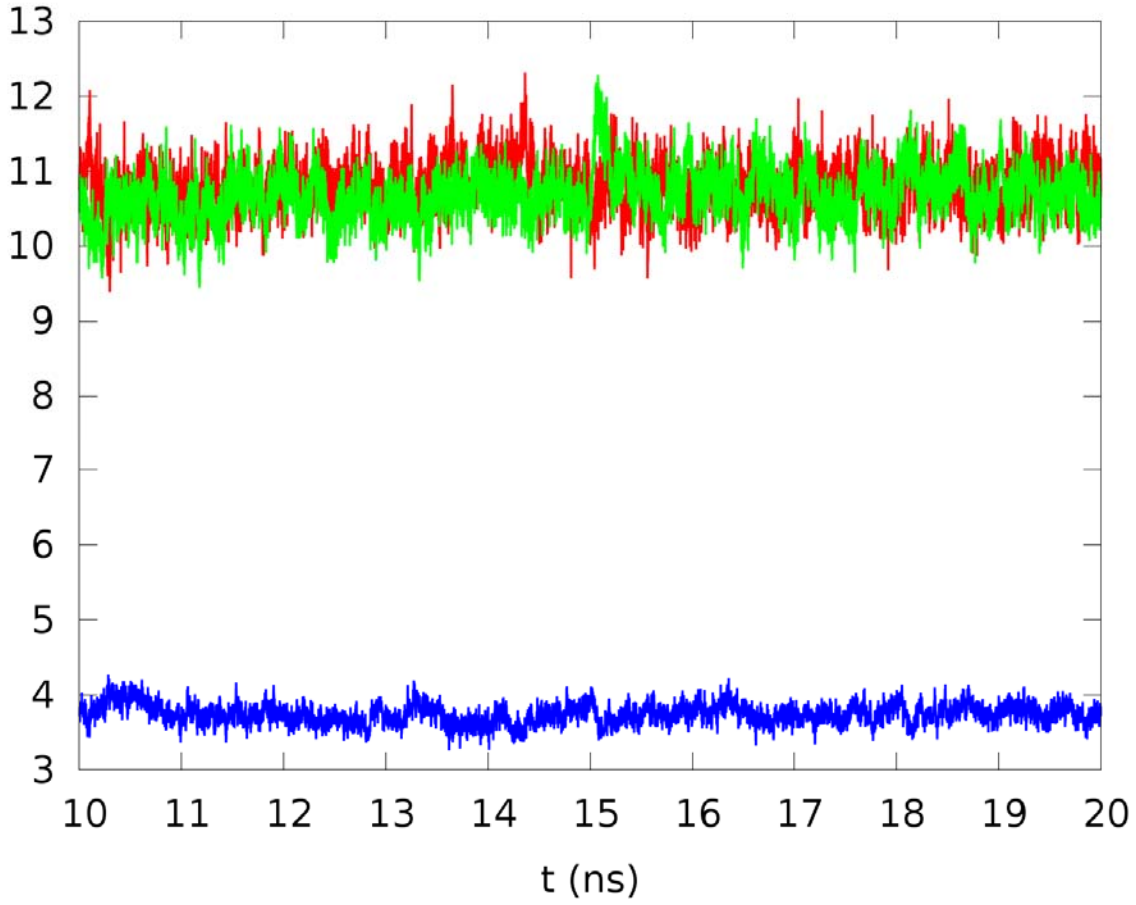
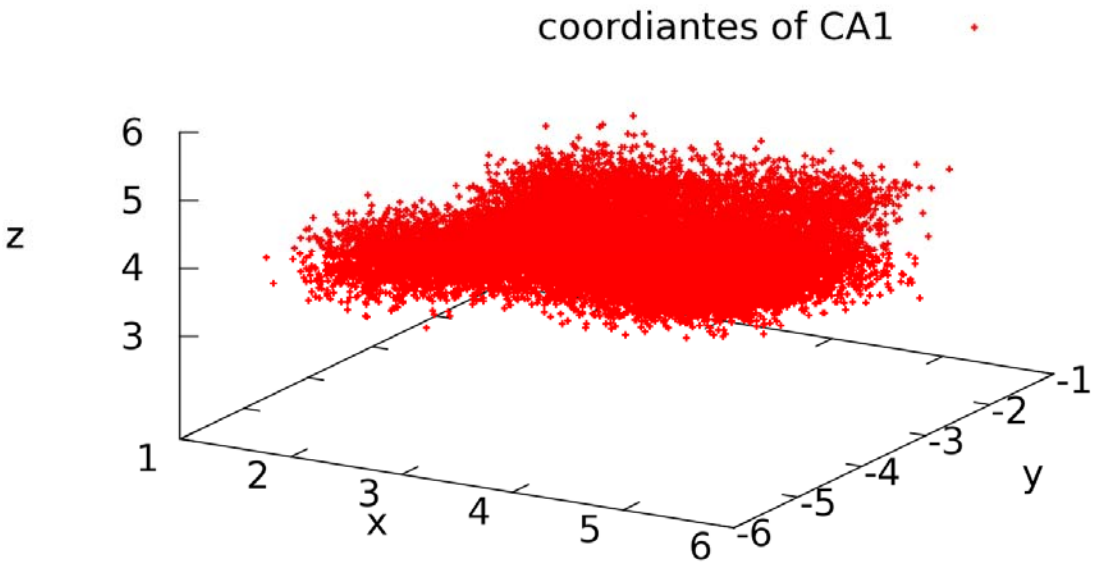
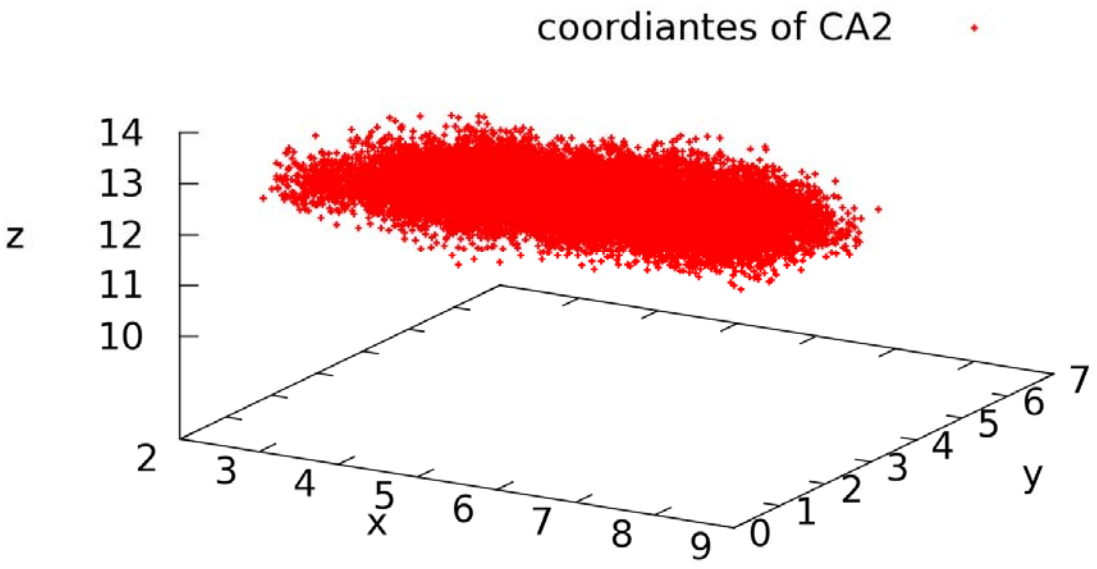


