Supporting Information

Miao and McCammon 10.1073/pnas.1800756115

SI Text

GaMD

GaMD enhances the conformational sampling of biomolecules by adding a harmonic boost potential to reduce the system energy barriers (1, 2). When the system potential $V(\vec{r})$ is lower than a reference energy E, the modified potential $V^*(\vec{r})$ of the system is calculated as:

$$V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r})$$

$$\Delta V(\vec{r}) = \begin{cases} \frac{1}{2}k\left(E - V(\vec{r})\right)^2, & V(\vec{r}) < E \\ 0, & V(\vec{r}) \geq E, \end{cases} \quad \text{[S1]}$$

where k is the harmonic force constant. The two adjustable parameters E and k are automatically determined based on three enhanced sampling principles (1). The reference energy needs to be set in the following range:

$$V_{\max} \leq E \leq V_{\min} + \frac{1}{k}, \quad \text{[S2]}$$

where $V_{\min}$ and $V_{\max}$ are the system minimum and maximum potential energies. To ensure that Eq. S2 is valid, k has to satisfy: $k \leq 1/V_{\max} - V_{\min}$. Let us define $k \equiv k_0 \bullet 1/V_{\max} - V_{\min}$, then $0 < k_0 \leq 1$. The SD of ΔV needs to be small enough (i.e., narrow distribution) to ensure accurate energetic reweighting (3): $\sigma_{\Delta V} = k(E - V_{\text{avg}})\sigma_V \leq \sigma_0$, where V_{avg} and σ_V are the average and SD of the system potential energies, and $\sigma_{\Delta V}$ is the SD of ΔV with σ_0 as a user-specified upper limit (e.g., $10k_BT$) for accurate reweighting. When E is set to the lower bound $E = V_{\max}$ according to Eq. S2, k_0 can be calculated as:

$$k_0 = \min(1.0, k_0') = \min\left(1.0, \frac{\sigma_0}{\sigma_V} \bullet \frac{V_{\max} - V_{\min}}{V_{\max} - V_{avg}}\right). \quad \text{[S3]}$$

Alternatively, when the threshold energy E is set to its upper bound $E = V_{\min} + 1/k$, k_0 is set to:

$$k_0 = k_0'' \equiv \left(1 - \frac{\sigma_0}{\sigma_V}\right) \bullet \frac{V_{\max} - V_{\min}}{V_{avg} - V_{\min}}, \quad \text{[S4]}$$

if k_0'' is found to be between *0* and *1*. Otherwise, k_0 is calculated by using Eq. **S3**.

For energetic reweighting of GaMD simulations, the probability distribution along a selected reaction coordinate $A(\vec{r})$ is written as $p^*(A)$, where $\vec{r}$ denotes the atomic positions $\{\boldsymbol{r}_1, \cdots, \boldsymbol{r}_N\}$. Given the boost potential $\Delta V(\vec{r})$ of each frame, $p^*(A)$ can be reweighted to recover the canonical ensemble distribution, $p(A)$, as:

$$p(A_j) = p^*(A_j) \frac{\left\langle e^{\beta \Delta V(\vec{r})} \right\rangle_j}{\sum_{i=1}^{M} \left\langle p^*(A_i) e^{\beta \Delta V(\vec{r})} \right\rangle_i}, \; j = 1, \cdots, M, \quad \text{[S5]}$$

where M is the number of bins, $\beta = k_BT$, and $\langle e^{\beta \Delta V(\vec{r})} \rangle_j$ is the ensemble-averaged Boltzmann factor of $\Delta V(\boldsymbol{r})$ for simulation frames found in the j^{th} bin. As shown earlier, when the boost potential follows near-Gaussian distribution, cumulant expansion to the second order provides the more accurate reweighting than the exponential average and Maclaurin series expansion methods (3). The ensemble-averaged reweighting factor is approximated by using cumulant expansion:

$$\left\langle e^{\beta \Delta V} \right\rangle = \exp\left\{ \sum_{k=1}^{\infty} \frac{\beta^k}{k!} C_k \right\}, \quad \text{[S6]}$$

where the first two cumulants are given by:

$$\begin{aligned} C_1 &= \langle \Delta V \rangle, \\ C_2 &= \left\langle \Delta V^2 \right\rangle - \langle \Delta V \rangle^2 = \sigma_{\Delta V}^2 \end{aligned}. \quad \text{[S7]}$$

