Supporting Information:

DeepBAR: A Fast and Exact Method for Binding Free Energy Computation

Xinqiang Ding and Bin Zhang*

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

E-mail: binz@mit.edu

Definition of Bound and Unbound States

To compute the binding free energy between CB7 and the guest molecules with DeepBAR, we first need to precisely define the bound state A and the unbound state B such that the binding free energy $\Delta F_{\text{binding}}$ equals $F_A - F_B$.

Based on previous studies,[S1,S2] the equilibrium binding constant between the host and a guest molecule, K_b, can be derived as

$$K_b = \frac{\int_{\text{site}} \mathrm{d}\boldsymbol{x}_g \int \mathrm{d}\boldsymbol{x}_h^* \; e^{-\beta U(\boldsymbol{x}_h^*, \boldsymbol{x}_g)}}{\left[\int_{\text{bulk}} \mathrm{d}\boldsymbol{x}_g \; \delta(\mathbf{r}_g - \mathbf{r}') \; e^{-\beta U(\boldsymbol{x}_g)}\right] \left[\int \mathrm{d}\boldsymbol{x}_h^* \; e^{-\beta U(\boldsymbol{x}_h^*)}\right]}. \tag{S1}$$

We use $\boldsymbol{x}_h$ and $\boldsymbol{x}_g$ for the full phase space of host and guest and $\boldsymbol{x}_h^*$ and $\boldsymbol{x}_g^*$ for the ones with translational and rotational degrees of freedom removed. $\mathbf{r}'$ indicates an arbitrary but fixed location in the bulk region that is far away from the host and $\mathbf{r}_g$ represents the guest's center of mass. $U(\boldsymbol{x}_h^*, \boldsymbol{x}_g)$ is the total potential energy of both host and guest molecules, while $U(\boldsymbol{x}_g)$ and $U(\boldsymbol{x}_h^*)$ are the individual energy. By definition, K_b has a unit of volume (e.g. nm^3). The standard binding free energy $\Delta F_{\text{binding}}$ is related to K_b as

$$\Delta F_{\text{binding}} = -k_b T \ln \left[K_b C^\circ\right], \tag{S2}$$

where C° is the standard concentration of 1 mol/L or equivalently $1/V^\circ$ with $V^\circ = 1.661$ nm^3. Combining Equation S1 and S2, we have

$$\begin{aligned}\Delta F_{\text{binding}} = &-k_b T \ln \Big[\int_{\text{site}} \mathrm{d}\boldsymbol{x}_g \int \mathrm{d}\boldsymbol{x}_h^* \; e^{-\beta U(\boldsymbol{x}_h^*, \boldsymbol{x}_g)} \Big] \\ &- \Big(- k_b T \ln \Big[V^\circ \cdot \int_{\text{bulk}} \mathrm{d}\boldsymbol{x}_g \; \delta(\mathbf{r}_g - \mathbf{r}') \; e^{-\beta U(\boldsymbol{x}_g)} \Big] - k_b T \ln \Big[\int \mathrm{d}\boldsymbol{x}_h^* \; e^{-\beta U(\boldsymbol{x}_h^*)} \Big] \Big).\end{aligned} \tag{S3}$$

Equation S3 shows that, to make $\Delta F_{\text{binding}} = F_{\text{A}} - F_{\text{B}}$ with

$$F_{\text{A}} = -k_b T \ln \Big[\int_{\text{site}} \mathrm{d}\boldsymbol{x}_g \int \mathrm{d}\boldsymbol{x}_h^* \; e^{-\beta U(\boldsymbol{x}_h^*, \boldsymbol{x}_g)} \Big], \tag{S4}$$

and

$$F_{\text{B}} = -k_b T \ln \Big[V^\circ \cdot \int_{\text{bulk}} \mathrm{d}\boldsymbol{x}_g \; \delta(\mathbf{r}_g - \mathbf{r}') \; e^{-\beta U(\boldsymbol{x}_g)} \Big] - k_b T \ln \Big[\int \mathrm{d}\boldsymbol{x}_h^* \; e^{-\beta U(\boldsymbol{x}_h^*)} \Big], \tag{S5}$$

we can define the bound and the unbound state as follows (see Figure 2a and 2b of the main text). The bound state A is the host-guest ensemble where the host has a fixed location and orientation, and the guest is inside the binding site. Therefore, the configurational space and energy function for state A can be denoted as $\boldsymbol{x}_{\text{A}} = (\boldsymbol{x}_h^*, \boldsymbol{x}_g)$ and $U_{\text{A}}(\boldsymbol{x}_{\text{A}}) = U(\boldsymbol{x}_h^*, \boldsymbol{x}_g)$. The unbound state B corresponds to the conformational ensemble in which the host has a fixed location and orientation, and the guest's center of mass is inside a box that is in the bulk region far away from the host and has a volume of V°. In state B, the integral corresponding to the guest can also be written as

$$V^\circ \cdot \int_{\text{bulk}} \mathrm{d}\boldsymbol{x}_g \; \delta(\mathbf{r}_g - \mathbf{r}') \; e^{-\beta U(\boldsymbol{x}_g)} = V^\circ \cdot 8\pi^2 \cdot \int \mathrm{d}\boldsymbol{x}_g^* \;\; e^{-\beta U(\boldsymbol{x}_g^*)}, \tag{S6}$$

where $8\pi^2$ is the result of integrating the guest's rotational degrees of freedom. Using Equation S6, F_{B} in Equation S5 can also be written as

$$F_{\text{B}} = -k_b T \ln \Big[\int \mathrm{d}\boldsymbol{x}_g^* \int \mathrm{d}\boldsymbol{x}_h^* \;\; e^{-\beta \left[U(\boldsymbol{x}_g^*) + U(\boldsymbol{x}_h^*) - \beta^{-1} \ln(8\pi^2 \cdot V^\circ) \right]} \Big]. \tag{S7}$$

Therefore for state B we have $\boldsymbol{x}_{\text{B}} = (\boldsymbol{x}_h^*, \boldsymbol{x}_g^*)$ and $U_{\text{B}}(\boldsymbol{x}_{\text{B}}) = U(\boldsymbol{x}_g^*) + U(\boldsymbol{x}_h^*) - \beta^{-1} \ln(8\pi^2 \cdot V^\circ)$.

Simulation Details for Bound and Unbound States

Biasing potentials were applied to the host to restrain its position and orientation in simulations of both state A and state B. Specifically, three anchor particles, represented as P1, P2, and P3 in Figure S25, were added to the host-guest system. These particles were fixed at (0.0, 0.0, 0.0)nm, (0.1, 0.0, 0.0)nm, and (0.0, 0.0, -0.1)nm, respectively. By definition, the vector P3-P1 aligns with the z-axis, and the vector P1-P2 aligns with the x-axis. Three virtual sites, represented as H1, H2, and H3 in Figure S25, were defined based on the positions of host atoms. They correspond to the average position of the 14 spiro carbon atoms of the host, the average position of the 14 nitrogen atoms on one side of the host and between H1 and P1, and the average position of the four carbon atoms belonging to one of the eight-atom rings. Three harmonic potentials were applied on H1 to restrain its position: $0.5 \times 10^5 \cdot (\text{distance}(\text{H1}, \text{P1}) - 0.5)^2$ on the distance between H1 and P1, $0.5 \times 10^3 \cdot (\text{angle}(\text{H1}, \text{P1}, \text{P3}) - \pi)^2$ on the angle H1-P1-P3, and $0.5 \times 10^3 \cdot (|\text{dihedral}(\text{H1}, \text{P1}, \text{P2}, \text{P3})| - \pi)^2$ on the dihedral angle H1-P1-P2-P3. Here and henceforth, distances, angles and energies are in units of nanometer, radius and kj/mol, respectively. Two biasing potentials were used to restrain H2 such that the vector H1-H2 aligns closely with the z-axis: $0.5 \times 10^3 \cdot (\text{angle}(\text{H2}, \text{H1}, \text{P1}))^2$ on the angle H2-H1-P1, and $0.5 \times 10^3 \cdot (\text{dihedral}(\text{H2}, \text{H1}, \text{P2}, \text{P1}))^2$ on the dihedral angle H2-H1-P2-P1. One biasing potential was added to restrain H3 such that the vector H1-H3 points to the same direction as the vector P1-P2: $0.5 \times 10^3 \cdot (\text{dihedral}(\text{H3}, \text{H1}, \text{H2}, \text{P2}))^2$ on the dihedral angle H3-H1-H2-P2. We note that the above restraining potentials do not affect the host's conformational ensemble.

No restrictions were applied to the guest molecules for simulations of state A, and they remain bound to the host throughout the simulations. In simulations of state B, the guest molecules were positioned at 2.8 nm away from the host, and there are no interactions between the guest and the host. We further restricted the translation and rotation of guest molecules using harmonic biasing potentials. As shown in Equation S7, these degrees of freedom are not needed for computing the absolute free energy. Details on the restraining

potentials applied to the guest molecules can be found in *Section: Restraining potential on guests' position and orientation.*

Generative Models for Reference State Construction

As mentioned in the main text, because the host and the guest interact with each other in state A, we modeled $\boldsymbol{x}_h^*$ and $\boldsymbol{x}_g$ as dependent variables in $q_{\mathrm{A}^\circ}(\boldsymbol{x}_h^*, \boldsymbol{x}_g)$ of state A°, i.e., $q_{\mathrm{A}^\circ}(\boldsymbol{x}_h^*, \boldsymbol{x}_g) = q_{\mathrm{A}^\circ}(\boldsymbol{x}_h^*) \cdot q_{\mathrm{A}^\circ}(\boldsymbol{x}_g|\boldsymbol{x}_h^*)$. In contrast, there are no interactions between the host and guest molecules in state B, and $\boldsymbol{x}_h^*$ and $\boldsymbol{x}_g^*$ were modeled as independent variables in the reference state B°: $q_{\mathrm{B}^\circ}(\boldsymbol{x}_h^*, \boldsymbol{x}_g^*) = q_{\mathrm{B}^\circ}(\boldsymbol{x}_h^*) \cdot q_{\mathrm{B}^\circ}(\boldsymbol{x}_g^*)$.

Because the host is relatively rigid and has small fluctuations in both state A and B, we modeled both $q_{\mathrm{A}^\circ}(\boldsymbol{x}_h^*)$ and $q_{\mathrm{B}^\circ}(\boldsymbol{x}_h^*)$ as results of applying an invertible linear transformation (ILT) to a standard normal random variable $\mathbf{u}_h^* \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. In detail, $\boldsymbol{x}_h^* = T(\mathbf{u}_h^*) = \mathbf{M}_{\mathrm{A}^\circ}\mathbf{u}_h^* + \boldsymbol{\mu}_{\mathrm{A}^\circ}$ and $\boldsymbol{x}_h^* = T(\mathbf{u}_h^*) = \mathbf{M}_{\mathrm{B}^\circ}\mathbf{u}_h^* + \boldsymbol{\mu}_{\mathrm{B}^\circ}$ for state A° and B°, respectively, where $\mathbf{M}_{\mathrm{A}^\circ}$ and $\mathbf{M}_{\mathrm{B}^\circ}$ are $3N_h \times 3N_h$ invertible matrices, $\boldsymbol{\mu}_{\mathrm{A}^\circ}$ and $\boldsymbol{\mu}_{\mathrm{B}^\circ}$ are vectors with $3N_h$ components, and $N_h = 126$ is the number of host atoms. As a result, the distribution of $\boldsymbol{x}_h^*$ in state A° can be derived as

$$
\begin{aligned}
q_{\mathrm{A}^\circ}(\boldsymbol{x}_h^*) &= p_u(\boldsymbol{u}_h^*)|\mathrm{det}J_T(\boldsymbol{u}_h^*)|^{-1} \\
&= p_u\left(\mathbf{M}_{\mathrm{A}^\circ}^{-1}(\boldsymbol{x}_h^* - \boldsymbol{\mu}_{\mathrm{A}^\circ})\right)|\mathrm{det}J_T(\boldsymbol{u}_h^*)|^{-1} \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^{3N_h} \exp\left\{-\frac{1}{2}(\boldsymbol{x}_h^* - \boldsymbol{\mu}_{\mathrm{A}^\circ})^{\mathrm{T}}(\mathbf{M}_{\mathrm{A}^\circ}\mathbf{M}_{\mathrm{A}^\circ}^{\mathrm{T}})^{-1}(\boldsymbol{x}_h^* - \boldsymbol{\mu}_{\mathrm{A}^\circ})\right\}|\mathrm{det}J_T(\boldsymbol{u}_h^*)|^{-1}
\end{aligned}
\tag{S8}
$$

The Jacobian $J_T(\boldsymbol{u})$ is the $3N_h \times 3N_h$ matrix of all partial derivatives of T given by $\mathbf{M}_{\mathrm{A}^\circ}$. It is evident that $q_{\mathrm{A}^\circ}(\boldsymbol{x}_h^*)$ is a multivariate normal distribution with mean $\boldsymbol{\mu}_{\mathrm{A}^\circ}$ and covariance matrix $\boldsymbol{\Sigma} = \mathbf{M}_{\mathrm{A}^\circ}\mathbf{M}_{\mathrm{A}^\circ}^{\mathrm{T}}$. Correspondingly,

$$
\log q_{\mathrm{A}^\circ}(\boldsymbol{x}_h^*) = -\frac{1}{2}(\boldsymbol{x}_h^* - \boldsymbol{\mu}_{\mathrm{A}^\circ})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_h^* - \boldsymbol{\mu}_{\mathrm{A}^\circ}) - \frac{1}{2}\log\det(\boldsymbol{\Sigma}) - \frac{3N_h}{2}\log 2\pi, \tag{S9}
$$

where $\det(\mathbf{\Sigma})$ is the determinant of the covariance matrix $\mathbf{\Sigma}$. A similar formula as Equation S9 applies to $\log q_{\mathrm{B}^\circ}(\boldsymbol{x}_h^*)$. The parameters $\{\mathbf{M}_{\mathrm{A}^\circ}, \boldsymbol{\mu}_{\mathrm{A}^\circ}\}$ and $\{\mathbf{M}_{\mathrm{B}^\circ}, \boldsymbol{\mu}_{\mathrm{B}^\circ}\}$ can be directly estimated based on host conformations sampled from state A and B, respectively. With estimated $\{\mathbf{M}_{\mathrm{A}^\circ}, \boldsymbol{\mu}_{\mathrm{A}^\circ}\}$ for state A° (or $\{\mathbf{M}_{\mathrm{B}^\circ}, \boldsymbol{\mu}_{\mathrm{B}^\circ}\}$ for state B°), we can easily calculate the value of $\log q_{\mathrm{A}^\circ}$ (or $\log q_{\mathrm{B}^\circ}$) for any conformation $\boldsymbol{x}_h^*$ using Equation S9 (or a similar equation for state B°). Moreover, we can draw independent samples of $\boldsymbol{x}_h^*$ from q_{A° (or q_{B°) by first sampling $\boldsymbol{u}_h^*$ from the standard normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$ and converting it into $\boldsymbol{x}_h^*$ using $\boldsymbol{x}_h^* = \mathbf{M}_{\mathrm{A}^\circ}\mathbf{u}_h^* + \boldsymbol{\mu}_{\mathrm{A}^\circ}$ (or $\boldsymbol{x}_h^* = \mathbf{M}_{\mathrm{B}^\circ}\mathbf{u}_h^* + \boldsymbol{\mu}_{\mathrm{B}^\circ}$).

