Dynamic Docking of a Medium-Sized Molecule to Its Receptor by Multicanonical MD Simulations

Gert-Jan Bekker,[†] Mitsugu Araki,[‡] Kanji Oshima,[§] Yasushi Okuno,[‡] and Narutoshi Kamiya,[*,∥]

[†]Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan
[‡]Graduate School of Medicine, Kyoto University, 53 Shogoin-Kawaharacho, Sakyo-ku, Kyoto 606-8507, Japan
[§]Biotechnology Research Laboratories, KANEKA CORPORATION, 1-8, Miyamae-cho, Takasago-cho, Takasago, Hyogo 676-8688, Japan
[∥]Graduate School of Simulation Studies, University of Hyogo, 7-1-28 Minatojima Minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

Supporting information

Figures S1-S8
Tables S1-S3
Movie S1 (a higher quality version is available at https://www.youtube.com/watch?v=_S4k6FotfUE)

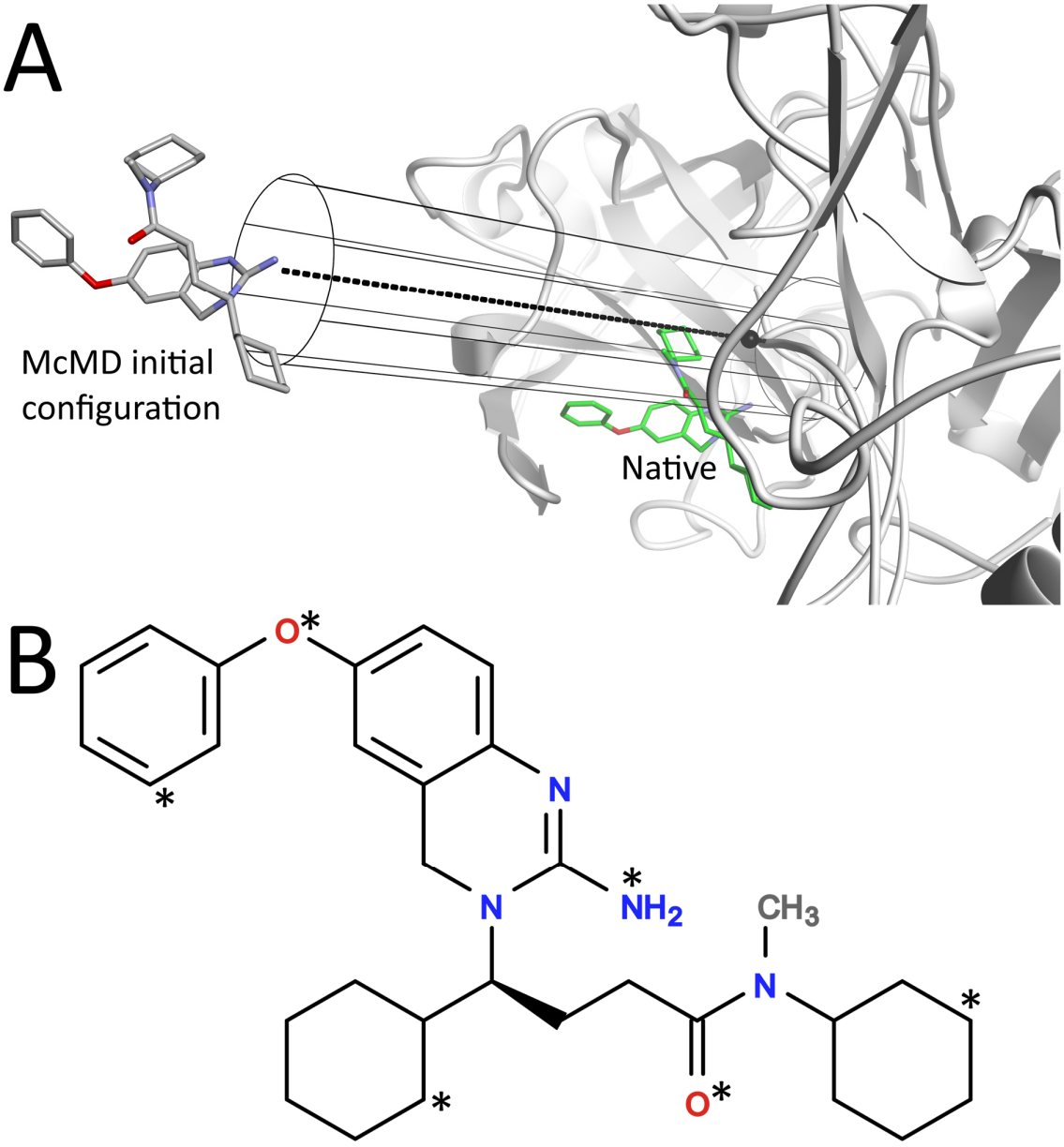

Figure S1. (A) Space restraints and the initial configuration for the McMD simulations. The center of mass (COM) of the ligand 3MR is restrained to stay inside the cylinder, which is defined based on the axis $\vec{\lambda}$ ranging from -5 Å to +29 Å as a black dotted line. The zero point is depicted as a black sphere. The ligand in the X-ray structure is colored in CPK with the carbon atoms colored green and is denoted by "Native" and hereon after referred to as the native configuration or structure, while BACE is rendered as cartoon in white. Also shown is the position of the ligand at the beginning of the McMD simulation denoted by "McMD initial configuration". (B) 2D structure of the inhibitor 3MR. The atoms of 3MR used during our PCA are denoted by "*".