When the boost potential follows near-Gaussian distribution, cumulant expansion to the second order (or "Gaussian approximation") provides the accurate approximation for free energy calculations (3). The reweighted free energy $F(A) = -k_BT\ln p(A)$ is calculated as:

$$F(A) = F^*(A) - \frac{1}{\beta} \sum_{k=1}^{2} \frac{\beta^k}{k!} C_k + F_c, \quad \text{[S8]}$$

where $F^*(A) = -k_BT\ln p^*(A)$ is the modified free energy obtained from GaMD simulation, and F_c is a constant.

System Setup

The X-ray structure of the agonist–nanobody-bound M_2 receptor (PDB ID code 4MQS; Fig. 1*A*) (4) was used for the simulations. The agonist IXO and G-protein mimetic nanobody Nb9-8 was initially displaced to be >20 Å away from the M_2 receptor for simulation. Then preparation of the simulation system followed a similar protocol as presented earlier (5). All chain termini were capped with neutral groups (acetyl and methylamide). Two disulfide bonds that were resolved in the crystal structure, i.e., $Cys96^{3.25}$–$Cys176^{ECL2}$ and $Cys413^{6.61}$–$Cys416^{7.29}$, were maintained in the simulations. By using the *psfgen* plugin in VMD (6), protein residues were set to the standard CHARMM protonation states at neutral pH, with the exception of $Asp69^{2.50}$ which is buried in the hydrophobic core and thus protonated (7).

The M_2 receptor was inserted into a palmitoyl-oleoyl-phosphatidyl-choline (POPC) bilayer with all overlapping lipid molecules removed by using the *Membrane* plugin in VMD (6). The system charges were then neutralized at 0.15 M NaCl by using the *Solvate* plugin in VMD (6). The agonist–receptor-nanobody system was solvated in a box of 96 × 100 × 133 $Å^3$ and contained 192 lipid molecules, 75 Na^+, 75 Cl^-, and 26,464 water molecules, for a total of ~111,600 atoms. Periodic boundary conditions were applied on the simulation system.

Simulation Protocol

The CHARMM36 parameter set (8) was used for the M_2 receptor, G-protein mimetic nanobody, and POPC lipids. For agonist IXO, the force field parameters were computed by using the General Automated Atomic Model Parameterization (GAAMP) tool (5, 9). With ab initio quantum mechanical calculations, GAAMP (9) generates force-field parameters that are compatible with CHARMM as used for protein and lipids.

Initial energy minimization, thermalization, and 100-ns cMD equilibration were performed by using NAMD2.10 (10). A cutoff distance of 12 Å was used for the van der Waals and short-range electrostatic interactions, and the long-range electrostatic interactions were computed with the particle-mesh Ewald summation method (11) using a grid point density of 1/Å. A 2-fs integration time step was used for all MD simulations, and a multiple-time-stepping algorithm was used, with bonded and short-range nonbonded interactions computed every time step and long-range electrostatic interactions every two time steps. The SHAKE algorithm was applied to all hydrogen-containing bonds. The NAMD simulation started with equilibration of the lipid tails. With all other atoms fixed, the lipid tails were energy-minimized for 1,000 steps by using the conjugate gradient algorithm and melted with an isothermal–isovolumetric (NVT) run for 0.5 ns at 310 K. The two systems were further equilibrated by using an isothermal–isobaric (NPT) run at 1 atm and 310 K for 10 ns with 5 kcal/(mol·Å^2) harmonic position restraints applied to the crystallographically identified atoms in the protein and ligand. The system volume was found to decrease with a flexible unit cell applied and level off within 10-ns NPT run, suggesting that solvent and lipid molecules in the system were well equilibrated. Final equilibration of each system was performed by using an NPT run at 1 atm and 310 K for 0.5 ns with all atoms unrestrained. After energy minimization and system equilibration, cMD simulation was performed on each system for 100 ns at 1 atm pressure and 310 K with a constant ratio constraint applied on the lipid bilayer in the *x*–*y* plane.