Compared to the host, the guest molecules are more flexible. For state B, we modeled the distribution of their Cartesian coordinates $q_{\mathrm{B}^\circ}(\boldsymbol{x}_g^*)$, through the distribution of its internal coordinates $\boldsymbol{z}_g^*$. Specifically, we first parameterized the distribution of $\boldsymbol{z}_g^*$ by applying a series of bijective transformations to a uniformly distributed random variable $\boldsymbol{u}_g^*$, i.e.,

$$\boldsymbol{z}_g^* = T_K^*(\cdots(T_2^*(T_1^*(\boldsymbol{u}_g^*)))) \tag{S10}$$

where $\boldsymbol{u}_g^*$ has the same dimension as $\boldsymbol{z}_g^*$. $T_k^*(\cdot)$, for $k = 1, ..., K$, are bijective transformations. Then we computed the Cartesian coordinates $\boldsymbol{x}_g^*$ from $\boldsymbol{z}_g^*$ based on molecular topology. Let us represent this transformation as $\boldsymbol{x}_g^* = T_{K+1}^*(\boldsymbol{z}_g^*) = T(\boldsymbol{u}_g^*)$. Because we could also compute $\boldsymbol{z}_g^*$ from $\boldsymbol{x}_g^*$, the transformation T_{K+1} is also bijective. Using the formula of changing variables in probability density functions, we have

$$q_{\mathrm{B}^\circ}(\boldsymbol{x}_g^*) = p_u(\boldsymbol{u}_g^*)|\det J_T(\boldsymbol{u}_g^*)|^{-1} \tag{S11}$$

Since all the transformations are invertible and differentiable, the Jacobian determinant of the overall transformation T can be computed from the Jacobian determinant of individual

transformations as

$$\det J_T(\boldsymbol{u}_g^*) = \prod_{k=1}^{K+1} \det J_{T_k^*}(\boldsymbol{u}_g^{k-1}) \tag{S12}$$

Correspondingly,

$$\log q_{\mathrm{B}^\circ}(\boldsymbol{x}_g^*) = \log p_u(\boldsymbol{u}_g^*) - \sum_{k=1}^{K+1} \log |\det(J_{T_k^*}(\boldsymbol{u}_g^{k-1}))|, \tag{S13}$$

where $\boldsymbol{u}_g^k = T_k^*(...T_1^*(\boldsymbol{u}_g^0))$, $\boldsymbol{u}_g^0 = \boldsymbol{u}_g^*$, $\boldsymbol{z}_g^* = \boldsymbol{u}_g^K$ and $\det(J_{T_k^*}(\cdot))$ is the determinant of the Jacobian matrix $J_{T_k^*}$ of the transformation T_k^*.

Unlike those in state B, guest molecules in state A can also translate and rotate as a whole inside the binding site. Hence the distributions of their Cartesian coordinates, $q_{\mathrm{A}^\circ}(\boldsymbol{x}_g|\boldsymbol{x}_h^*)$ were modeled through the distributions of position, orientation and internal coordinates, the union of which is represented using $\boldsymbol{z}_g$. Similarly as state B°, $\boldsymbol{z}_g$ is modeled as the result of multiple bijective transformations of a uniform random variable $\boldsymbol{u}_g$. Because the distribution of $\boldsymbol{x}_g$ in $q_{\mathrm{A}^\circ}(\boldsymbol{x}_g|\boldsymbol{x}_h^*)$ is conditional on $\boldsymbol{x}_h^*$, we made the transformations between $\boldsymbol{u}_g$ and $\boldsymbol{z}_g$ dependent on $\boldsymbol{x}_h$ too, i.e.,

$$\boldsymbol{z}_g|\boldsymbol{x}_h = T_K(\cdots(T_1(T_0(\boldsymbol{u}_g;\boldsymbol{x}_h);\boldsymbol{x}_h));\boldsymbol{x}_h) \tag{S14}$$

where $T_k(\boldsymbol{u};\boldsymbol{x}_h)$ are bijective transformations of $\boldsymbol{u}$ that depends on $\boldsymbol{x}_h$. Based on guest molecules' topology, $\boldsymbol{z}_g$ can be reversibly converted into $\boldsymbol{x}_g$ for state A°, say with $\boldsymbol{x}_g = T_{K+1}(\boldsymbol{z}_g)$. Then we can compute $\log q_{\mathrm{A}^\circ}(\boldsymbol{x}_g|\boldsymbol{x}_h^*)$ using a similar equation as Equation S13 by replacing $T_k^*(\cdot)$ with $T_k(\cdot;\boldsymbol{x}_h^*)$ for $k = 1, ..., K+1$.

For both state A° and state B°, the bijective transformations of both $T_k(\cdot;\boldsymbol{x}_h^*)$ and $T_k^*(\cdot)$, for $k = 1, ..., K$, were modeled using rational quadratic neural spline flows with coupling layers (RQ-NSF(C))[S3–S5] and learned based on samples from state A and state B, respectively. The detailed architecture of the RQ-NSF(C) layer is as follows. To simplify the notation, we will use $\boldsymbol{u}_g$ and $\boldsymbol{v}_g$ as the input and output of the RQ-NSF(C) layer, respectively. In

each RQ-NSF(C) layer, components of the input vector $\boldsymbol{u}_g$ were randomly split into two groups: $\boldsymbol{u}_g^I$ and $\boldsymbol{u}_g^C$. The components in $\boldsymbol{u}_g^I$ were kept the same in the transformation, i.e., the corresponding components of the output vector, $\boldsymbol{v}_g^I$, is equal to $\boldsymbol{u}_g^I$. The components in $\boldsymbol{u}_g^C$ were transformed by applying a component-wise monotonic rational-quadratic splines,[S4] where both positions of knots and derivatives on the knots for the spline were parameterized using residual neural networks[S6] with $\boldsymbol{u}_g^I$ and $\boldsymbol{x}_h^*$ as the input for $T_k(\cdot;\boldsymbol{x}_h^*)$ and with just $\boldsymbol{u}_g^I$ as the input for $T_k^*(\cdot)$ (Fig. S2). In detail, the transformation applied by a RQ-NSF(C) layer is as follows:

$$\boldsymbol{v}_g^I = \boldsymbol{u}_g^I \tag{S15}$$

$$\boldsymbol{v}_g^C = g_\theta(\boldsymbol{u}_g^C) \tag{S16}$$

where $\boldsymbol{u}_g = (\boldsymbol{u}_g^I, \boldsymbol{u}_g^C)$ and $\boldsymbol{v}_g = (\boldsymbol{v}_g^I, \boldsymbol{v}_g^C)$ are the input and the output of the RQ-NSF(C) layer, respectively, and the transformation g_θ is a component-wise monotonic rational-quadratic splines with 4 knots. The parameters θ of the spline include positions of the 4 knots and derivatives on the 4 knots and these parameters are parameterized using a neural network. For the transformation $T_k(\cdot;\boldsymbol{x}_h^*)$ used in state A°, the input to the neural network include both $\boldsymbol{u}_g^I$ and $\boldsymbol{x}_h^*$, i.e., $\theta = \mathrm{NN}(\boldsymbol{u}_g^I, \boldsymbol{x}_h^*)$. For the transformation $T_k(\cdot)$ used in state B°, the input to the neural network is just $\boldsymbol{u}_g^I$, i.e., $\theta = \mathrm{NN}(\boldsymbol{u}_g^I)$. The Jacobian matrix of the RQ-NSF(C) layer is a triangular matrix, so the value of its determinant is the product of its diagonal terms and the diagonal terms can be calculated using the backpropagation algorithm. Because g_θ is a component-wise monotonic rational-quadratic splines, its inverse g_θ^{-1} can be solved analytically. With g_θ^{-1}, the inverse of the RQ-NSF(C) layer can be evaluated as follows:

$$\boldsymbol{u}_g^I = \boldsymbol{v}_g^I \tag{S17}$$

$$\boldsymbol{u}_g^C = g_\theta^{-1}(\boldsymbol{v}_g^C), \tag{S18}$$

where $\theta = \mathrm{NN}(\boldsymbol{v}_g^I, \boldsymbol{x}_h^*)$ for $T_k(\cdot;\boldsymbol{x}_h^*)$ and $\theta = \mathrm{NN}(\boldsymbol{v}_g^I)$ for $T_k(\cdot)$. For more details on construct-

ing monotonic rational-quadratic splines and its inverse, see the reference S4 and S7.

Bijective transformations between Cartesian and internal coordinates

From simulated configurations of state B, we removed the translational and rotational degrees of freedom of guest molecules to obtain $\boldsymbol{x}_g^*$ as follows. Three reference particles, represented as R1, R2 and R3, were chosen from the guests' atoms such that R1 is bonded to R2 and R3. Based on the three reference particles, we aligned their conformations such that (1) the position of R1 is (0.0, 0.0, 0.0); (2) the position of R2 is (0.0, 0.0, z_{R2}); (3) the position of R3 is (0.0, y_{R3}, z_{R3}). After alignment, the x, y, z coordinates of R1, the x, y coordinates of R2 and the x coordinates of R3 were all fixed at zeros. The Cartesian coordinates $\boldsymbol{x}_g^*$ include z_{R2}, y_{R3}, z_{R3} and the x, y, z coordinates of other guest atoms that were not chosen as reference particles.

With the three reference particles specified as above, we can also represent the guest conformations using internal coordinates $\boldsymbol{z}_g^*$. z_{R2} can be determined using the length of the bond R1-R2 ($b_{\mathrm{R1-R2}}$). y_{R3} and z_{R3} can be specified using the length of the bond R1-R3 ($b_{\mathrm{R1-R3}}$) and the angle R2-R1-R3 ($a_{\mathrm{R2-R1-R3}}$). Therefore, the Cartesian coordinates of R1, R2 and R3 can be specified using the internal coordinates ($b_{\mathrm{R1-R2}}$, $b_{\mathrm{R1-R3}}$, $a_{\mathrm{R2-R1-R3}}$). To specify the Cartesian coordinates of other guest atoms using internal coordinates, we label R1, R2 and R3 particles as atom 1, 2 and 3 respectively, and incrementally label other atoms in the order of querying by a depth-first-search with R1 as the root node and the guest molecular topology as the tree structure. The internal coordinates of atom $l (l \geq 4)$ include the bond length b_l, the angle a_l, and the dihedral angle d_l. If atom l is directly bonded with R1, its internal coordinates are defined with respect to R1, R2 and R3, i.e., the bond length of l-R1, the angle of l-R1-R2 and the dihedral of l-R1-R2-R3. If atom l is directly bonded with R2 (or R3), its internal coordinates are defined with respect to R2 (or R3), R1, and R3 (or R2). For atom l that is not directly bonded to R1, R2 or R3, its

coordinates are defined with respect to atoms i, j, k such that $i < j < k < l$ and the four atoms are consecutively bonded, i.e., atom i is bonded with atom j which is bonded with atom k and atom k is further bonded with atom l. Because atoms are labeled using the depth-first-search order, we can alway identify atoms i, j, k that satisfy these two conditions for atom l that is not directly bonded to R1, R2 or R3. Therefore, the internal coordinates $\boldsymbol{z}_g^*$ of a guest include $b_{\mathrm{R1-R2}}$, $b_{\mathrm{R1-R3}}$, $a_{\mathrm{R2-R1-R3}}$ and the bond b_l, the angle a_l, and the dihedral angle d_l of atom l for $l \geq 4$.

We included the full set of Cartesian coordinates to model guest molecules in the bound state A, and the corresponding $\boldsymbol{z}_g$ was defined as follows. Similarly to the case for state B, three reference particles, represented as R1, R2 and R3, were chosen from the guests' atoms and other guest atoms were labeled using the same depth-first-search order. The position of R1 was directly specified using its xyz coordinates as in $\boldsymbol{x}_g$, i.e., $(x_{\mathrm{R1}}, y_{\mathrm{R1}}, z_{\mathrm{R1}})$. The position of R2 was specified using a spherical coordinate system with R1 as the origin, i.e., using r_{R2}, θ_{R2}, and φ_{R2} which are the radial distance, polar angle, and azimuthal angle of R2, respectively. The position of R3 could be similarly determined using r_{R3}, θ_{R3}, and φ_{R3} with R1 as the origin. After specifying the positions of R1, R2 and R3, we defined the position of atom l using the bond length b_l, the angle a_l, and the dihedral angle d_l with respect to atoms i, j, k as in the case for state B. Therefore, the coordinates $\boldsymbol{z}_g$ for state A include $x_{\mathrm{R1}}, y_{\mathrm{R1}}, z_{\mathrm{R1}}$, r_{R2}, θ_{R2}, φ_{R2}, r_{R3}, θ_{R3}, φ_{R3} and b_l, a_l, and d_l of atom l for $l \geq 4$.

Details on PMF calculations

As mentioned in the main text, intermediate states for PMF calculations were designed in three phases. *In the first, attachment phase*, a restraining potential was applied to the host CB7 to slowly enlarge its cavity portal. A positional restraining potential was also applied on the guest molecules to keep them inside the binding pocket. *In the second, pulling phase*, the guest molecules were restrained at increasing distances from the host along the axis that

passes through the host's center of mass and is perpendicular to the host's middle section plane. In addition, the orientation of guest molecules was also restrained to the same pose they adopt when bound with the host. The cavity opening restraints introduced in the first phase facilitates the exit of guest molecules. In the last state of the pulling phase, the guest molecules were positioned far enough from the host so that there are no interactions between them. *In the final, release phase*, the restraining potential on the host that was turned on during the attachment phase and kept on during the pulling phase, was slowly turned off.

In addition to the various restraining potentials in different phases, all PMF simulations were subject to the restraints on translational and rotational degrees of freedom specified in *Section: Simulation Details for Bound and Unbound States.*

Restraining potential for host cavity opening

A restraining potential was applied to the host to enlarge its cavity portal and facilitate guest existing. Towards that end, we defined a virtual site H4 on the host as the average position of the seven carbon atoms that are bonded to oxygen atoms around the cavity portal and on the opposite side of the host from the anchor particles (Figure S26). Harmonic biasing potentials were applied on the distances (d, Figure S26) of H4 from each of the seven carbon atoms that were used in determining the position of H4. The form of the potentials is $\frac{k}{2} \cdot (d - d^c)^2$ with $k = 5 \times 10^4$. d^c is the equilibrium distance that the potential is biasing towards.

For the 10 intermediate states from the attachment phase, the equilibrium distance, d^c, was linearly increased from 0.48 to 0.55 nm. d_c was kept at 0.55 nm for all the 75 intermediate states of the pulling phase. During the release phase, d_c was linearly decreased from 0.55 to 0.48 nm for the 10 intermediate states.

Restraining potential on guests' position and orientation

For all intermediate states in the attachment phase, a biasing potential was applied on the guest molecules to keep them inside the binding pocket (Figure S27). The potential adopts the form of $\frac{k}{2} \cdot \mathcal{H}(d_{g_1} - d^c_{g_1}) \cdot (d_{g_1} - d^c_{g_1})^2$, with $k = 2000$. The Heaviside step function $\mathcal{H}(x)$ is 1 when $x \geq 0$ and 0 when $x < 0$. d_{g_1} is the distance between G1 and P1. The virtual site G1 and the value of $d^c_{g_1}$ for the four guest molecules were chosen as follows:

- GI: G1 is the average position of the six carbon atoms and $d^c_{g_1} = 0.60$ nm.
- GII: G1 is the average position of the six carbon atoms in the benzene ring and $d^c_{g_1} = 0.575$ nm.
- GIII: G1 is the average position of the ten spiro carbon atoms and $d^c_{g_1} = 0.56$ nm.
- GIV: G1 is the average position of the ten spiro carbon atoms and $d^c_{g_1} = 0.58$ nm.