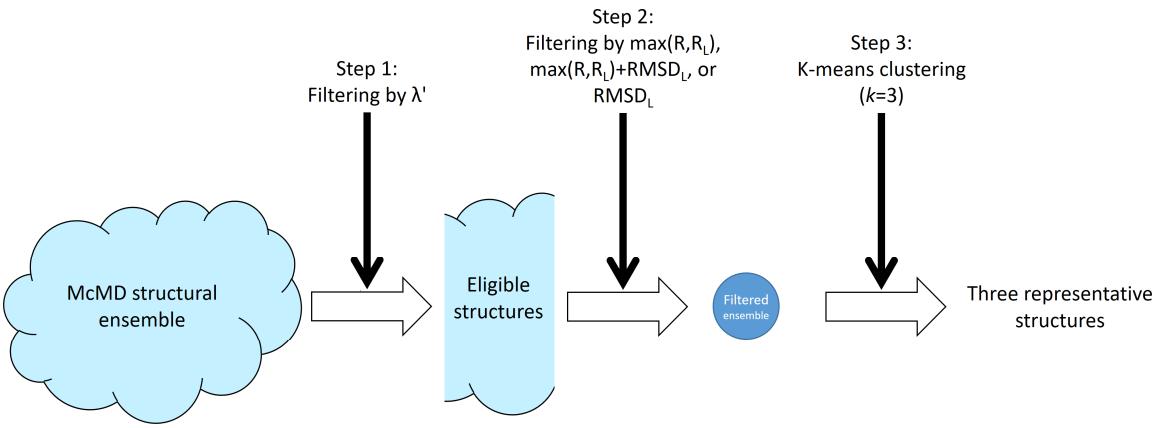

Figure S2. Schematic representation of the picking method applied for each window. For the picking process, each window is defined by the parameters λ'_l, λ'_c, λ'_u along $\vec{\lambda}$ as described in Table 1, corresponding to the lower range, center and upper range of each window, respectively. For each window, we obtain a set of eligible structures by filtering the entire McMD ensemble based on the λ' range from λ'_l to λ'_u (step 1). Next, we filter them based on the structural similarity via a nearest neighbor like approach with respect to a set of reference structures (see the main text for a detailed explanation), obtaining a filtered ensemble (step 2). Using this subset, we performed K-means clustering (with k=3) where we take a representative structure from each cluster (step 3) by taking the nearest-to-average structure, which then serve as the reference structures for the next window. The reference structure for window 2 corresponds to $\vec{s}$. Performing this procedure for each window defined in Table 1 starting from the bound state (window 2), we obtain a set of connected representative structures along the binding/unbinding direction $\vec{\lambda}$.