With the NAMD output structure, along with the system topology and CHARMM36 force field files, the *ParmEd* tool in the AMBER package was used to convert the simulation files into the AMBER format (12). The GaMD module implemented in the GPU version of AMBER14 (1, 12) was then applied to perform the GaMD simulation, which included 10-ns short cMD simulation used to collect the potential statistics for calculating GaMD acceleration parameters, 50-ns equilibration after adding the boost potential, and, finally, multiple independent GaMD production simulations with randomized initial atomic velocities. All GaMD simulations were run at the "dual-boost" level by setting the reference energy to the lower bound, i.e., $E = V_{max}$ (1). One boost potential is applied to the dihedral energetic term and another to the total potential energetic term. The average and SD of the system potential energies were calculated in every 400,000 (800 ps). The upper limit of the boost potential SD, σ_0, was set to 6.0 kcal/mol for both the dihedral and total potential energetic terms. Similar temperature and pressure parameters were used as in the NAMD simulations. A list of the GaMD production simulations lasting ∼4,500 ns is provided in Table 1.

Simulation Analysis

CPPTRAJ (13) and VMD (6) were used to analyze the GaMD simulation trajectories. Particularly, distances were calculated between the C_α atoms of Arg121$^{3.50}$–Thr386$^{6.34}$, and similarly for the perimeter of triangle formed by the C_α atoms of Tyr104$^{3.33}$–Tyr403$^{6.51}$–Tyr426$^{7.39}$. rmsds were calculated for the diffusing agonist IXO and G-protein mimetic nanobody Nb9-8 relative to the 4MQS X-ray conformation. The β_2, β_3, β_6, β_7, and β_8 strands that represent the core domain of Nb9-8 are selected for calculating rmsd of the nanobody and heavy atoms for the agonist IXO.

The *PyReweighting* (3) toolkit was used to reweight the atom distances, the perimeter of the tyrosine triangle, and the ligand and nanobody rmsds to compute the PMF profiles. A bin size of 1 Å was used for the atom distances, perimeter of the tyrosine triangle, and rmsds. The cutoff was set to 500 for 2D PMF calculations. The 2D PMF profiles were obtained for each complex system regarding the nanobody Nb9-8 rmsd vs. Arg121$^{3.50}$–Thr386$^{6.34}$ distance or the Tyr104$^{3.33}$–Tyr403$^{6.51}$–Tyr426$^{7.39}$ triangle perimeter.

1. Miao Y, Feher VA, McCammon JA (2015) Gaussian accelerated molecular dynamics: Unconstrained enhanced sampling and free energy calculation. *J Chem Theory Comput* 11:3584–3595.
2. Pang YT, Miao Y, Wang Y, McCammon JA (2017) Gaussian accelerated molecular dynamics in NAMD. *J Chem Theory Comput* 13:9–19.
3. Miao Y, et al. (2014) Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *J Chem Theory Comput* 10:2677–2689.
4. Kruse AC, et al. (2013) Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* 504:101–106.
5. Miao Y, McCammon JA (2016) Graded activation and free energy landscapes of a muscarinic G-protein-coupled receptor. *Proc Natl Acad Sci USA* 113:12162–12167.
6. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14:33–38, 27–28.
7. Kruse AC, et al. (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* 482:552–556.
8. Vanommeslaeghe K, MacKerell AD, Jr (2014) CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochim Biophys Acta*.
9. Huang L, Roux B (2013) Automated force field parameterization for nonpolarizable and polarizable atomic models based on Ab initio target data. *J Chem Theory Comput* 9:3543–3556.
10. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26: 1781–1802.
11. Essmann U, et al. (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–8593.
12. Case D, et al. (2014) Amber 14 (University of California, San Francisco).
13. Roe DR, Cheatham TE, 3rd (2013) PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* 9:3084–3095.