During the pulling and release phases, the biasing potential on guest molecules used in the attachment phase was removed. New biasing potentials were defined to restrain guests at specific positions and orientations using two virtual sites, G1 and G2 (Figure S27). Biasing potentials of the following forms were applied on the distance d_{g_1} between G1 and P1, the angle θ_{g_1} between G1, P1, and P3, and the angle θ_{g_2} between G2, G1, and P1:

$$\begin{aligned} \frac{1}{2} \cdot 2000 \cdot (d_{g_1} - d^c_{g_1})^2 \text{ for } d_{g_1}, \\ \frac{1}{2} \cdot 1000 \cdot (\theta_{g_1} - \theta^c_{g_1})^2 \text{ for } \theta_{g_1}, \\ \mathcal{H}(\theta^c_{g_2} - \theta_{g_2}) \cdot \frac{1}{2} \cdot 200 \cdot (\theta_{g_2} - \theta^c_{g_2})^2 \text{ for } \theta_{g_2}. \end{aligned}$$

The definition of the virtual site G1 is the same as that used in the attachment phase. The virtual site G2 and the values for $d^c_{g_1}$, $\theta^c_{g_1}$, and $\theta^c_{g_2}$ were chosen as follows:

- GI: G2 is the carbon atom that is connected to a nitrogen atom; In both the pulling and the release phase, $\theta^c_{g_1} = 3.07$ and $\theta^c_{g_2} = 2.70$. In the pulling phase, $d^c_{g_1}$ linearly increases

from 0.53 to 1.5nm over the first 40 intermediate states, and further linearly increases to 2.8nm over the rest of 35 intermediate states. In the release phase, $d^c_{g_1} = 2.8$ nm.

- GII: G2 is the carbon atom that is connected to a nitrogen atom; In both the pulling and the release phase, $\theta^c_{g_1} = 3.08$ and $\theta^c_{g_2} = 2.90$. In the pulling phase, $d^c_{g_1}$ linearly increases from 0.50 to 1.5nm over the first 40 intermediate states, and further linearly increases to 2.8nm over the rest of 35 intermediate states. In the release phase, $d^c_{g_1} = 2.8$ nm.

- GIII: G2 is the carbon atom that is connected to oxygen atoms; In both the pulling and the release phase, $\theta^c_{g_1} = 3.11$ and $\theta^c_{g_2} = 2.90$. In the pulling phase, $d^c_{g_1}$ linearly increases from 0.53 to 1.5nm over the first 40 intermediate states, and further linearly increases to 2.8nm over the rest of 35 intermediate states. In the release phase, $d^c_{g_1} = 2.8$ nm.

- GIV: G2 is the nitrogen atom that is connected to one of the spiro carbon atoms; In both the pulling and the release phase, $\theta^c_{g_1} = 3.11$ and $\theta^c_{g_2} = 2.70$. In the pulling phase, $d^c_{g_1}$ linearly increases from 0.51 to 1.5nm over the first 40 intermediate states, and further linearly increases to 2.8nm over the rest of 35 intermediate states. In the release phase, $d^c_{g_1} = 2.8$ nm.

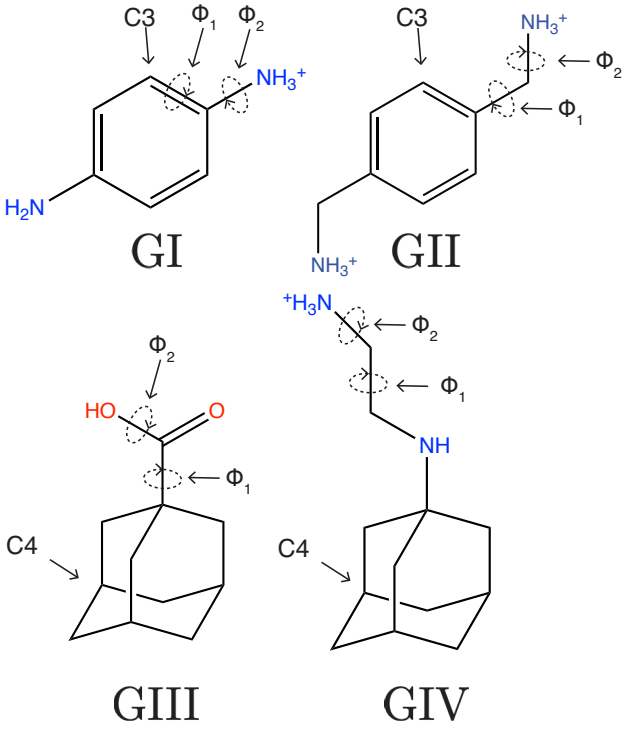

Figure S1: The specific atoms and dihedral angles of guest molecules mentioned in both the main text and the Supporting Information.

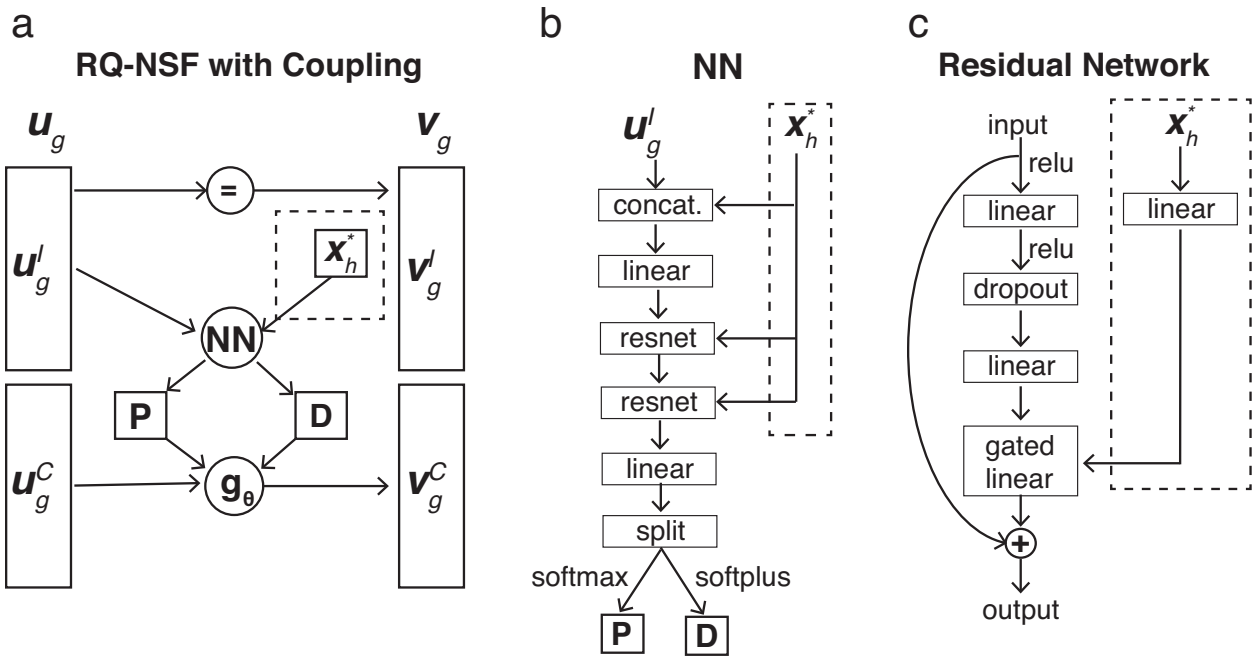

Figure S2: Illustration of the RQ-NSF(C) layer used to define the bijective transformations of $T_k(\cdot; \boldsymbol{x}_h^*)$ and $T_k^*(\cdot)$ for state A° and state B°, respectively. The transformation $T_k(\cdot; \boldsymbol{x}_h^*)$ depends on $\boldsymbol{x}_h^*$ and includes the components enclosed in the dashed box, whereas the transformation $T_k^*(\cdot)$ does not. (a) The input random variable $\boldsymbol{u}_g$ is randomly split into two components, $\boldsymbol{u}_g^I$ and $\boldsymbol{u}_g^C$. $\boldsymbol{u}_g^I$ is kept identical through the layer and $\boldsymbol{u}_g^C$ is transformed via a component-wise monotonic rational quadratic spline $\boldsymbol{g}_\theta$. The spline is defined using 4 knots and the parameters θ of the spline include the positions (P) of the 4 knots and derivatives (D) on the 4 knots. (b) The knots positions (P) and derivatives (D) are parameterized using a neural network (NN) with both $\boldsymbol{u}_g^I$ and $\boldsymbol{x}_h^*$ as inputs for $T_k(\cdot; \boldsymbol{x}_h^*)$ and with just $\boldsymbol{u}_g^I$ as input for $T_k^*(\cdot)$. The concatenation layer concatenates its two input vectors into one vector and the split layer does the opposite, splitting a vector into two vectors. The component labeled as resnet is a residue network building block and its detailed architecture is shown in (c).

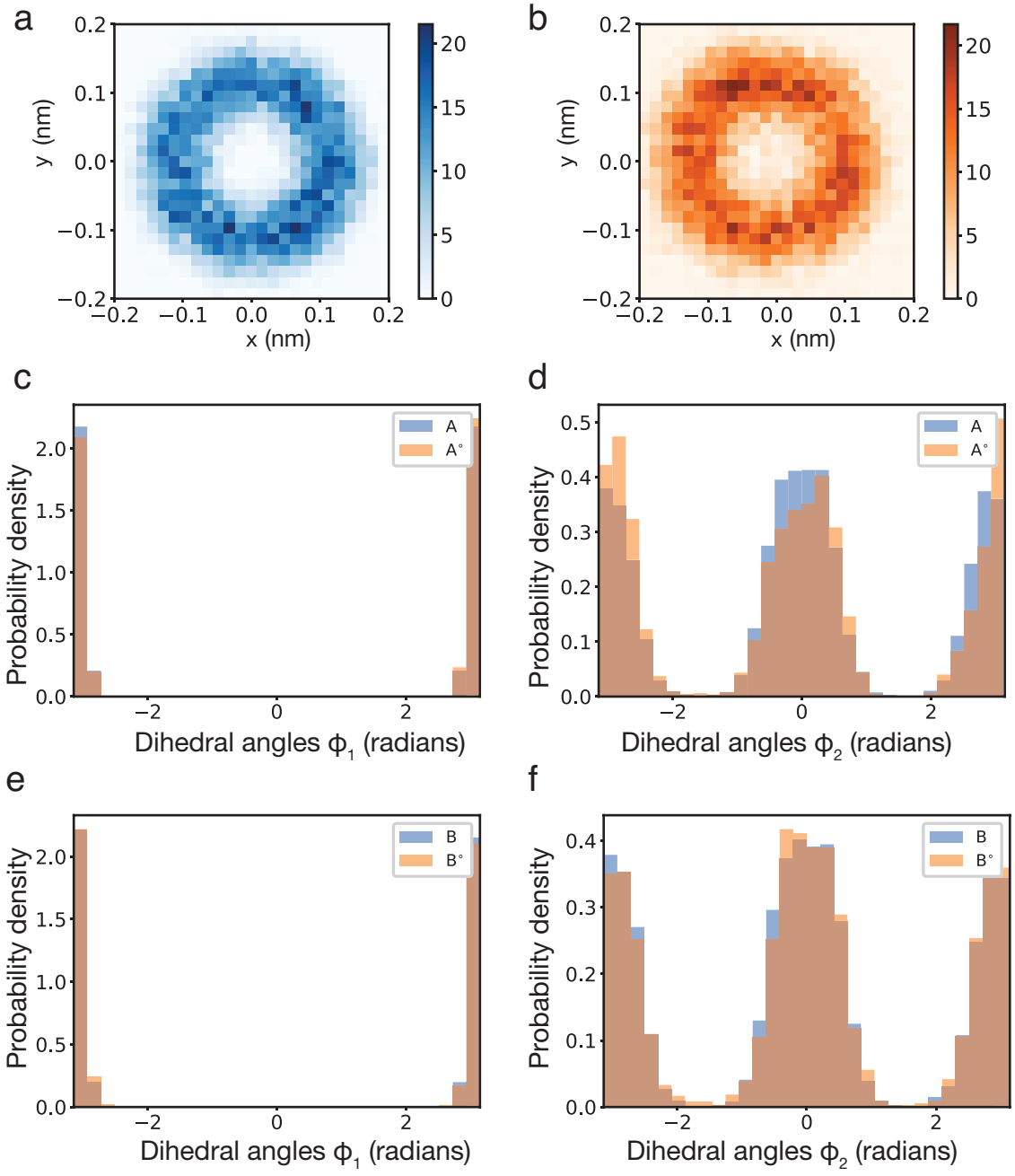

Figure S3: Marginal distributions of specific degrees of freedom from state A/B and state A°/B° for the CB7-GI host-guest system. (a and b) The distribution of x-y coordinates of atom C3 for state A (a) and state A° (b). (c) The distribution of the dihedral angle ϕ_1 for state A (blue) and state A° (orange). (d) The distribution of the dihedral angle ϕ_2 for state A (blue) and state A° (orange). (e) The distribution of the dihedral angle ϕ_1 for state B (blue) and state B° (orange). (f) The distribution of the dihedral angle ϕ_2 for state B (blue) and state B° (orange).

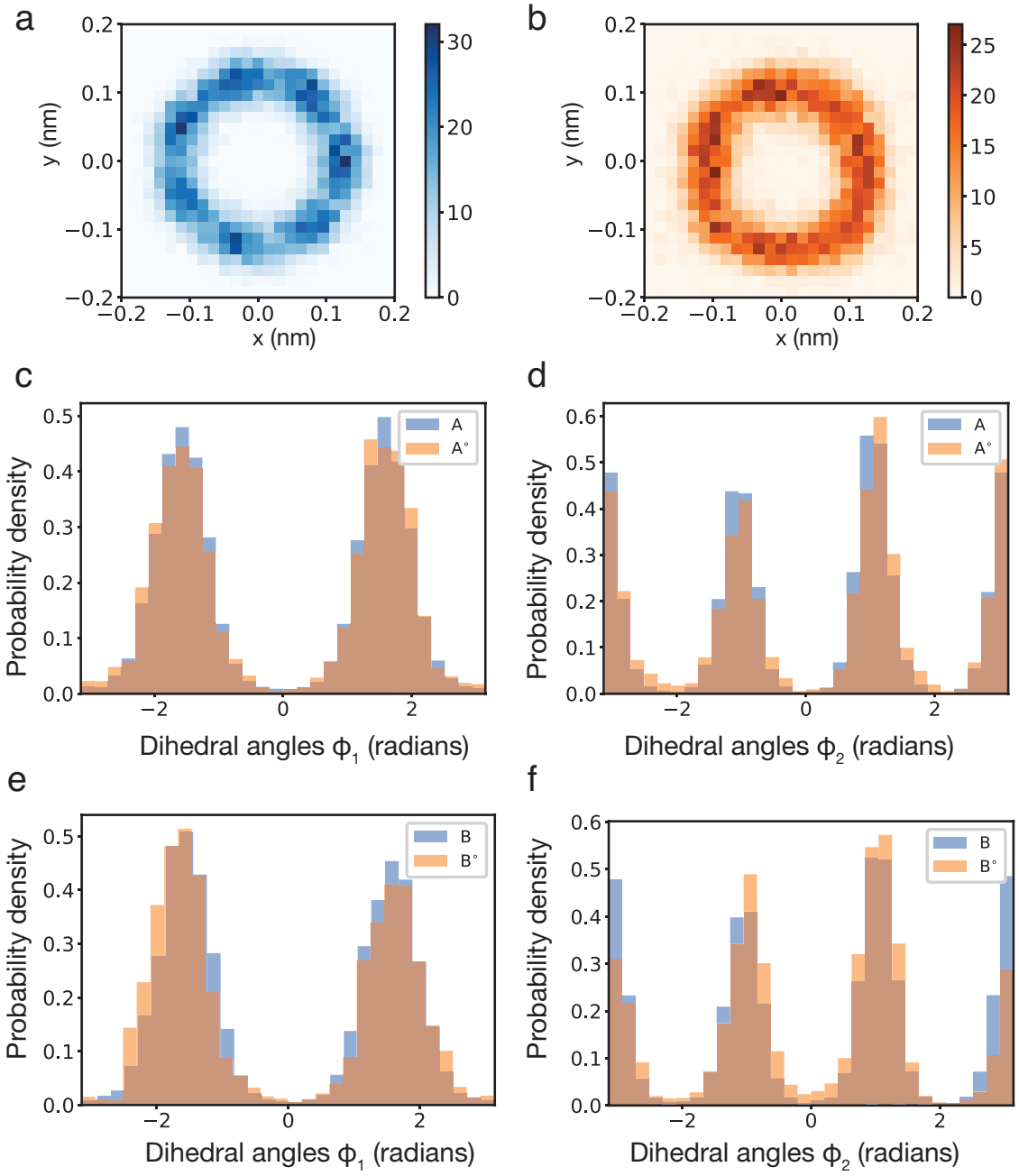

Figure S4: Marginal distributions of specific degrees of freedom from state A/B and state A°/B° for the CB7-GII host-guest system. (a and b) The distribution of x-y coordinates of atom C3 for state A (a) and state A° (b). (c) The distribution of the dihedral angle ϕ_1 for state A (blue) and state A° (orange). (d) The distribution of the dihedral angle ϕ_2 for state A (blue) and state A° (orange). (e) The distribution of the dihedral angle ϕ_1 for state B (blue) and state B° (orange). (f) The distribution of the dihedral angle ϕ_2 for state B (blue) and state B° (orange).