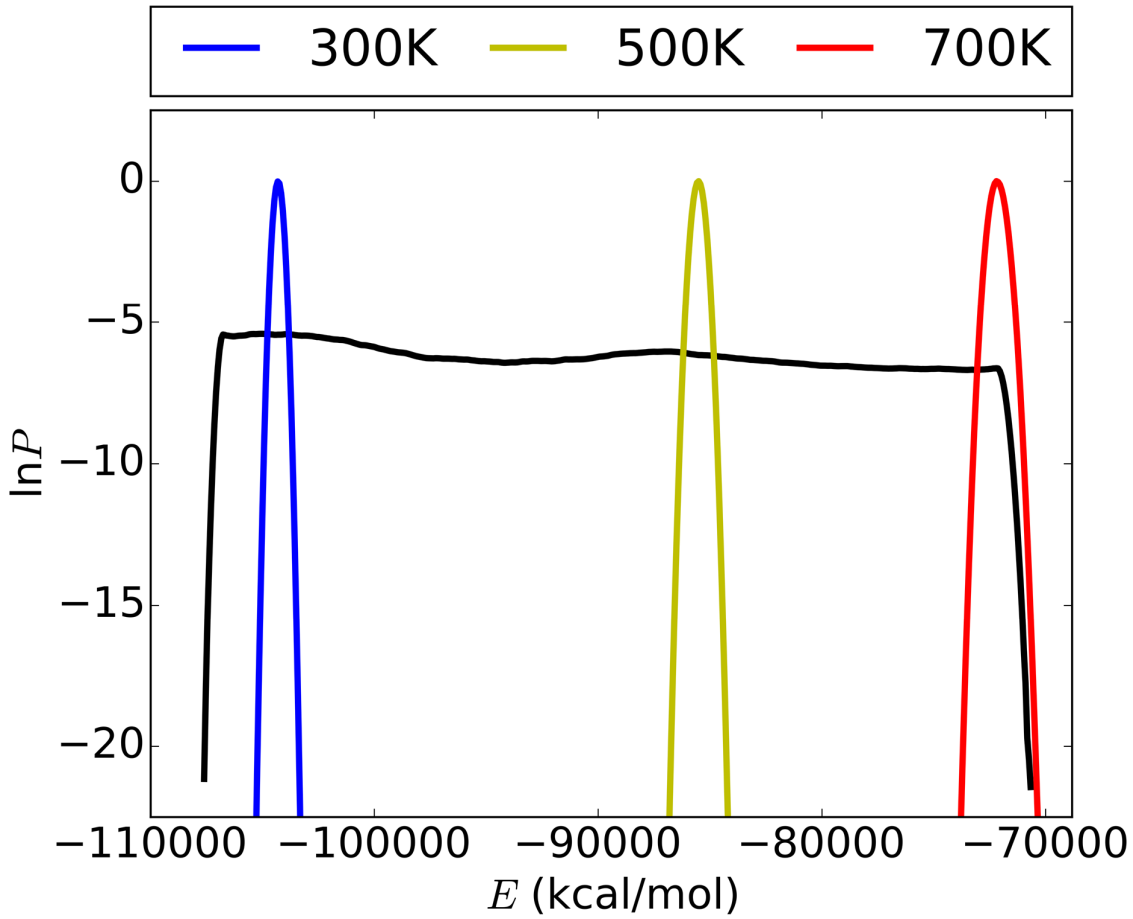

Figure S3. Potential energy distribution of the McMD ensemble in black as the log probability ln$P$ for each energetic state. Also shown are the canonical distributions $P_c(E, T)$ at 300 K, 500 K and 700 K in blue, yellow and red, respectively.

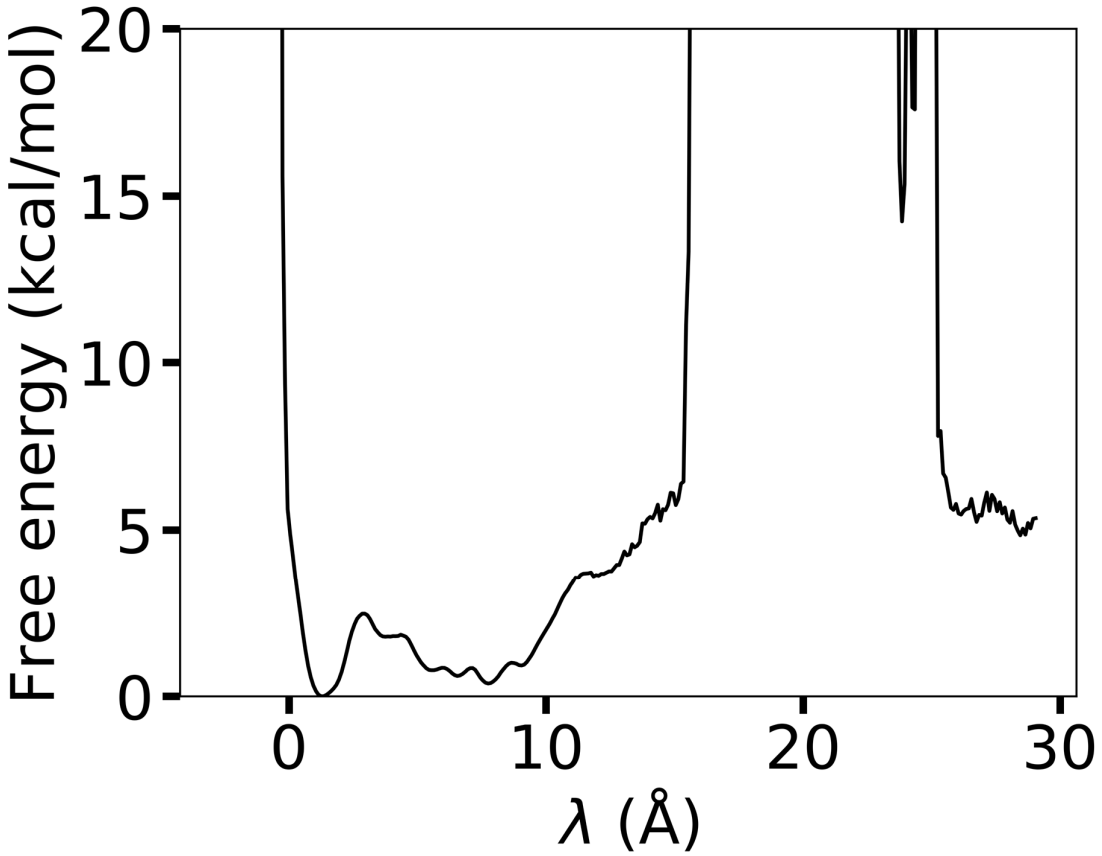

Figure S4. Free energy landscape at 300 K along λ obtained from the McMD simulation. The native structure corresponds to λ = 0.10 Å, while the minimum corresponds to 1.35 Å.