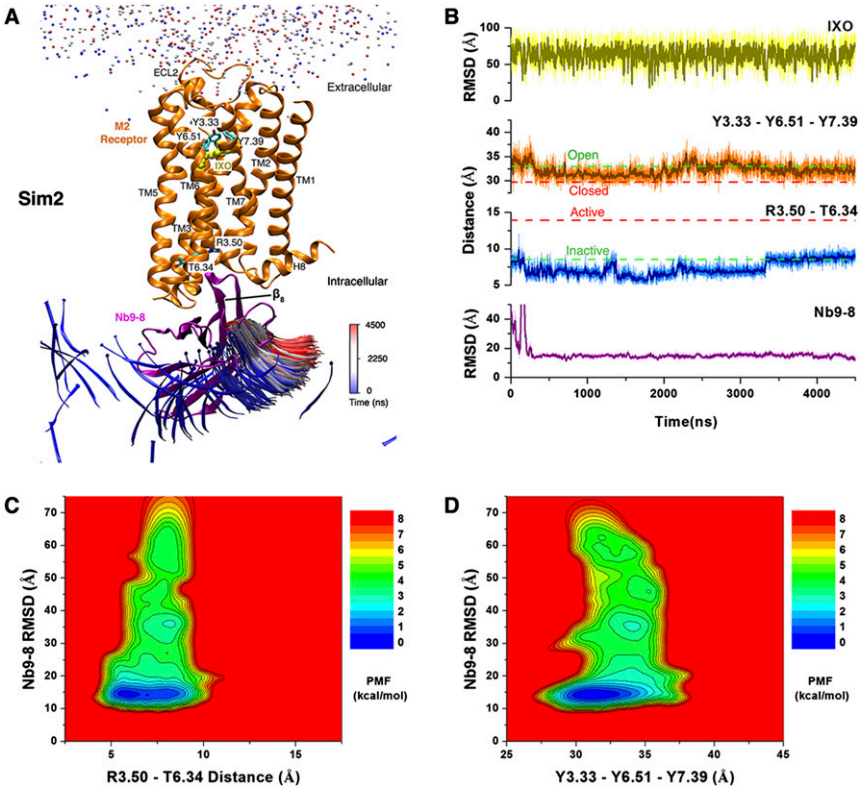


Fig. S1. One 4,500-ns GaMD simulation of the M_2 muscarinic receptor (Sim2 in Table 1), during which IXO and Nb9-8 bind to the receptor with the smallest rmsds of 15.12 and 10.02 Å, respectively, compared with the 4MQS X-ray conformation. (*A*) Trajectories of the nitrogen atom in the trimethylamine group of IXO (beads) and the β_8 strand of Nb9-8 (ribbons) that are colored by the simulation time in a blue (0 ns)–white (2,250 ns)–red (4,500 ns) scale. The structural representations are similar to Fig. 1*A*. (*B*) The rmsds of the IXO and Nb9-8 relative to the 4MQS X-ray conformations, Tyr104$^{3.33}$–Tyr403$^{6.51}$–Tyr426$^{7.39}$ triangle perimeter, and Arg121$^{3.50}$–Thr386$^{6.34}$ distance are plotted as a function of time during the 4,500-ns GaMD simulation. The dashed lines indicate the X-ray structural values of the M_2 receptor (3UON, green; 4MQS, red). (*C*) The 2D PMF calculated with the Arg121$^{3.50}$–Thr386$^{6.34}$ distance and rmsd of the Nb9-8 relative to the 4MQS X-ray conformation. (*D*) The 2D PMF calculated with the Tyr104$^{3.33}$–Tyr403$^{6.51}$–Tyr426$^{7.39}$ triangle perimeter and rmsd of the Nb9-8 relative to the 4MQS X-ray conformation.