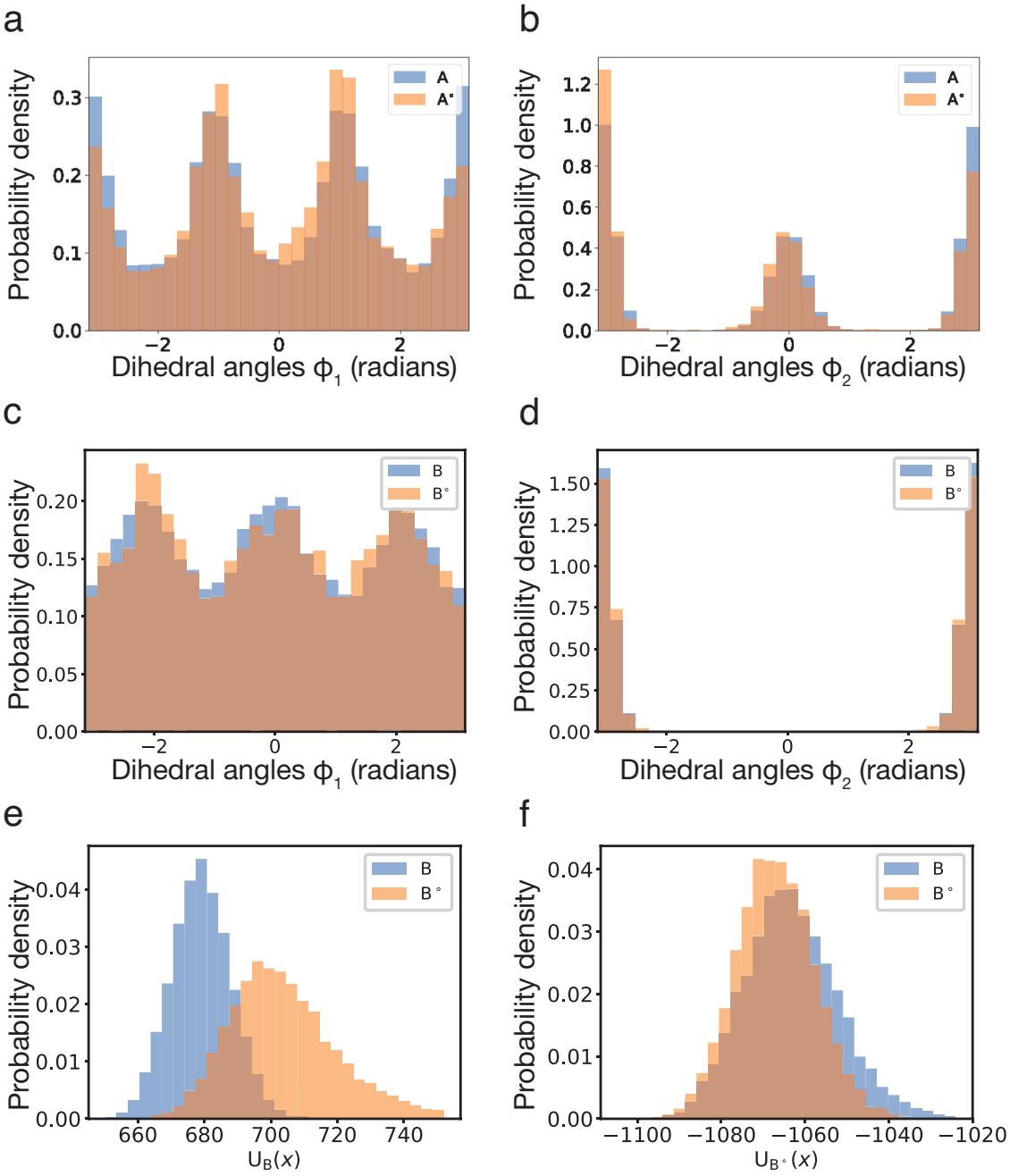

Figure S5: Marginal distributions of specific degrees of freedom from state A/B and state A°/B° for the CB7-GIII host-guest system. (a) The distribution of the dihedral angle ϕ_1 for state A (blue) and state A° (orange). (b) The distribution of the dihedral angle ϕ_2 for state A (blue) and state A° (orange). (c) The distribution of the dihedral angle ϕ_1 for state B (blue) and state B° (orange). (d) The distribution of the dihedral angle ϕ_2 for state B (blue) and state B° (orange). (e and f) Distributions of the value (in kcal/mol) of energy functions U_{B} (e) and U_{B° (f) on samples from state B (blue) and state B° (orange).

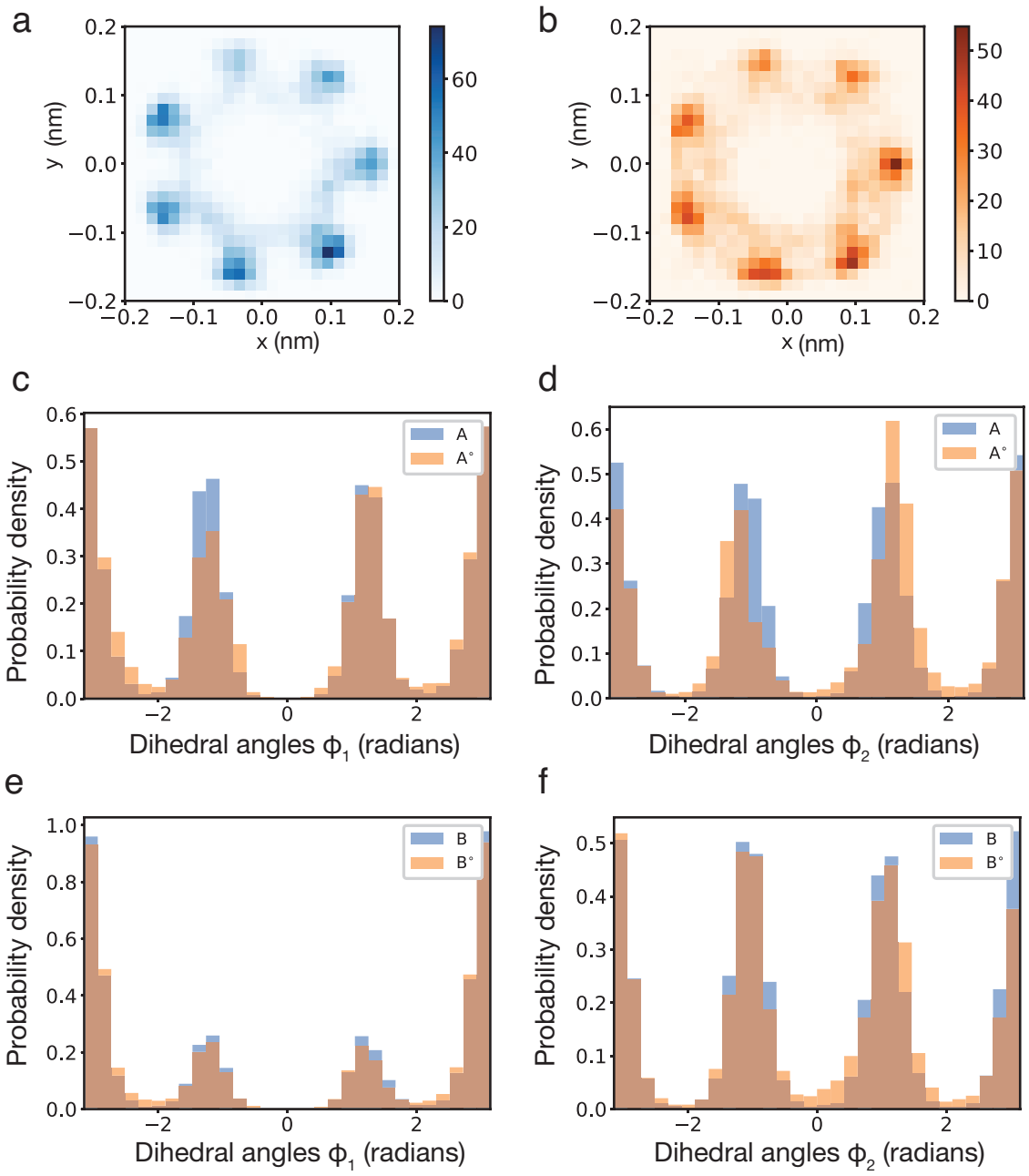

Figure S6: Marginal distributions of specific degrees of freedom from state A/B and state A°/B° for the CB7-GIV host-guest system. (a and b) The distribution of x-y coordinates of atom C4 for state A (a) and state A° (b). (c) The distribution of the dihedral angle ϕ_1 for state A (blue) and state A° (orange). (d) The distribution of the dihedral angle ϕ_2 for state A (blue) and state A° (orange). (e) The distribution of the dihedral angle ϕ_1 for state B (blue) and state B° (orange). (f) The distribution of the dihedral angle ϕ_2 for state B (blue) and state B° (orange).

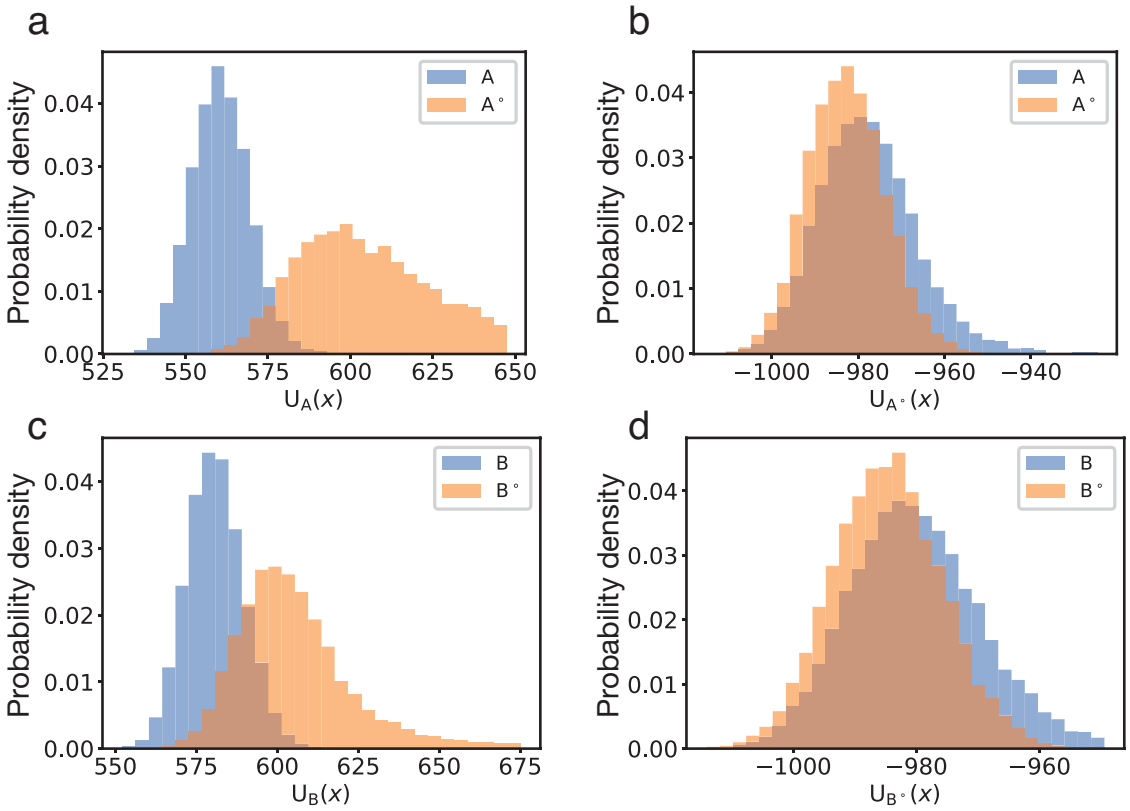

Figure S7: Energy function overlap between target state A/B and reference state A°/B° for the guest GI. (a and b) Distributions of the value (in kcal/mol) of energy functions U_{A} (a) and U_{A° (b) on samples from state A (blue) and state A° (orange). (c and d) Distributions of the value (in kcal/mol) of energy functions U_{B} (c) and U_{B° (d) on samples from state B (blue) and state B° (orange).

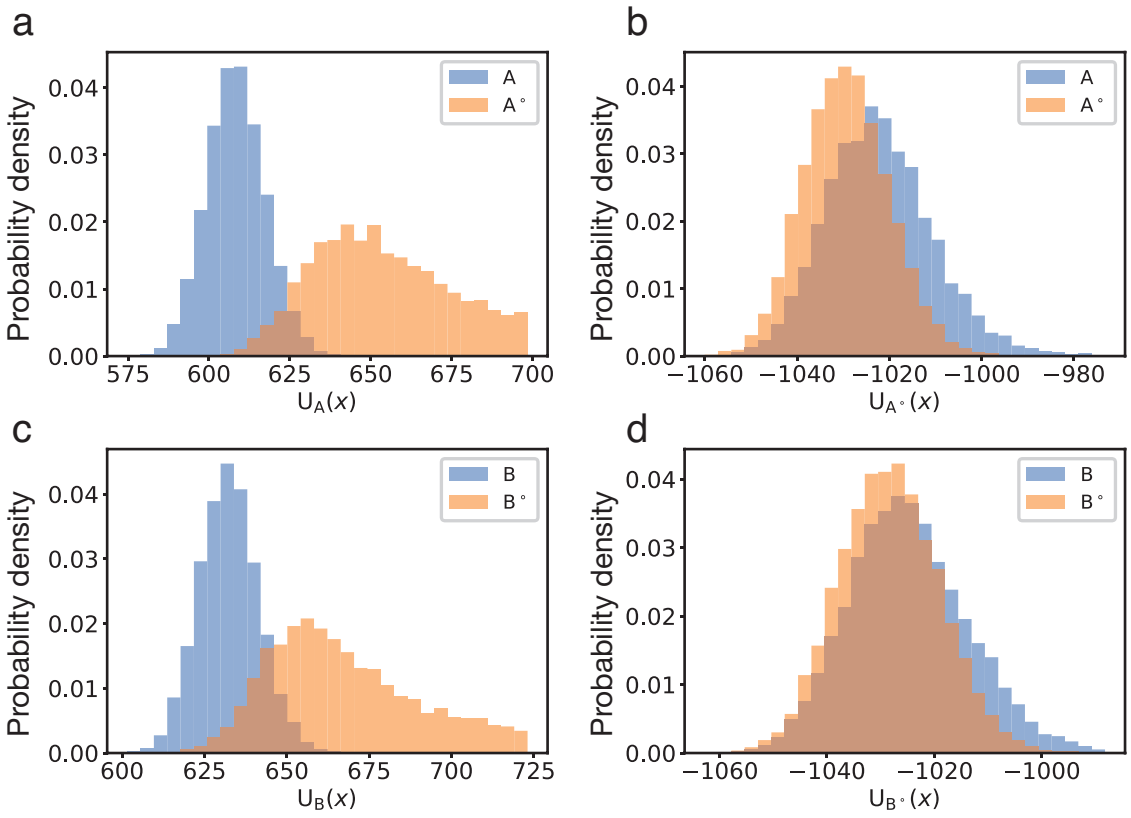

Figure S8: Energy function overlap between target state A/B and reference state A°/B° for the guest GII. (a and b) Distributions of the value (in kcal/mol) of energy functions U_{A} (a) and U_{A° (b) on samples from state A (blue) and state A° (orange). (c and d) Distributions of the value (in kcal/mol) of energy functions U_{B} (c) and U_{B° (d) on samples from state B (blue) and state B° (orange).