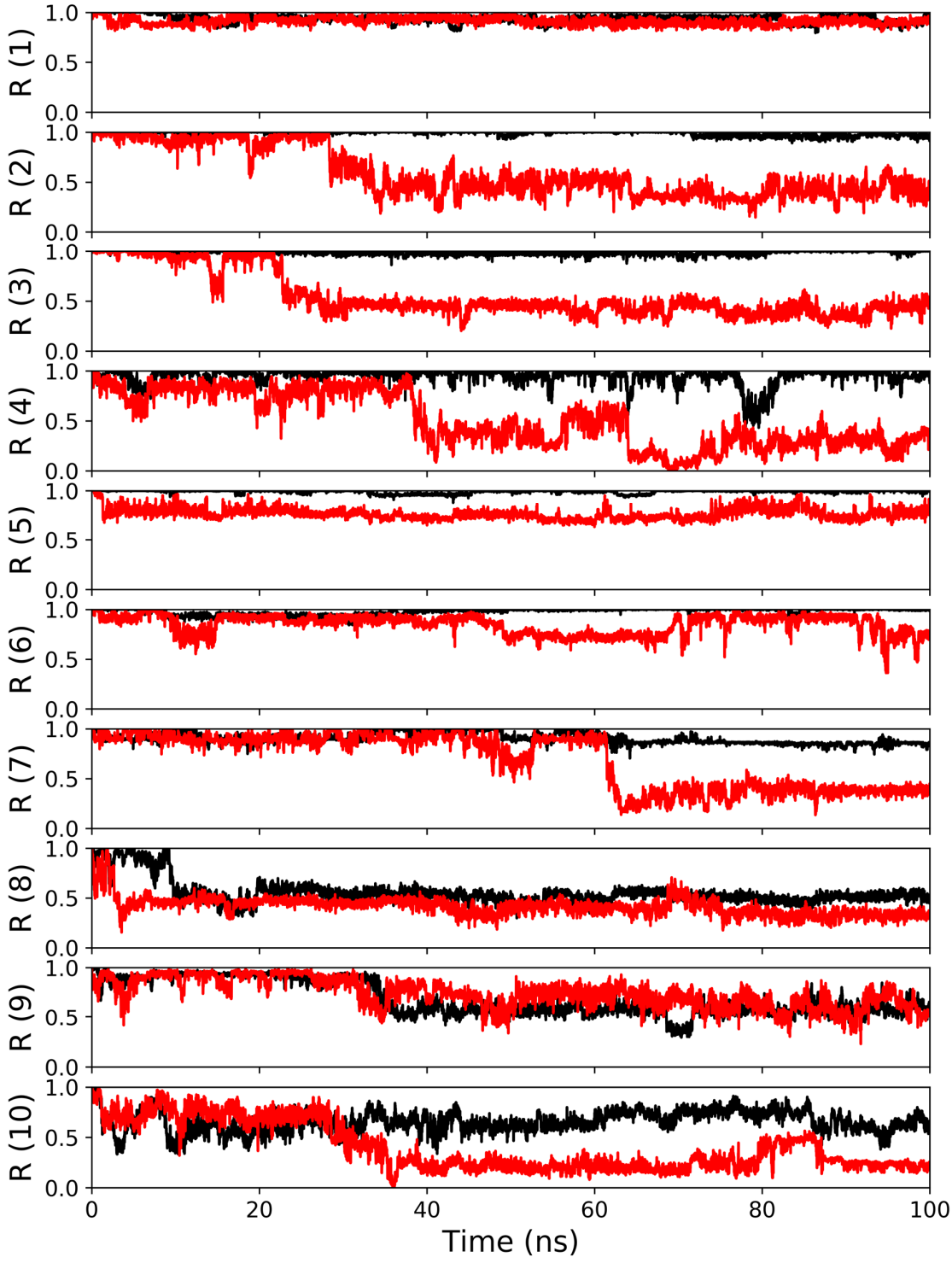

Figure S5. Stability of the $mcmd_k^r$ structures at 300 K (black) and at 400 K (red) along the trajectories. For each representative structure $mcmd_k^r$ shown in Figure 1C, 100 ns canonical MD simulations at 300 K and 400 K were performed and the R-value, i.e. the fraction of reference contacts with respect to the starting structure of each MD simulation, was plotted along each trajectory. Higher R-values correspond to more stable structures.