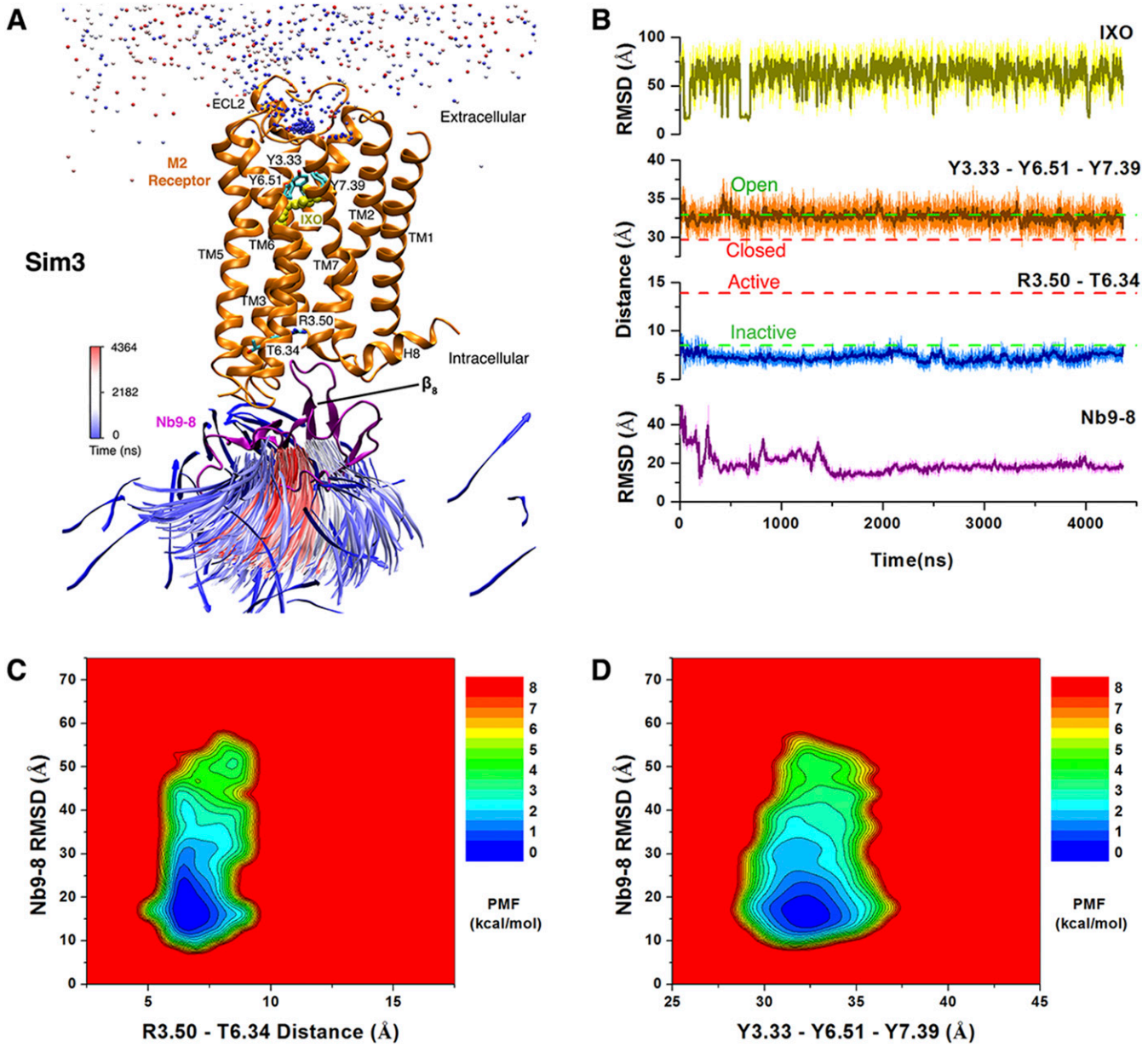


Fig. S2. One 4,364-ns GaMD simulation of the M_2 muscarinic receptor (Sim3 in Table 1), during which IXO and Nb9-8 bind to the receptor with the smallest rmsds of 11.37 and 8.38 Å, respectively, compared with the 4MQS X-ray conformation. (*A*) Trajectories of the nitrogen atom in the trimethylamine group of IXO (beads) and the β_8 strand of Nb9-8 (ribbons) that are colored by the simulation time in a blue (0 ns)–white (2,182 ns)–red (4,364 ns) scale. The structural representations are similar to Fig. 1*A*. (*B*) The rmsds of the IXO and Nb9-8 relative to the 4MQS X-ray conformations, Tyr104$^{3.33}$–Tyr403$^{6.51}$–Tyr426$^{7.39}$ triangle perimeter, and Arg121$^{3.50}$–Thr386$^{6.34}$ distance are plotted as a function of time during the 4,500-ns GaMD simulation. The dashed lines indicate the X-ray structural values of the M_2 receptor (3UON, green; 4MQS, red). (*C*) The 2D PMF calculated with the Arg121$^{3.50}$–Thr386$^{6.34}$ distance and rmsd of the Nb9-8 relative to the 4MQS X-ray conformation. (*D*) The 2D PMF calculated with the Tyr104$^{3.33}$–Tyr403$^{6.51}$–Tyr426$^{7.39}$ triangle perimeter and rmsd of the Nb9-8 relative to the 4MQS X-ray conformation.