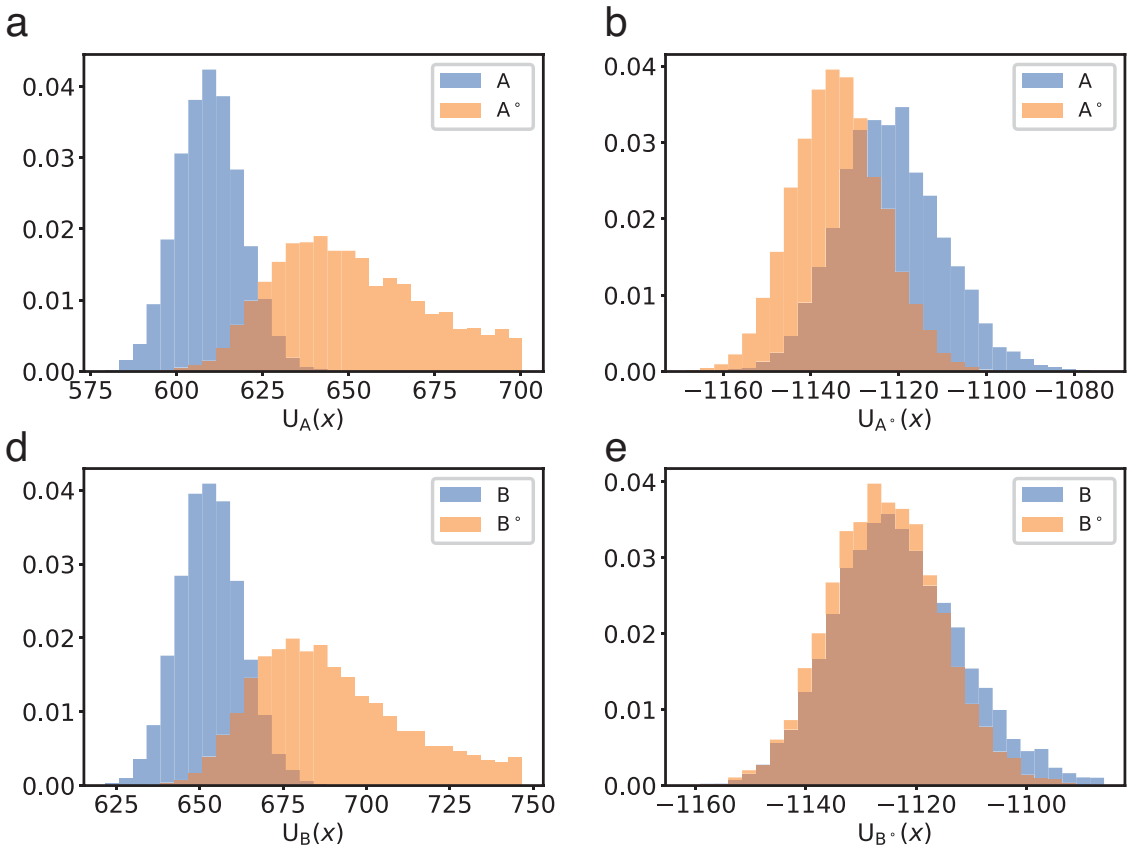

Figure S9: Energy function overlap between target state A/B and reference state A°/B° for the guest GIV. (a and b) Distributions of the value (in kcal/mol) of energy functions U_{A} (a) and U_{A° (b) on samples from state A (blue) and state A° (orange). (c and d) Distributions of the value (in kcal/mol) of energy functions U_{B} (c) and U_{B° (d) on samples from state B (blue) and state B° (orange).

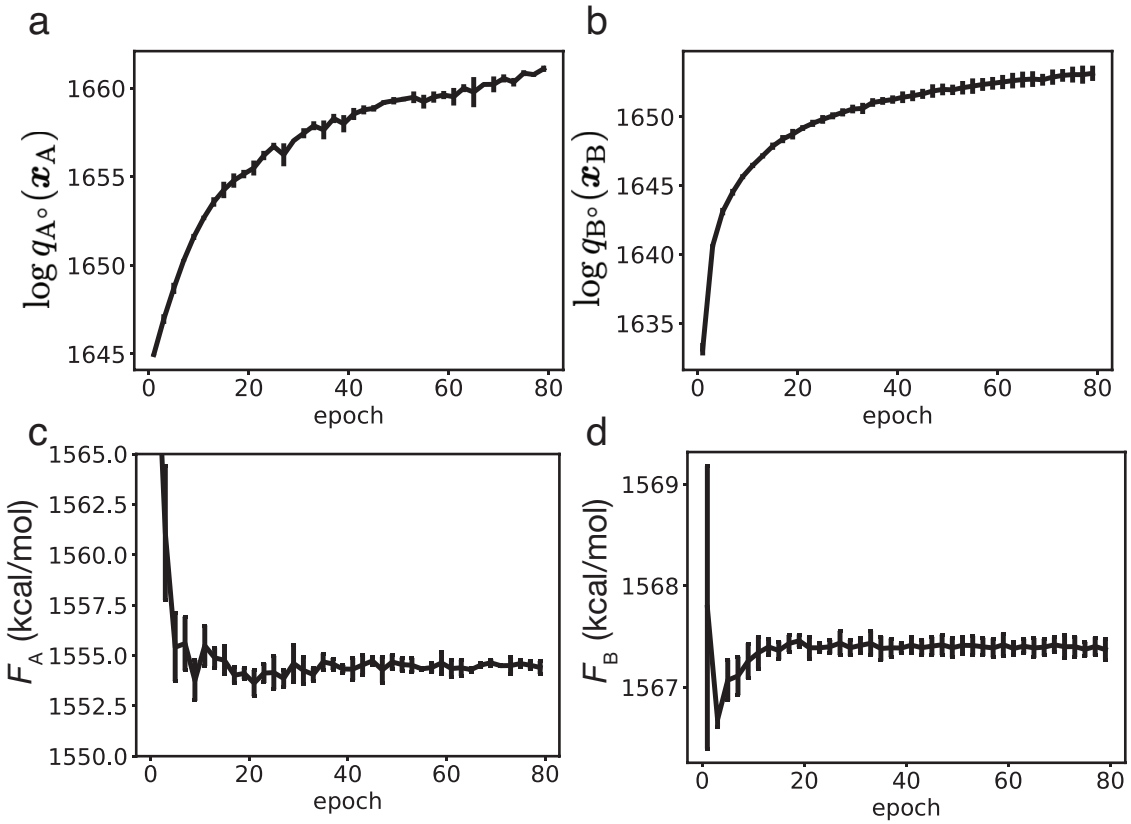

Figure S10: Similar results as in Figure 4 of the main text but for guest GI.

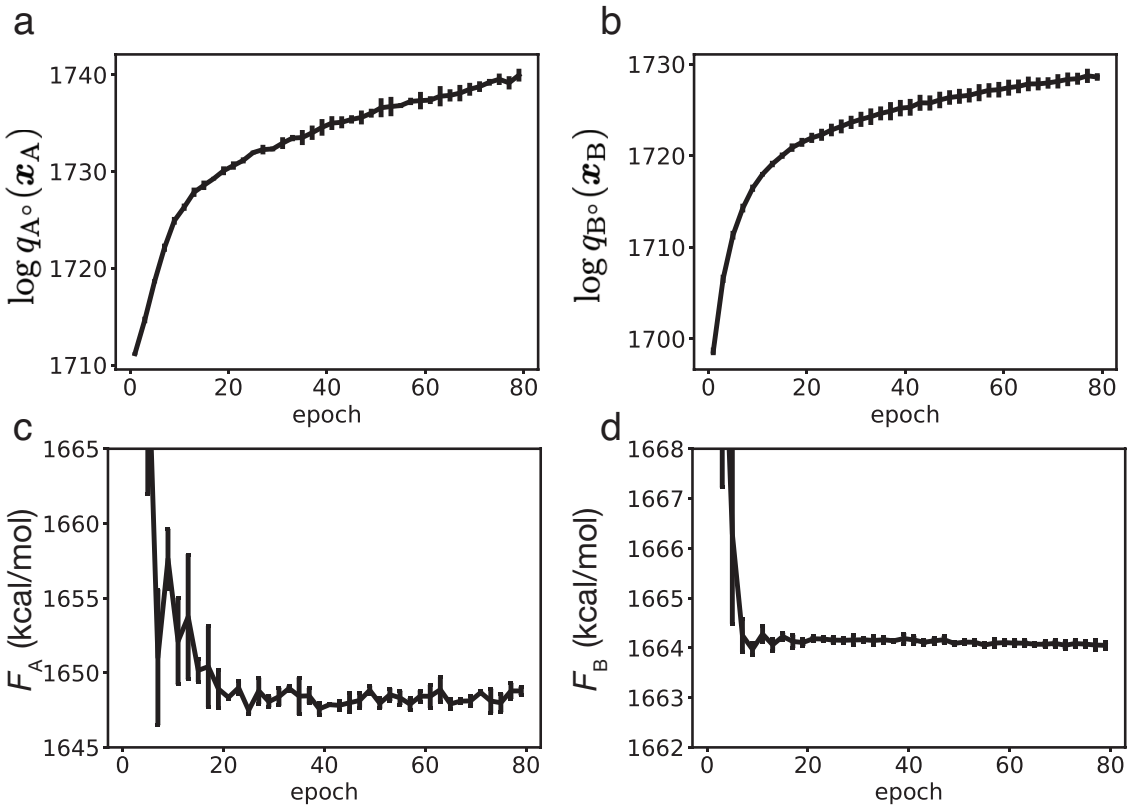

Figure S11: Similar results as in Figure 4 of the main text but for guest GII.

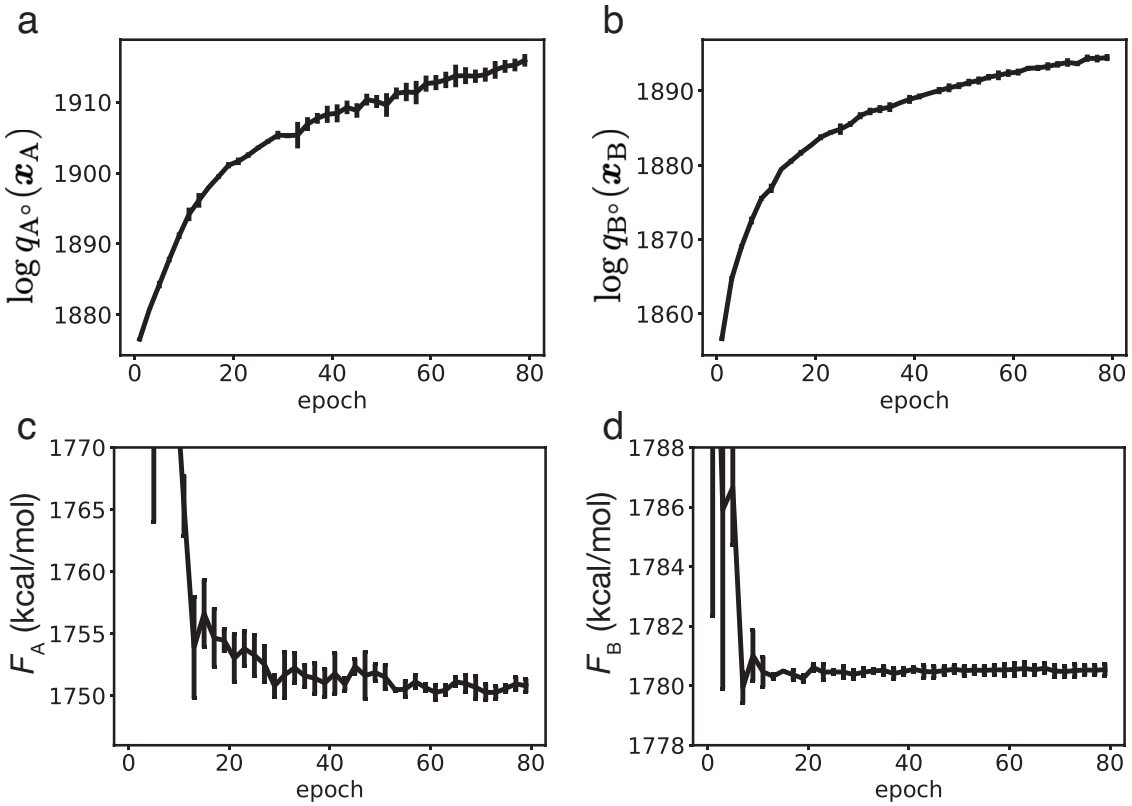

Figure S12: Similar results as in Figure 4 of the main text but for guest GIV.

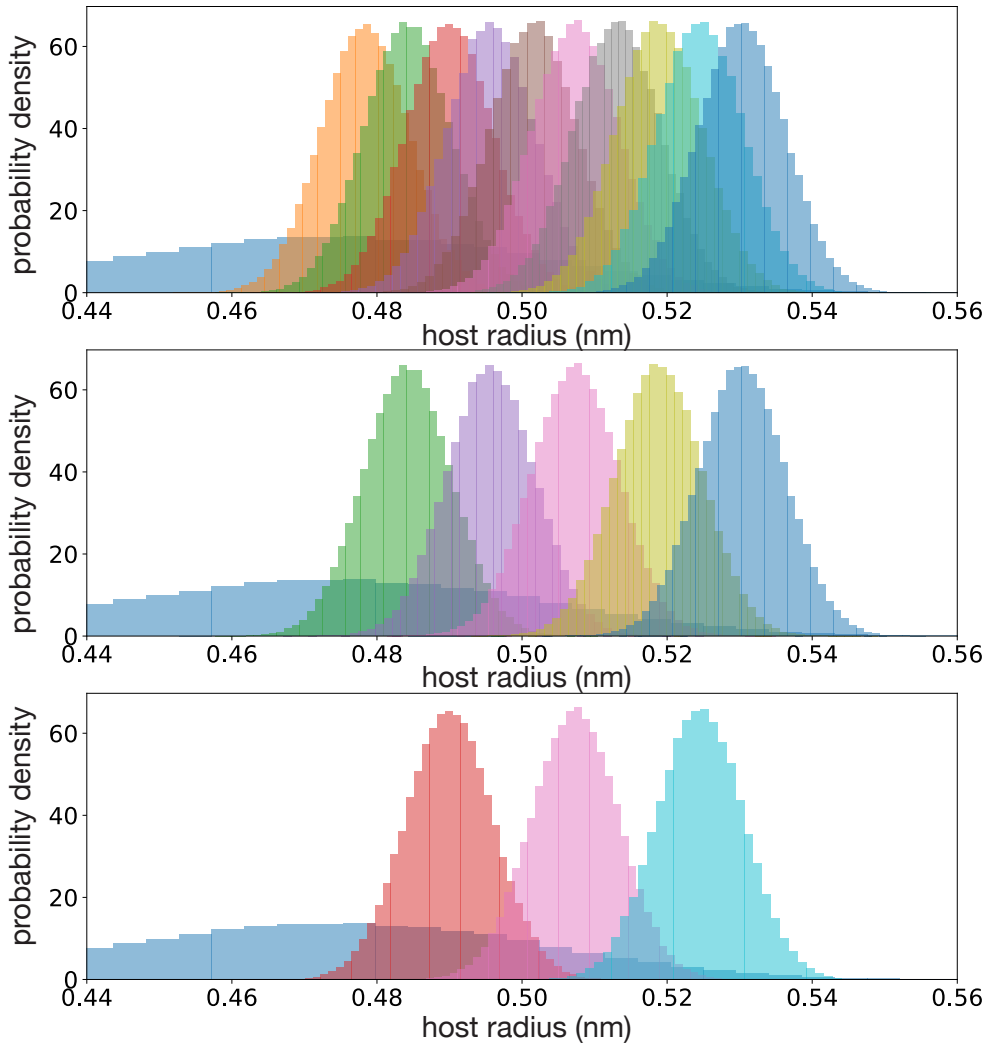

Figure S13: Distributions of the host's radius for states in the attachment phase of the PMF method with 97 windows (top), 49 windows (middle), and 33 windows (bottom) for the guest GI. The most left distribution is for state A and other distributions are for intermediate states of the attachment phase.