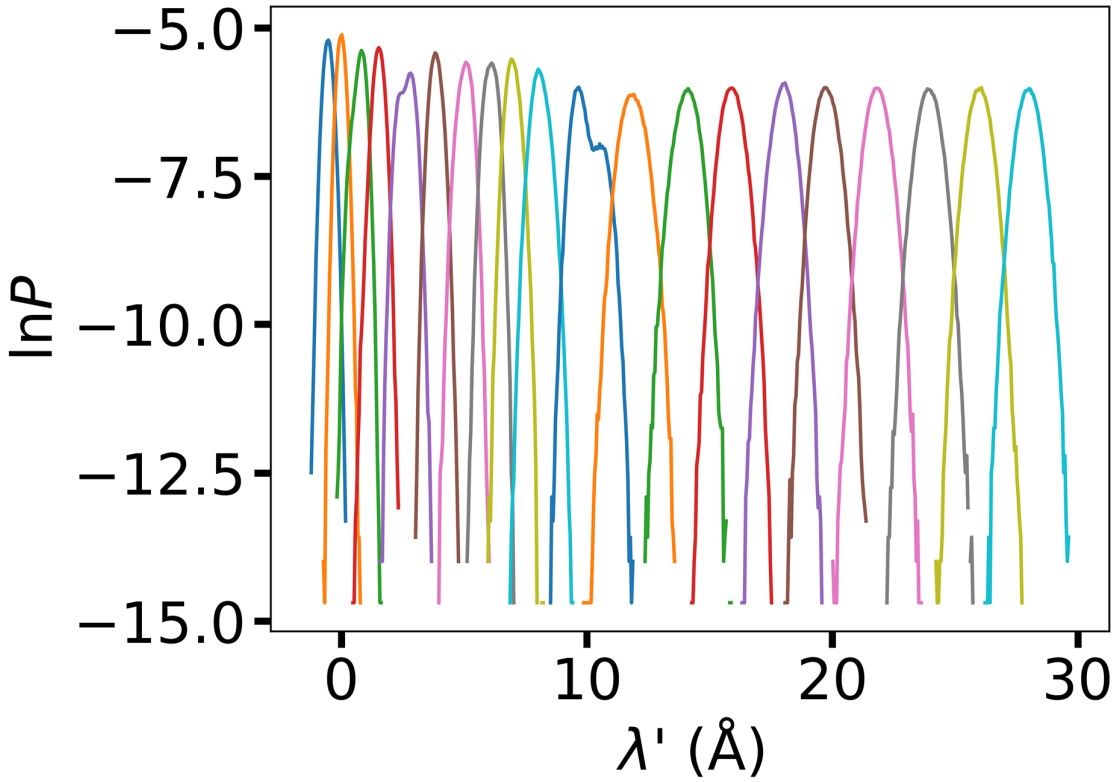

Figure S6. Probability distributions along λ' obtained during the US simulations.

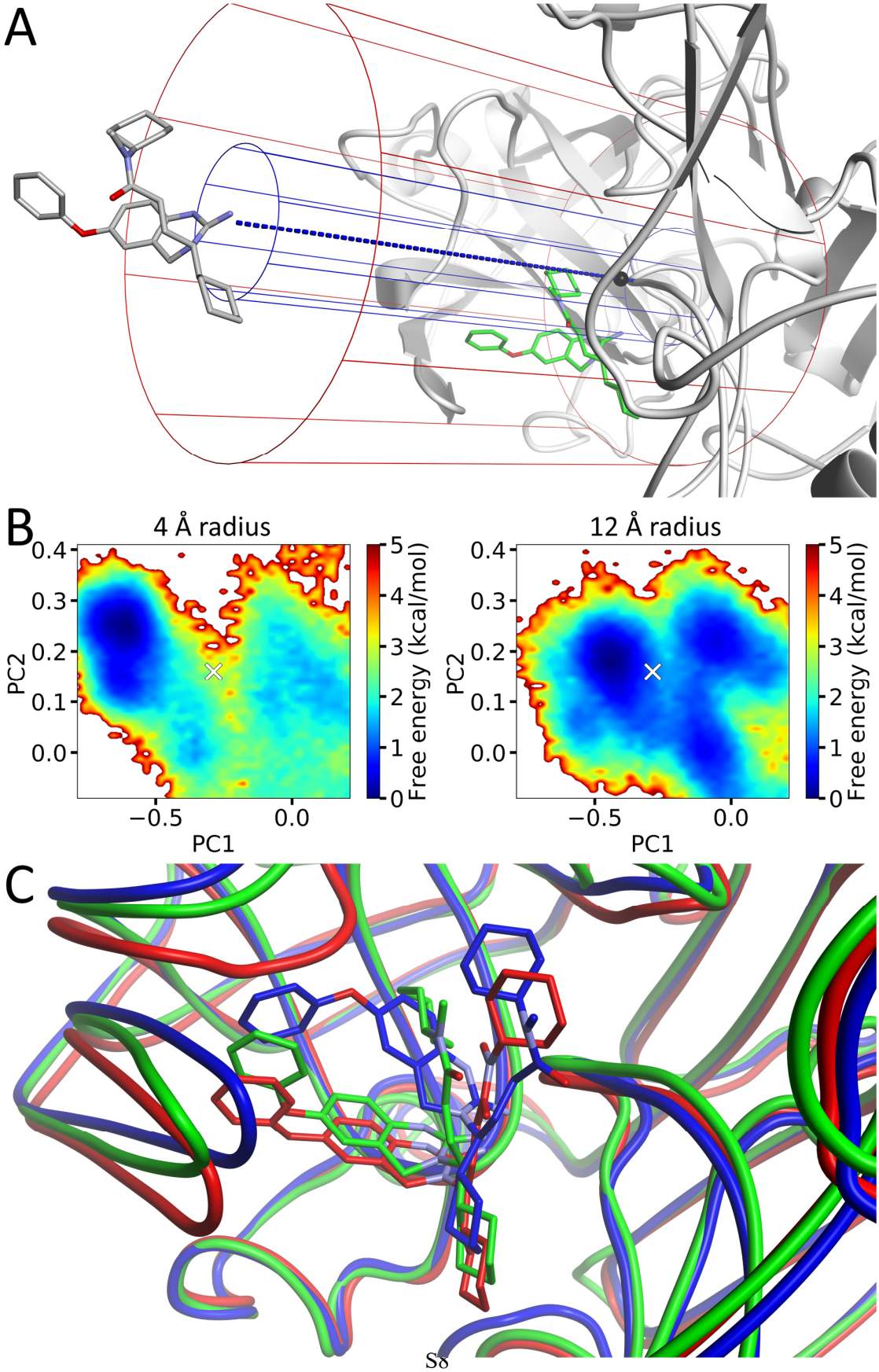
A
B
4 Å radius
12 Å radius
PC2
PC1
Free energy (kcal/mol)
0.4
0.3
0.2
0.1
0.0
−0.5
0.0
5
4
3
2
1
0
C