Fig. S8. The fluctuations of the three pulling centers on Barstar in the dissociated state.


coordiantes of CA1
z
6
5
4
3
1
2
3
4
5
6
x
-1
-2
-3
-4
-5
-6
y


coordiantes of CA2
z
14
13
12
11
10
2
3
4
5
6
7
8
9
x
0
1
2
3
4
5
6
7
y

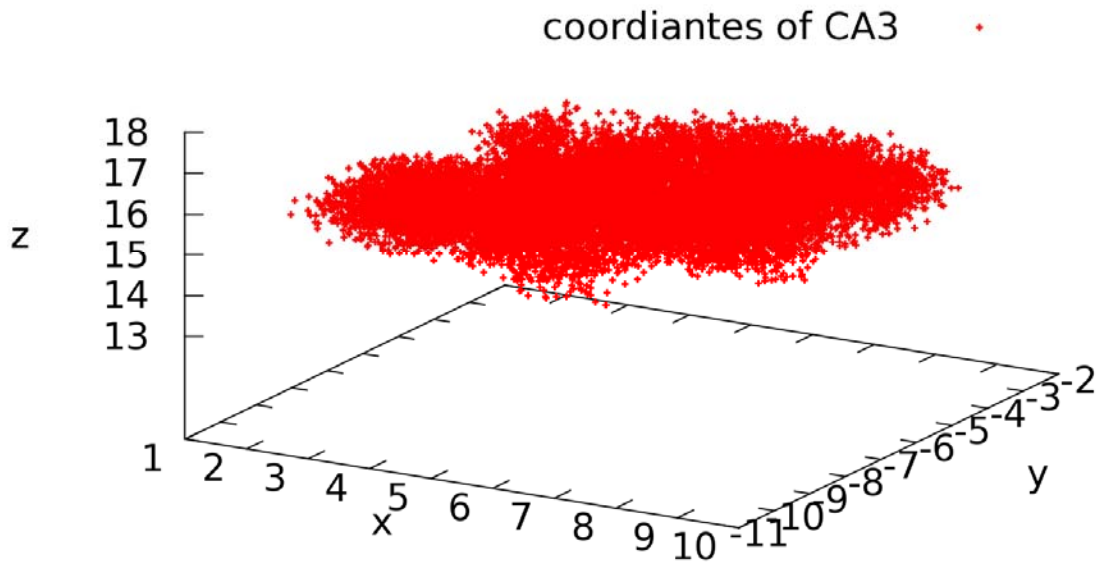


Fig. S9. The fluctuations of the three pulling centers on Barstar in the associated state.