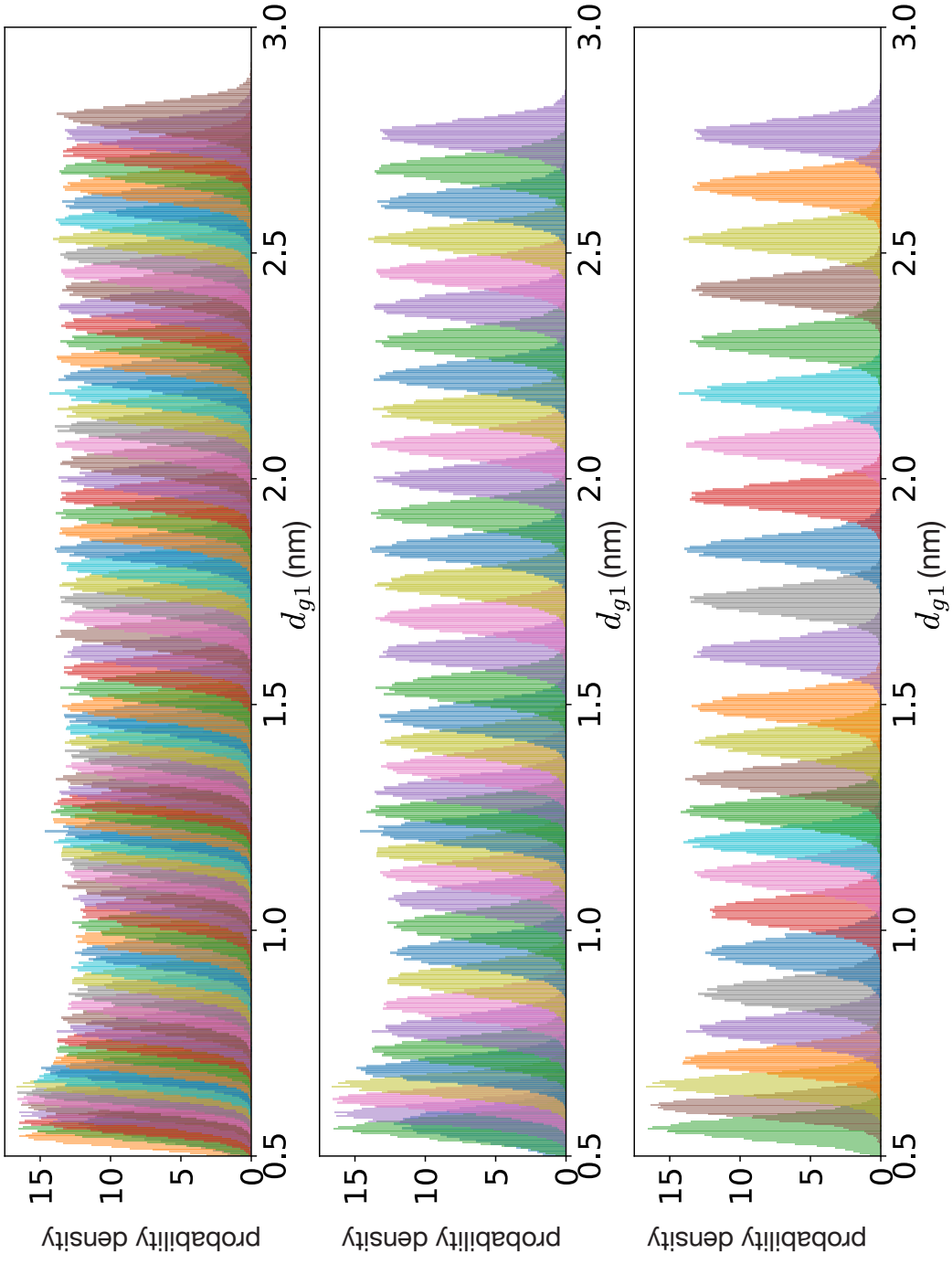

Figure S14: Distributions of the distance d_{g_1} between the guest GI's G1 site and the anchor particle P1 for states in the pulling phase of the PMF method with 97 windows (left), 49 windows (middle), and 33 windows (right).

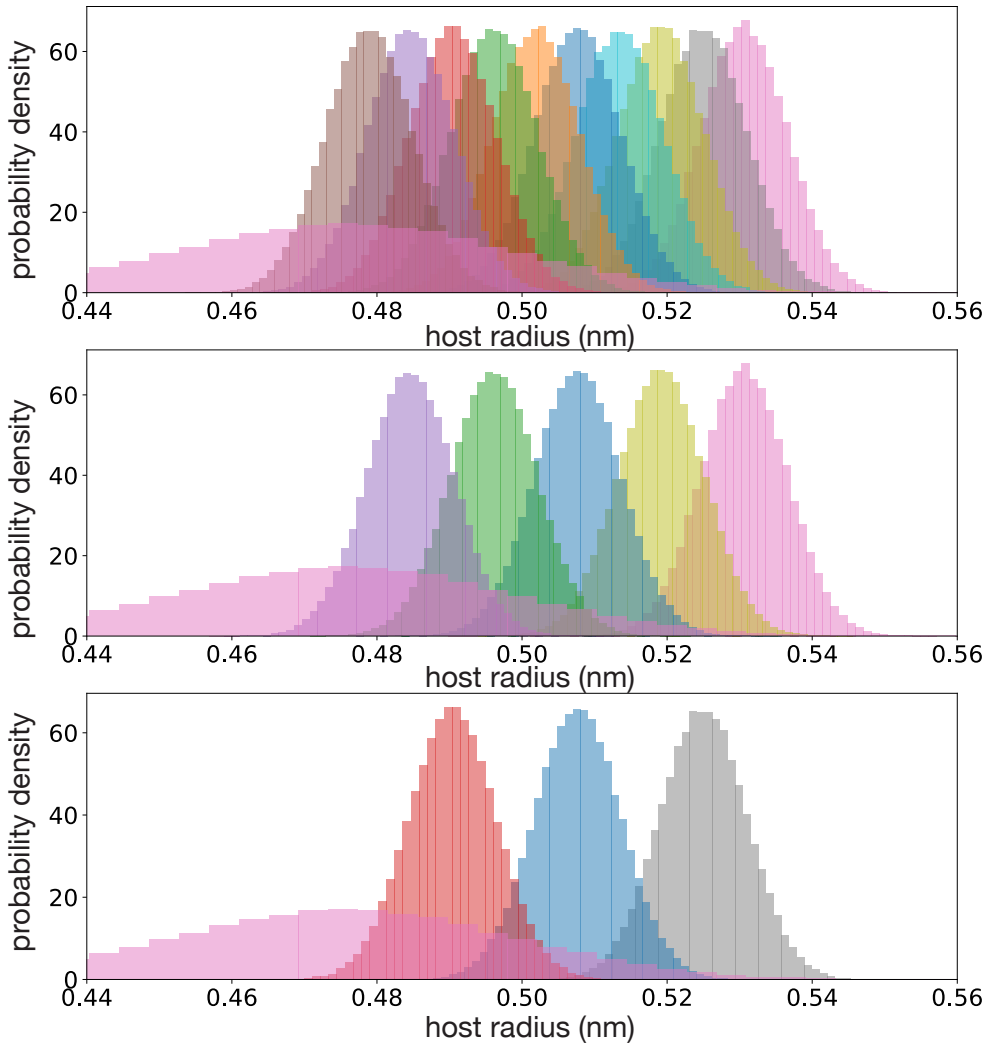

Figure S15: Distributions of the host's radius for states in the release phase of the PMF method with 97 windows (top), 49 windows (middle), and 33 windows (bottom) for the guest GI. The most left distribution is for state B and other distributions are for intermediate states of the release phase.

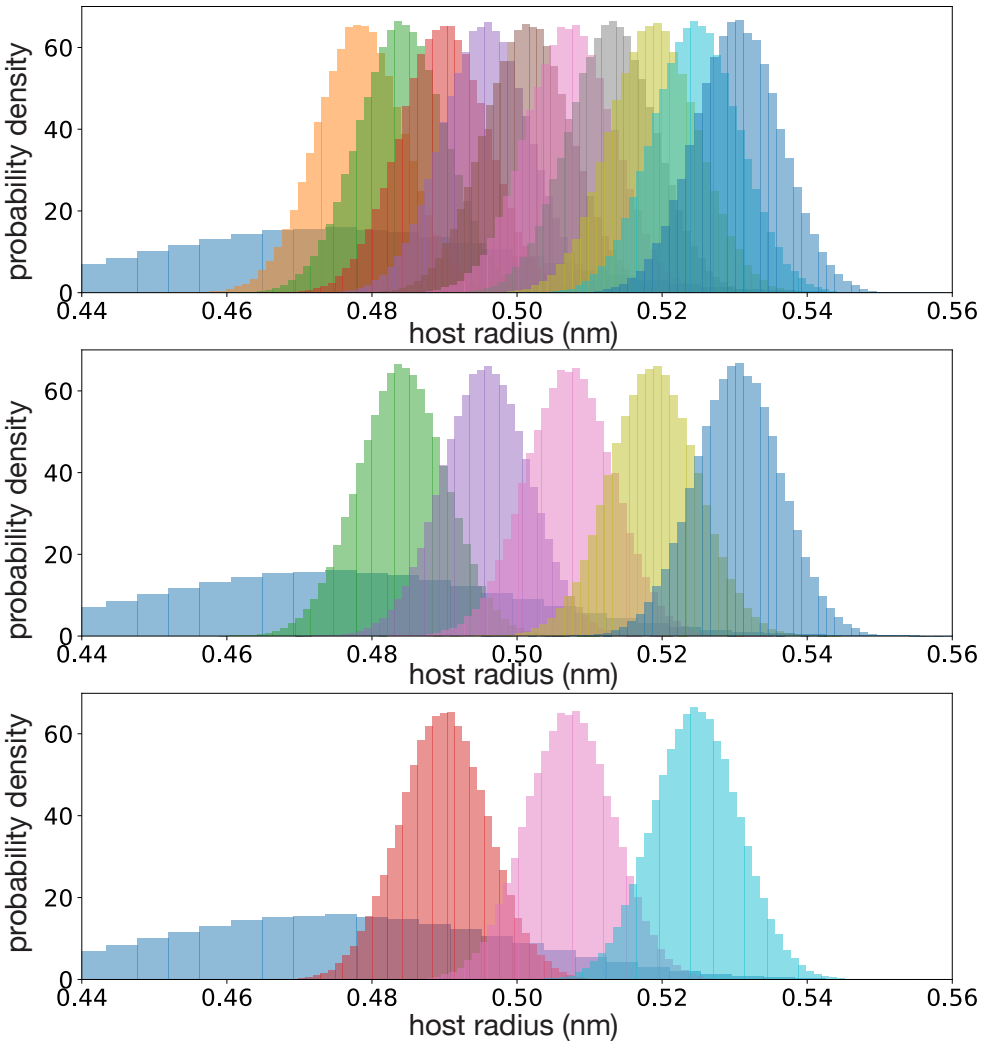

Figure S16: Distributions of the host's radius for states in the attachment phase of the PMF method with 97 windows (top), 49 windows (middle), and 33 windows (bottom) for the guest GII. The most left distribution is for state A and other distributions are for intermediate states of the attachment phase.

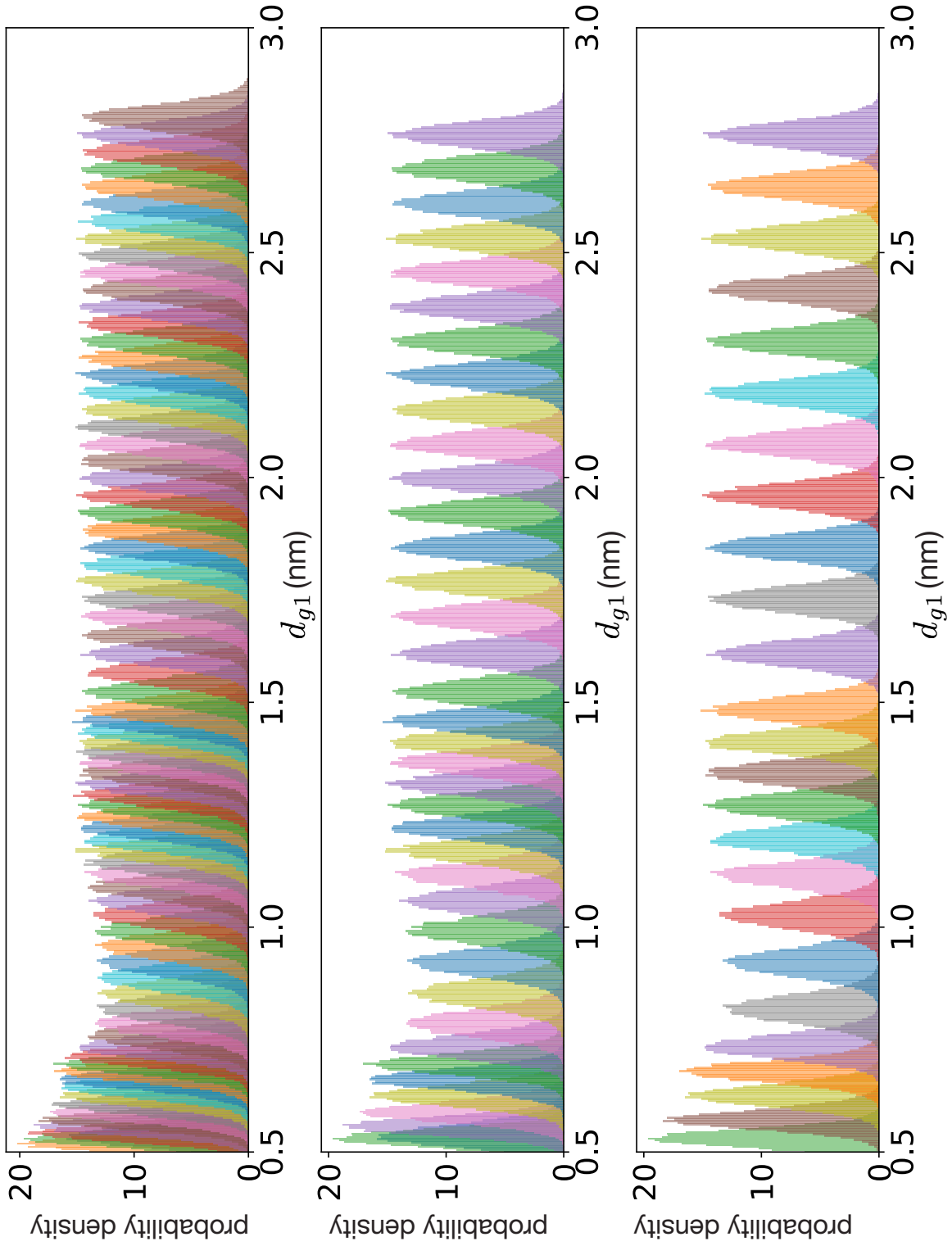

Figure S17: Distributions of the distance d_{g_1} between the guest GII's G1 site and the anchor particle P1 for states in the pulling phase of the PMF method with 97 windows (left), 49 windows (middle), and 33 windows (right).