Figure S7. Analysis of the effects of the cylinder radius during the McMD docking simulations. (A) Comparison of the initial 4 Å radius cylinder, which is the same as shown in Figure S1A, and the second 12 Å radius cylinder with respect to the native structure. (B) Comparison of the FEL of the initial and second McMD simulation around the native structure. (C) Comparison of the native X-ray structure (green), the representative structure from our first McMD simulation (blue) at the global minimum (k=1) and the representative structure from our second McMD simulation (red) at the global minimum (k=1).

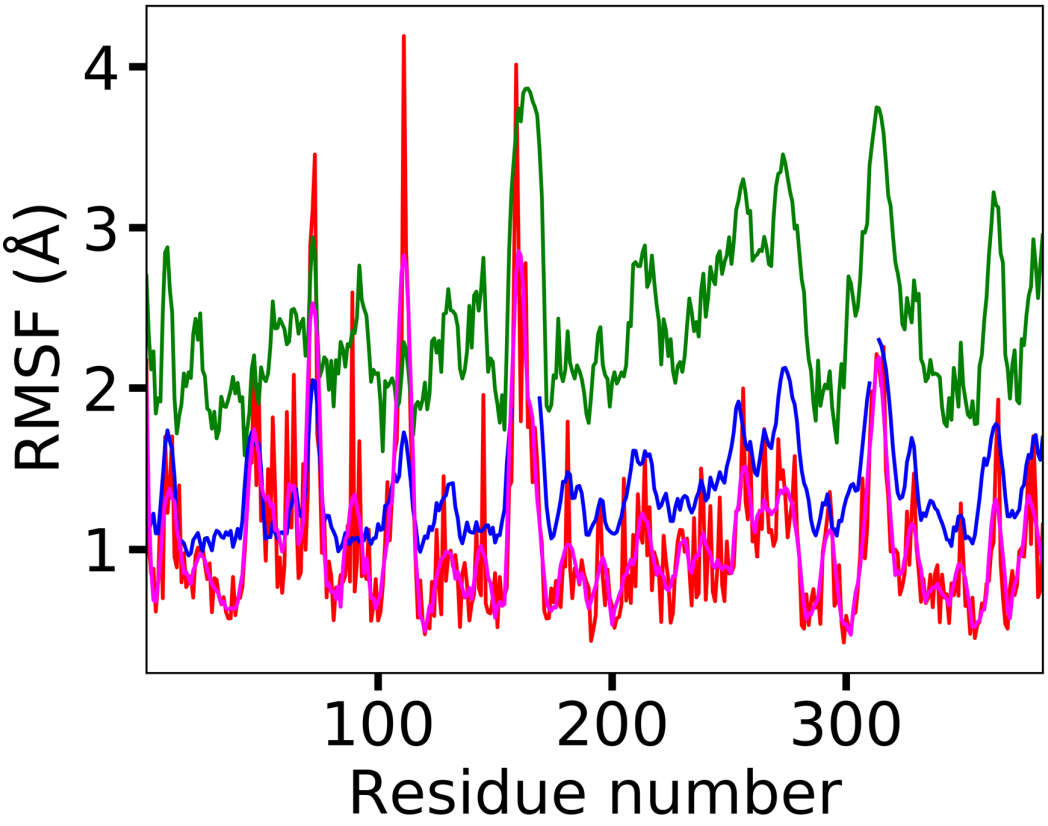

Figure S8. Root mean square fluctuation (RMSF) profiles of BACE's backbone α-carbons. Shown are the reweighted McMD ensemble at 300 K, the holo structure (PDB ID 2q15), the apo structure (PDB ID 3tpj) and a smoothed version of the reweighted McMD ensemble at 300 K in red, green, blue and magenta, respectively. The B-factor (B) of the apo and the holo structures were converted to the RMSF via $\sqrt{3B/8\pi}$. The smoothing was performed using a Savitzky-Golay filter with a window-size of 11 residues and a third order polynomial.