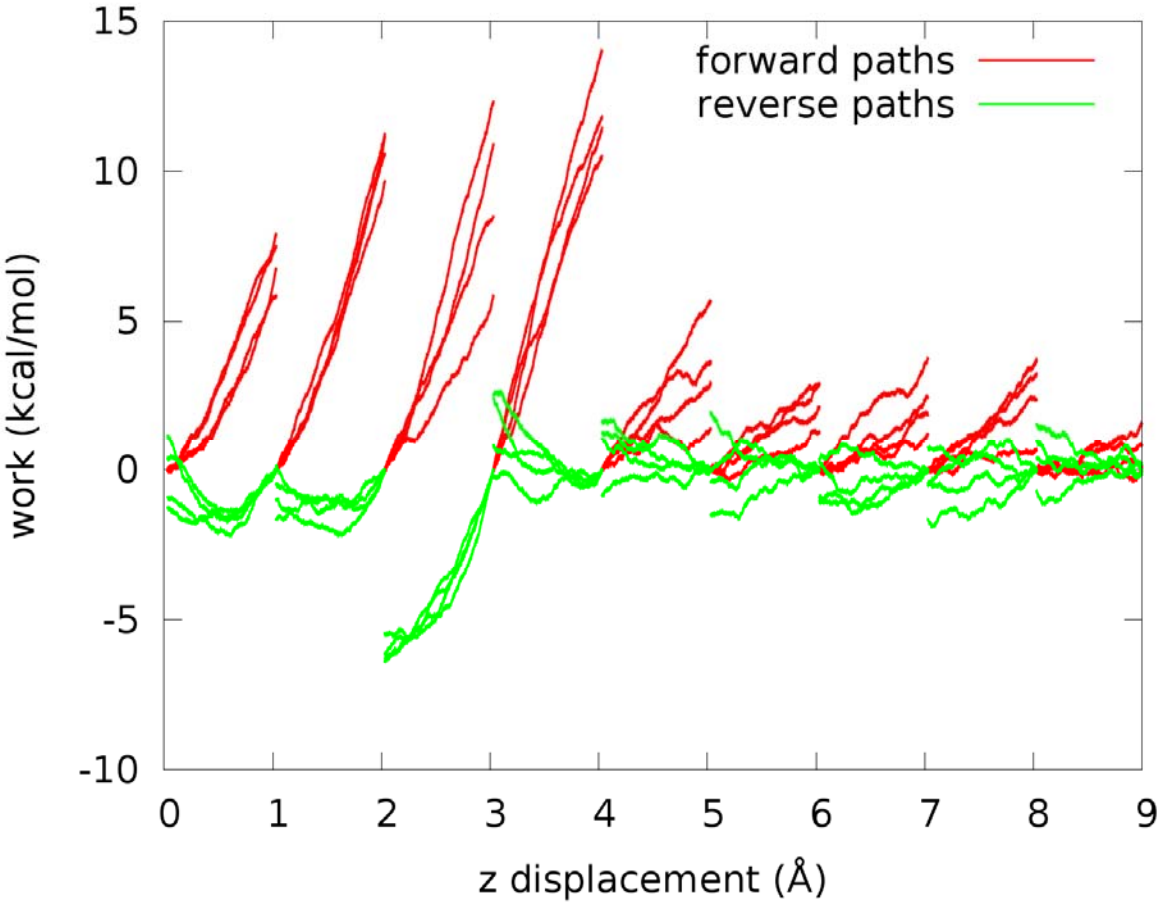


Fig. S10. Work curves along the pulling paths that were used to extract the PMF curve shown in Fig. S6.

References

(1) Hartley, R. W., Barnase and barstar: two small proteins to fold and fit together. *Trends Biochem. Sci.* **1989,** *14*, 450-454.
(2) Schreiber, G.; Fersht, A. R., Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry* **1993,** *32*, 5145-5150.
(3) Schreiber, G. F., A. R., Rapid, electrostatically assisted association of proteins *Nat. Struct. Biol.* **1996,** *3*, 427-431.
(4) Wang, T.; Tomic, S.; Gabdoulline, R. R.; Wade, R. C., How Optimal Are the Binding Energetics of Barnase and Barstar? *Biophys. J.* **2004,** *87*, 1618-1630.
(5) Buckle, A. M.; Schreiber, G.; Fersht, A. R., Protein-protein recognition: Crystal structural analysis of a barnase-barstar complex at 2.0-.ANG. resolution. *Biochemistry* **1994,** *33*, 8878-8889.
(6) 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.