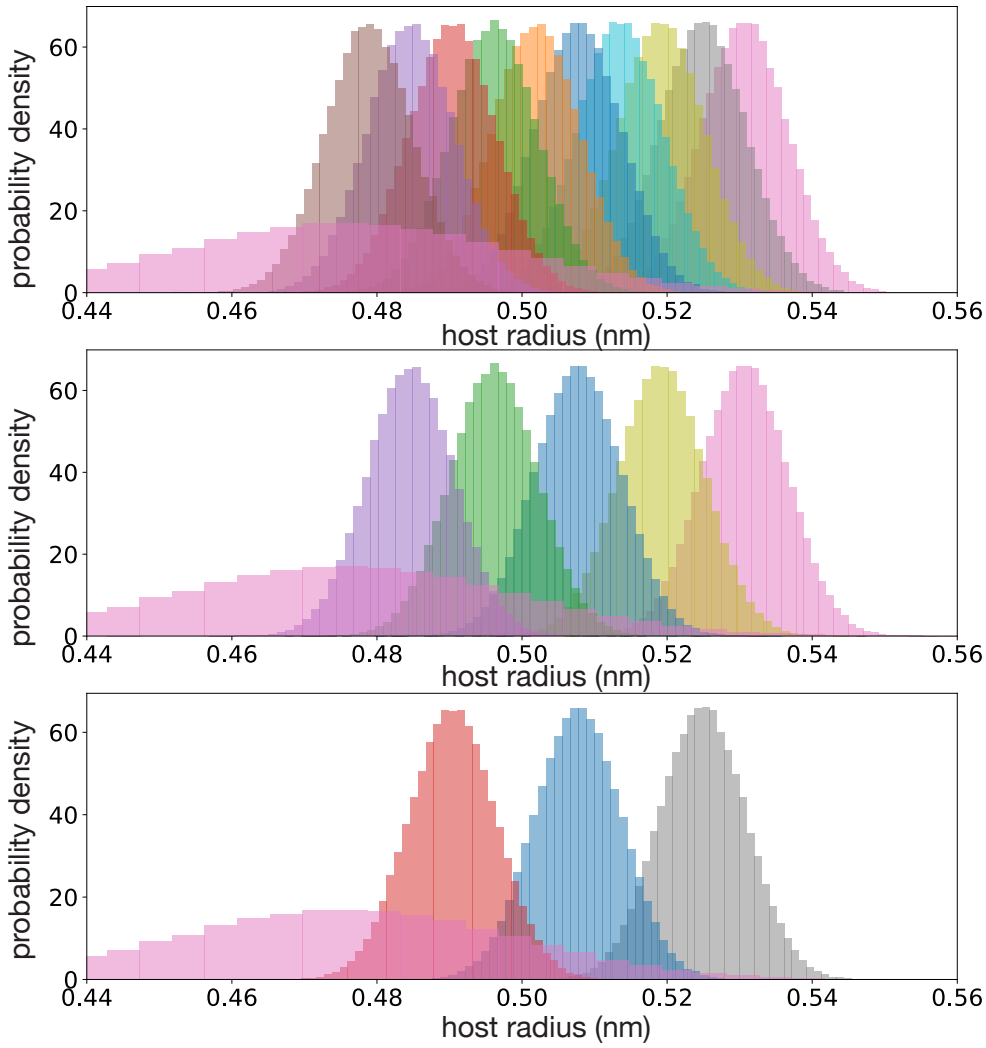

Figure S18: Distributions of the host's radius for states in the release phase of the PMF method with 97 windows (top), 49 windows (middle), and 33 windows (bottom) for the guest GII. The most left distribution is for state B and other distributions are for intermediate states of the release phase.

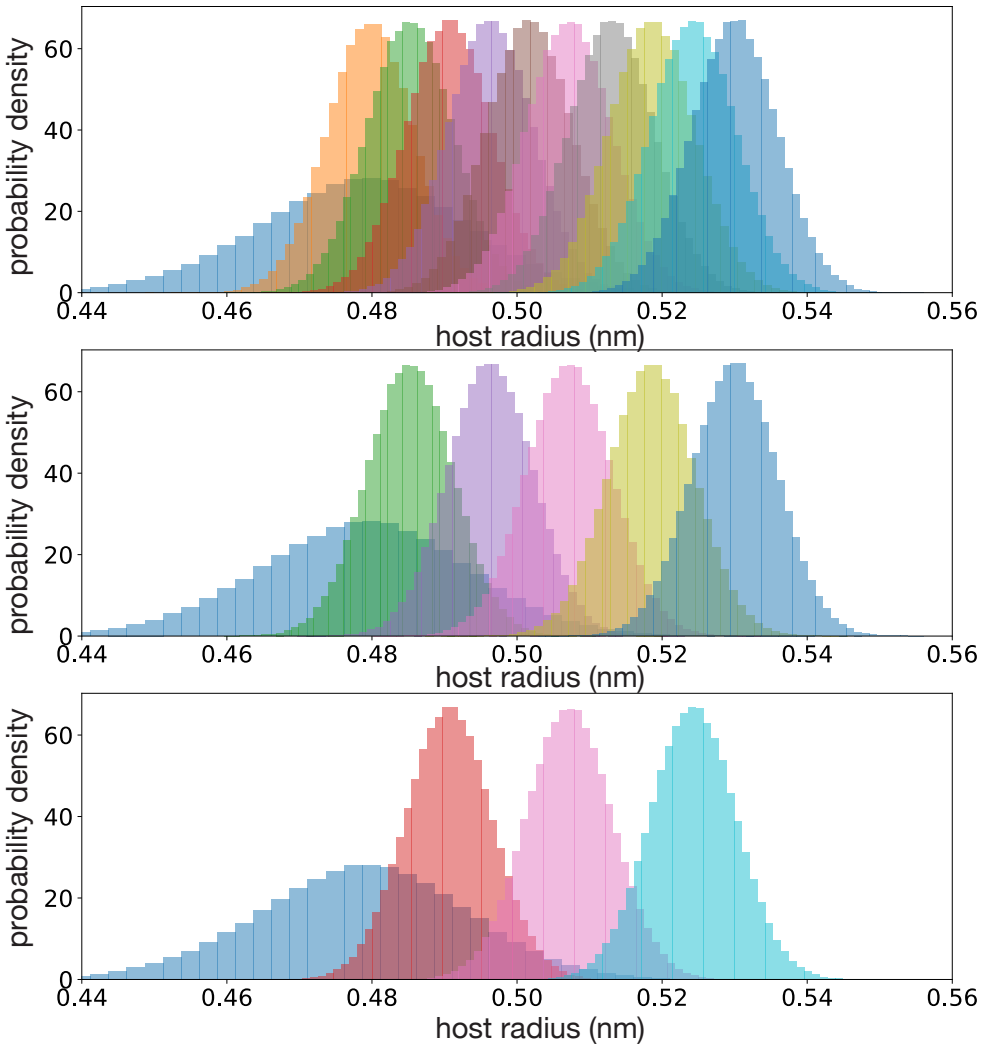

Figure S19: Distributions of the host's radius for states in the attachment phase of the PMF method with 97 windows (top), 49 windows (middle), and 33 windows (bottom) for the guest GIII. The most left distribution is for state A and other distributions are for intermediate states of the attachment phase.

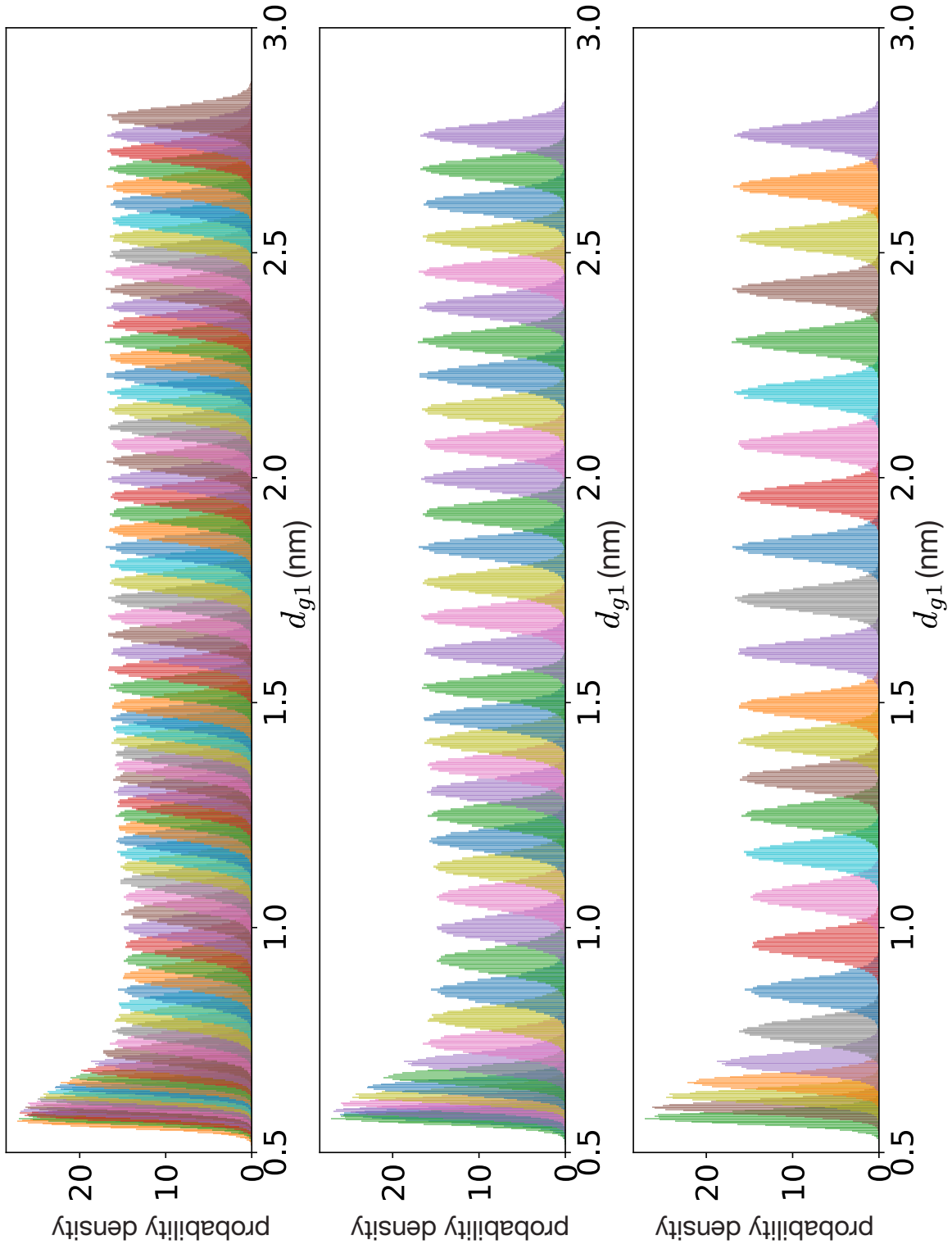

Figure S20: Distributions of the distance d_{g_1} between the guest GIII's G1 site and the anchor particle P1 for states in the pulling phase of the PMF method with 97 windows (left), 49 windows (middle), and 33 windows (right).

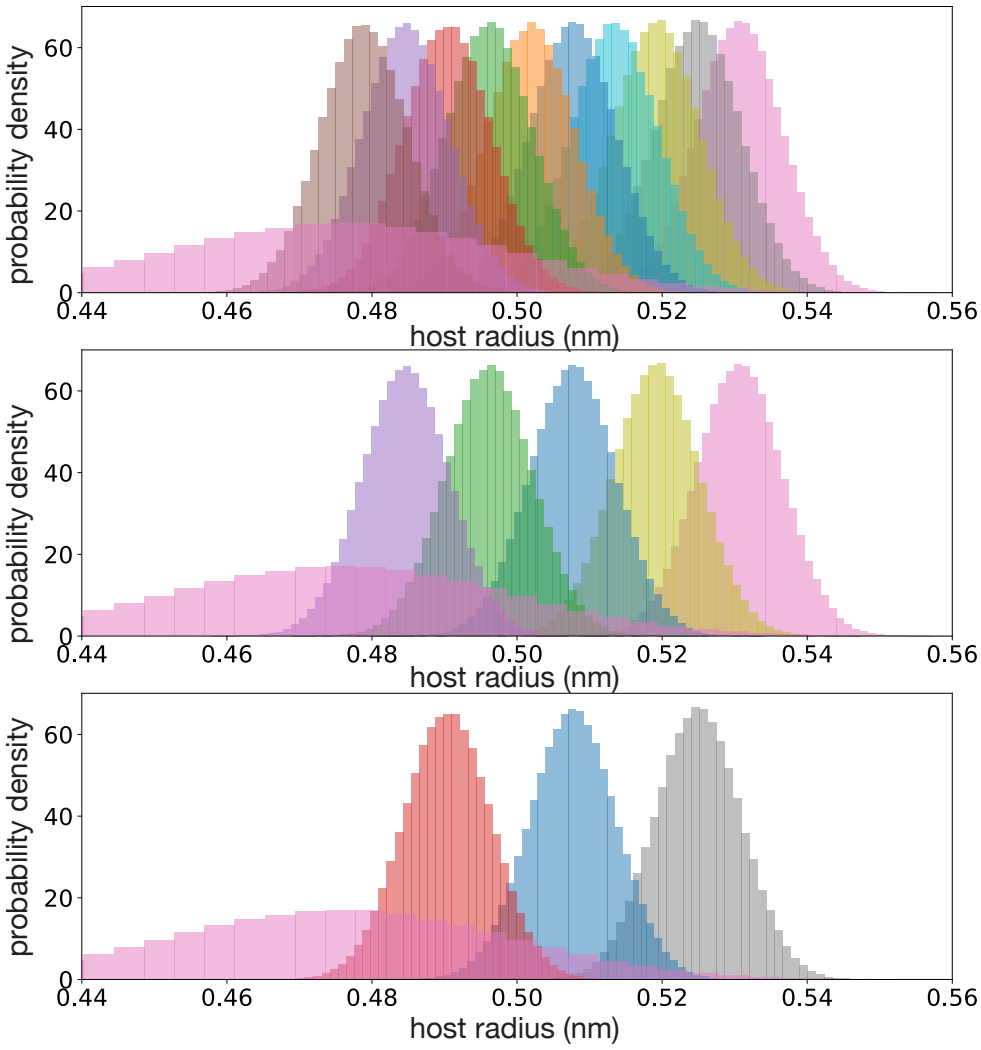

Figure S21: Distributions of the host's radius for states in the release phase of the PMF method with 97 windows (top), 49 windows (middle), and 33 windows (bottom) for the guest GIII. The most left distribution is for state B and other distributions are for intermediate states of the release phase.