Table S1. Comparison of the R-values and RMSDs of the reference (before the arrow) and picked structures (after the arrow) of each window during the picking process. With three reference and picked structures each, there are nine combinations. For structures with no protein-ligand contacts, "X" is used instead of the R-value. For the initial window (i.e. window 2), there is only one reference structure, i.e. $\vec{s}$ and the comparison of the three listed structure is made with respect to that one. These structures serve as the starting structures of the US simulations.

Window	Previous window's structure -> current window's structure R-value (RMSD in Å)								
	1->1	1->2	1->3	2->1	2->2	2->3	3->1	3->2	3->3
1	0.979 (1.567)	0.775 (1.996)	0.821 (2.353)	0.970 (1.015)	0.741 (1.629)	0.761 (2.189)	0.968 (2.461)	0.702 (3.026)	0.770 (2.932)
2	0.989 (2.222)	0.964 (2.329)	0.967 (2.736)						
3	0.964 (1.421)	0.946 (2.545)	0.995 (1.658)	0.983 (1.727)	0.887 (2.812)	0.946 (2.007)	0.948 (2.867)	0.927 (3.521)	0.957 (2.927)
4	0.904 (2.233)	0.864 (2.682)	0.865 (2.083)	1.000 (0.764)	0.972 (1.343)	0.871 (2.212)	0.982 (1.663)	0.923 (2.330)	0.907 (1.636)
5	0.925 (1.980)	0.795 (3.431)	0.909 (2.009)	0.974 (1.504)	0.827 (3.251)	0.952 (1.296)	0.927 (1.903)	0.901 (2.470)	0.944 (1.878)
6	0.730 (2.733)	0.898 (2.956)	0.857 (1.991)	0.799 (3.377)	0.944 (2.379)	0.855 (2.574)	0.854 (2.849)	0.897 (3.346)	0.869 (2.296)
7	0.677 (2.835)	0.576 (3.678)	0.713 (3.703)	0.798 (3.896)	0.994 (2.003)	0.982 (2.374)	0.834 (2.930)	0.876 (2.808)	0.917 (2.925)
8	0.763 (3.229)	0.968 (2.642)	0.647 (4.743)	0.975 (2.046)	0.705 (4.992)	0.571 (3.102)	0.915 (2.057)	0.900 (4.344)	0.580 (3.042)
9	0.563 (4.175)	0.655 (4.381)	0.499 (5.175)	0.555 (5.478)	0.445 (5.739)	0.375 (6.029)	1.000 (1.467)	0.978 (1.791)	0.961 (2.349)
10	0.947 (1.497)	1.000 (1.733)	0.886 (3.479)	0.988 (1.020)	0.998 (1.708)	0.876 (3.558)	0.992 (1.278)	0.961 (2.211)	0.993 (2.274)
11	0.992 (2.313)	0.718 (4.574)	0.955 (3.231)	0.841 (2.220)	0.570 (5.006)	0.825 (3.297)	0.907 (3.335)	0.774 (2.896)	0.894 (2.931)
12	0.215 (4.817)	0.491 (3.483)	0.399 (4.898)	0.085 (6.242)	0.027 (6.288)	0.448 (3.487)	0.253 (4.472)	0.201 (3.717)	0.494 (3.777)
13	0.063 (4.484)	0.484 (4.273)	0.000 (5.941)	0.051 (7.157)	0.138 (6.997)	0.080 (7.444)	0.452 (4.696)	0.472 (4.583)	0.159 (4.660)
14	0.410 (5.869)	0.154 (5.223)	0.492 (4.460)	0.000 (6.190)	0.123 (4.860)	0.192 (5.273)	0.236 (4.007)	0.437 (3.338)	0.304 (3.112)
15	0.109 (5.148)	0.079 (5.841)	0.021 (6.035)	0.043 (5.550)	0.187 (5.519)	0.060 (6.022)	0.068 (4.254)	0.581 (4.622)	0.311 (4.628)
16	0.541 (5.975)	0.397 (5.754)	0.000 (5.392)	0.000 (4.375)	0.002 (5.224)	0.789 (5.207)	0.000 (4.181)	0.150 (3.589)	0.000 (3.655)
17	0.500 (4.461)	0.000 (5.344)	0.000 (4.981)	X (7.690)	X (4.069)	X (8.065)	0.671 (6.770)	0.030 (3.504)	0.451 (7.181)
18	0.195 (5.682)	0.000 (4.664)	0.000 (4.451)	0.000 (8.292)	0.000 (8.629)	0.951 (8.349)	0.250 (4.113)	0.000 (2.843)	0.000 (3.664)
19	X (5.717)	X (6.087)	X (5.744)	X (4.900)	X (6.067)	X (4.421)	0.000 (3.968)	0.000 (3.694)	0.000 (3.030)
20	X (6.342)	X (5.929)	X (6.370)	X (7.050)	X (5.198)	X (6.812)	X (3.899)	X (5.151)	X (4.040)

Table. S2. Structural overlap between neighboring windows of the US simulations. For each snapshot of each window w_i, the R-value with respect to each snapshot in the previous window (i.e. w_{i-1}) and the next window (i.e. w_{i+1}) were calculated and averaged per window permutation, calculating the average all-vs-all R-values ($\langle R(w_i, w_{i-1})\rangle$ and $\langle R(w_i, w_{i+1})\rangle$, respectively). This gives the overal structural similarity between the ensembles of neighbouring windows along the binding/unbinding direction. Shown are the averages with the standard deviation for both sides of the window and the rounded average number of contacts <N> between BACE and 3MR over all structures in the trajectories for each window. For the windows where there is no preceeding (w_i=1) or subsequent (w_i=20) one, - is noted. Only the trajectory data over the final 40 ns was used. The average R-value is direction dependent, e.g. w_i=1 (reference) to w_i=2 has an R-value of 0.838, while w_i=2 (reference) to w_i=1 has an R-value of 0.822. Since the set of reference structures is different depending on the direction, the values are slightly different, but still in the same order. As the ligand dissociates, the rounded average number of contacts <N> decreases and with it the average R value, while it's standard deviation increases, suggesting an increase in the number of unique conformations. The final two windows have barely any contacts, indicating that the ligand is in the bulk region.

Window ID *i*	$\langle R(w_i, w_{i-1})\rangle$ (std)	$\langle R(w_i, w_{i+1})\rangle$ (std)	<N>
1	-	0.838 (0.026)	176
2	0.822 (0.128)	0.931 (0.029)	182
3	0.916 (0.027)	0.898 (0.025)	175
4	0.942 (0.039)	0.745 (0.100)	163
5	0.795 (0.039)	0.638 (0.112)	131
6	0.672 (0.082)	0.526 (0.090)	121
7	0.584 (0.099)	0.507 (0.195)	102
8	0.422 (0.089)	0.467 (0.106)	105
9	0.461 (0.244)	0.456 (0.130)	96
10	0.390 (0.073)	0.367 (0.274)	128
11	0.409 (0.063)	0.112 (0.088)	94
12	0.114 (0.085)	0.107 (0.105)	63
13	0.085 (0.086)	0.131 (0.119)	72
14	0.190 (0.147)	0.066 (0.089)	54
15	0.073 (0.114)	0.247 (0.137)	44
16	0.247 (0.195)	0.330 (0.270)	46
17	0.487 (0.179)	0.077 (0.150)	21
18	0.122 (0.182)	0.010 (0.057)	11
19	0.079 (0.206)	0.009 (0.067)	1
20	0.030 (0.144)	-	0

Table S3. PMF and error estimates using different time ranges of the US trajectories. All WHAM calculations were executed using the same parameters as the calculations used in the main text (i.e. a $\Delta\lambda$ of 0.05 Å, a tolerance of 1×10^{-8} and with 1000 bootstraps). Here, start = 60 with end = 100 ns corresponds to the data used in the main text. ΔG is the average PMF over the final 50 bins (2.5 Å), with σ its standard deviation. ε corresponds to the average error (via bootstrapping) taken over the same range.

Start (ns)	End (ns)	ΔG	σ	ε
0	10	19.73	0.03	0.12
0	20	18.26	0.03	0.08
0	30	18.86	0.02	0.07
0	40	18.47	0.02	0.06
0	50	18.18	0.02	0.05
0	60	17.93	0.01	0.05
0	70	17.58	0.01	0.04
0	80	17.34	0.01	0.04
0	90	17.28	0.01	0.04
0	100	17.16	0.01	0.04
10	100	16.78	0.01	0.04
20	100	16.86	0.01	0.04
30	100	16.55	0.01	0.04
40	100	16.44	0.01	0.05
50	100	16.39	0.01	0.05
60	100	16.43	0.02	0.06
70	100	16.71	0.02	0.07
80	100	17.28	0.03	0.08
90	100	17.36	0.03	0.11

Movie S1. Representative binding/unbinding trajectory from the US ensemble along λ' starting from the bound state. The ligand 3MR is colored in black-CPK with the sidechains colored in grey-CPK. The sidechains of the residues other than Asp32, Asp228, Tyr71, Arg235 and Thr231 are displayed in thin sticks. The λ' value along the trajectory is shown in the bottom left corner. At the beginning and the end of the movie, every tenth residue is annotated, with the movie focussed on the binding site and the surrounding region.
A higher quality version is available at https://www.youtube.com/watch?v=_S4k6FotfUE.