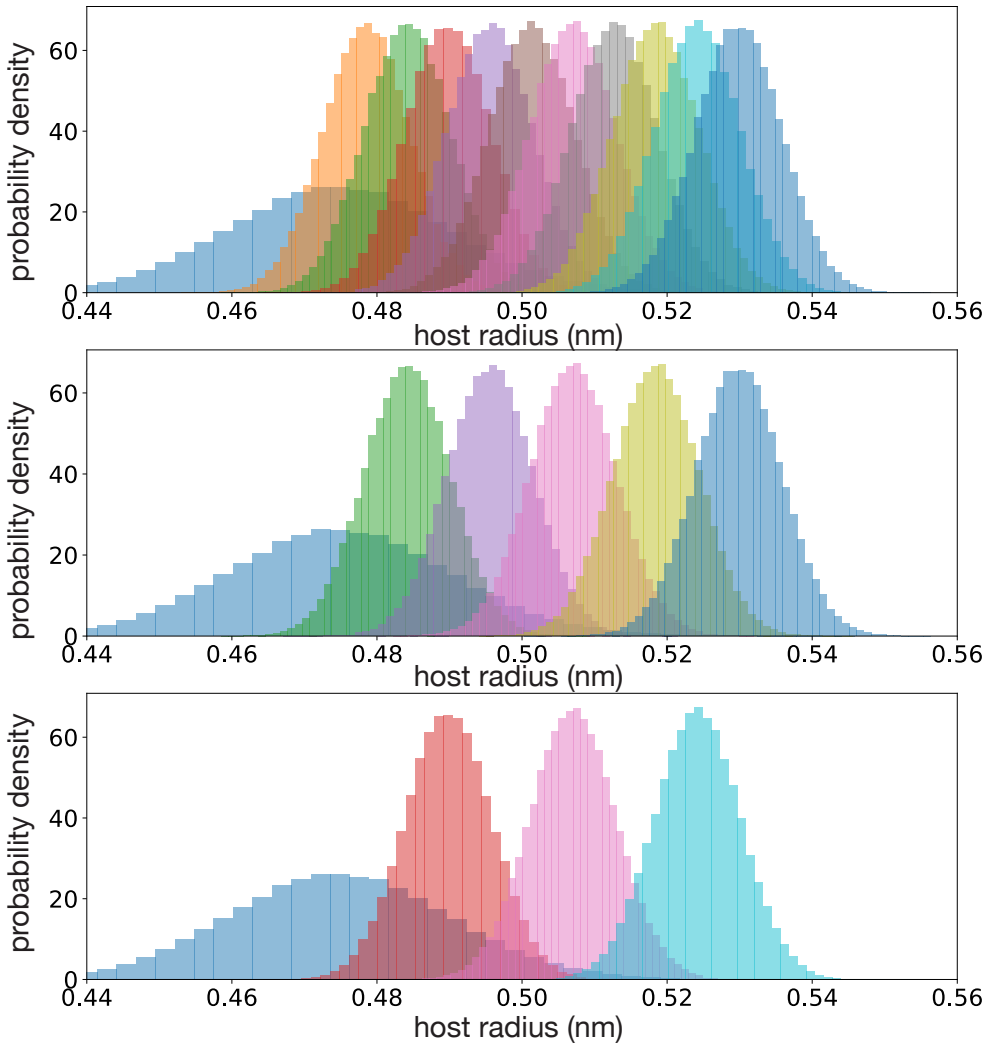

Figure S22: Distributions of the host's radius for states in the attachment phase of the PMF method with 97 windows (top), 49 windows (middle), and 33 windows (bottom) for the guest GIV. The most left distribution is for state A and other distributions are for intermediate states of the attachment phase.

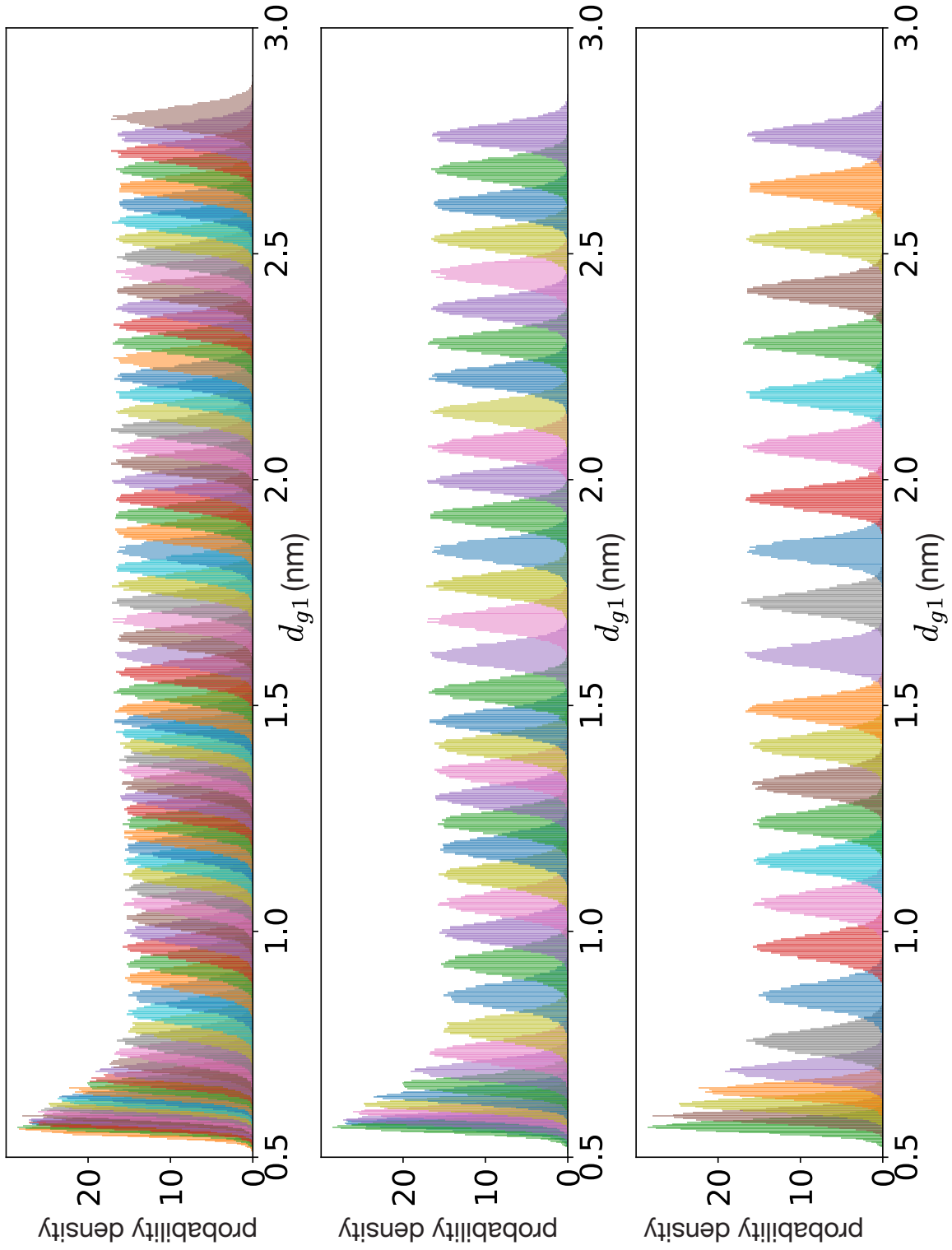

Figure S23: Distributions of the distance d_{g_1} between the guest GIV's G1 site and the anchor particle P1 for states in the pulling phase of the PMF method with 97 windows (left), 49 windows (middle), and 33 windows (right).

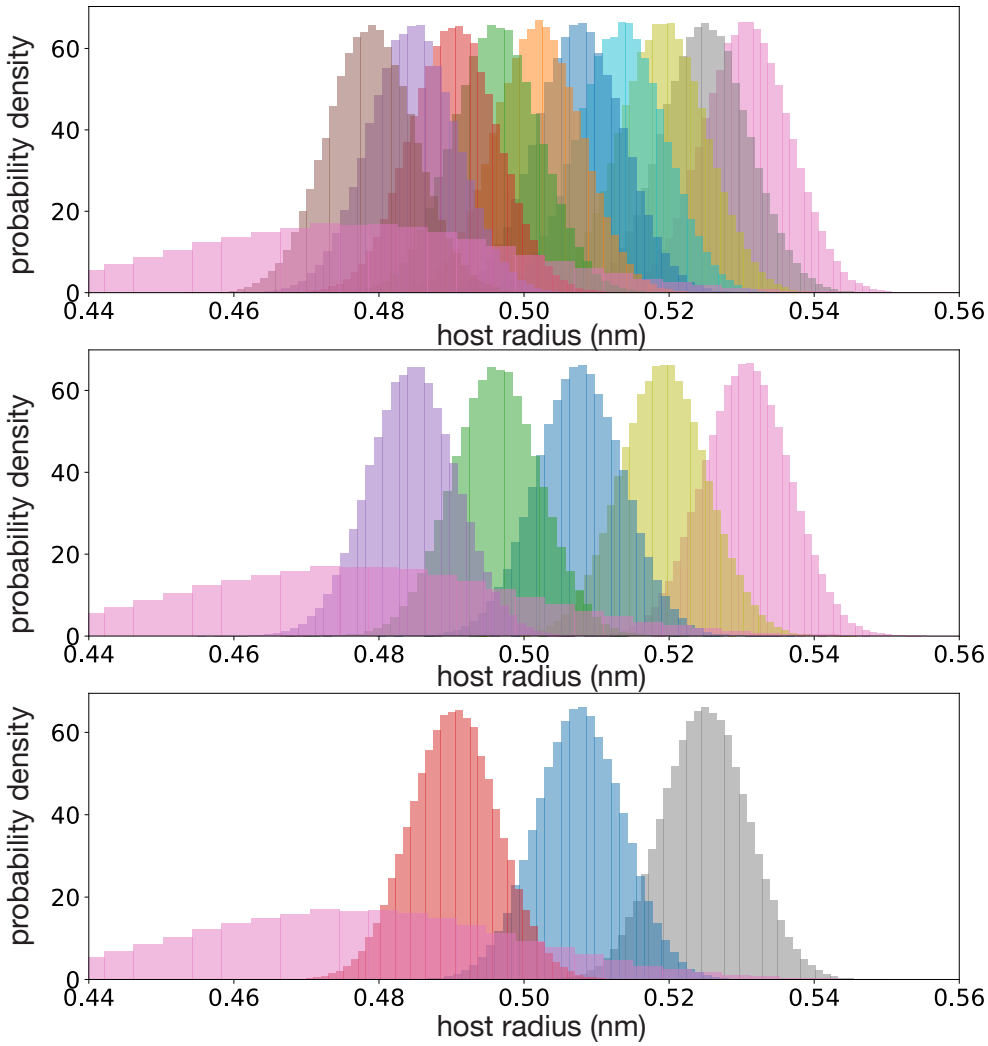

Figure S24: Distributions of the host's radius for states in the release phase of the PMF method with 97 windows (top), 49 windows (middle), and 33 windows (bottom) for the guest GIV. The most left distribution is for state B and other distributions are for intermediate states of the release phase.

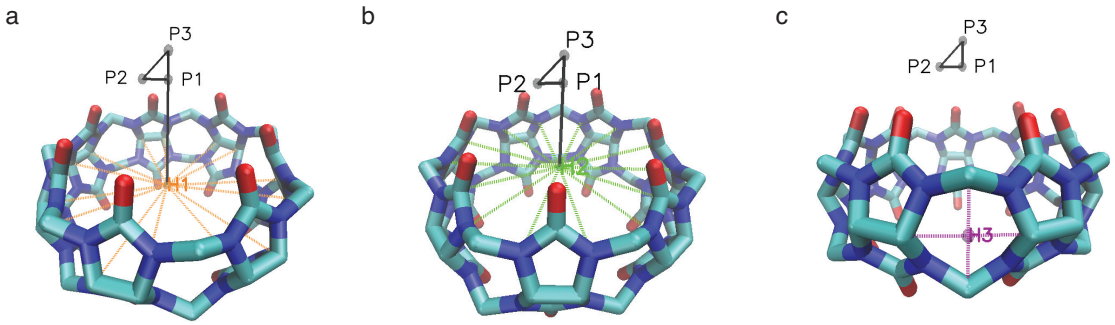

Figure S25: Illustration of the fixed anchor particles (P1, P2, and P3) and virtual sites (H1, H2, and H3) introduced for restraining the host's position and orientation. The virtual sites (H1, H2, and H3) were defined based on positions of host atoms. (a), (b), and (c) show the definition of the three virtual sites and the position of the three anchor particles. Hydrogen atoms of the host are not shown for clarity.

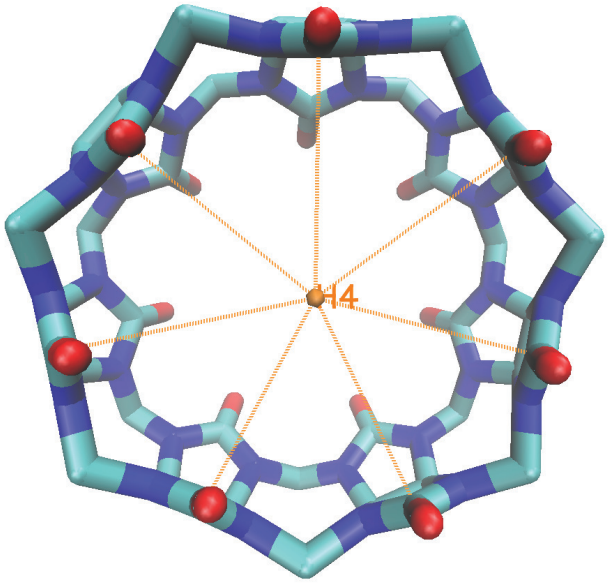

Figure S26: The virtual site H4 of the host was defined as the average position of the seven carbon atoms that are bonded to oxygen atoms around the cavity portal and on the opposite side of the host from the anchor particles. Harmonic biasing potentials were applied to the seven distances (shown as orange dashed line) between H4 and the carbon atoms.

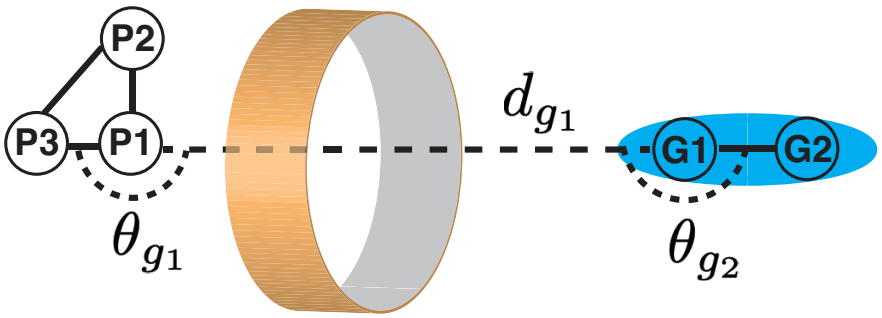

Figure S27: Illustration of the collective variables introduced to restrain the position and orientation of guest molecules. G1 and G2 are two virtual sites defined on the guest molecules. d_{g1} is the distance between G1 and P1, θ_{g1} is the angle between G1, P1, and P3, and θ_{g2} is the angle between G2, G1, and P1.

Table S1: Binding free energy [a] of the four guests calculated using the MM/GBSA method.

guest	MM/GBSA		
	ΔU	$T\Delta S$	$\Delta F_{\text{binding}}$
GI	-20.17±0.19	-16.20±0.01	-3.97±0.19
GII	-24.11±0.13	-17.58±0.02	-6.53±0.15
GIII	-40.25±0.22	-19.64±0.30	-20.61±0.16
GIV	-43.42±0.13	-20.77±0.01	-22.65±0.14

[a] Standard deviations are computed using three independent repeats;

References

(S1) 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.

(S2) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *The Journal of Physical Chemistry B* **2003**, *107*, 9535–9551.

(S3) Dinh, L.; Sohl-Dickstein, J.; Bengio, S. *Density Estimation Using Real NVP*. 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings. 2017.

(S4) Durkan, C.; Bekasov, A.; Murray, I.; Papamakarios, G. In *Advances in Neural Information Processing Systems 32*; Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc., 2019; pp 7511–7522.

(S5) Rezende, D. J.; Papamakarios, G.; Racanière, S.; Albergo, M. S.; Kanwar, G.; Shanahan, P. E.; Cranmer, K. Normalizing flows on tori and spheres. *arXiv preprint arXiv:2002.02428* **2020**,

(S6) He, K.; Zhang, X.; Ren, S.; Sun, J. *Deep residual learning for image recognition*. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2016.

(S7) Gregory, J. A.; Delbourgo, R. Piecewise Rational Quadratic Interpolation to Monotonic Data. *IMA Journal of Numerical Analysis* **1982**, *2*, 123–130.