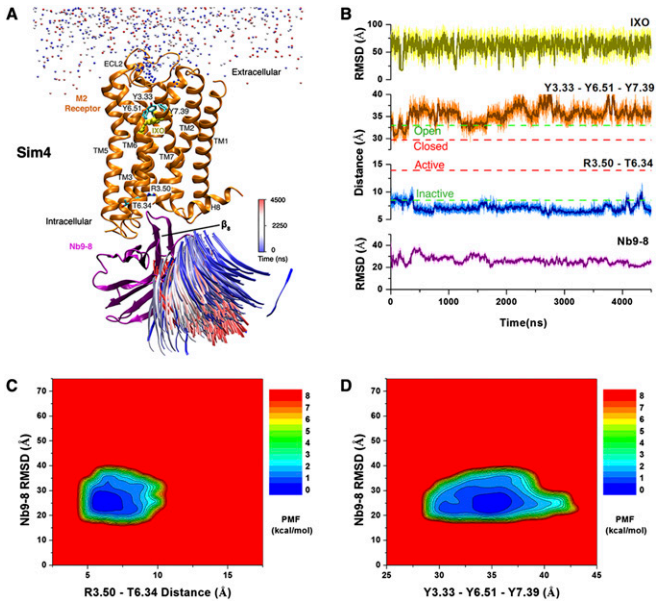
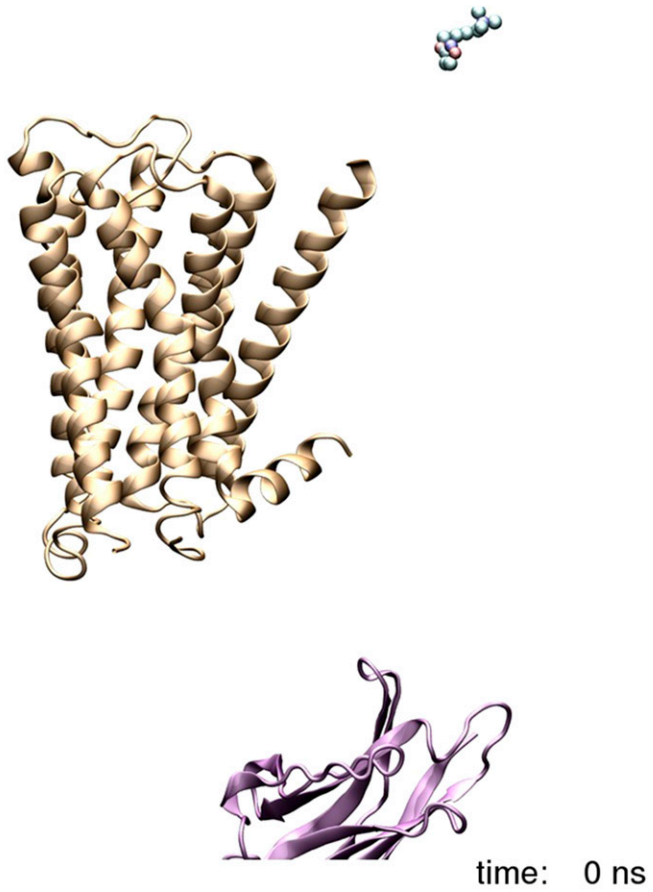


Fig. S3. One 4,500-ns GaMD simulation of the M_2 muscarinic receptor (Sim4 in Table 1), during which IXO and Nb9-8 bind to the receptor with the smallest rmsds of 14.57 and 17.19 Å, respectively, compared with the 4MQS X-ray conformation. (*A*) Trajectories of the nitrogen atom in the trimethylamine group of IXO (beads) and the β_8 strand of Nb9-8 (ribbons) that are colored by the simulation time in a blue (0 ns)–white (2,250 ns)–red (4,500 ns) scale. The structural representations are similar to Fig. 1*A*. (*B*) The rmsds of the IXO and Nb9-8 relative to the 4MQS X-ray conformations, Tyr104$^{3.33}$–Tyr403$^{6.51}$–Tyr426$^{7.39}$ triangle perimeter, and Arg121$^{3.50}$–Thr386$^{6.34}$ distance are plotted as a function of time during the 4,500-ns GaMD simulation. The dashed lines indicate the X-ray structural values of the M_2 receptor (3UON, green; 4MQS, red). (*C*) The 2D PMF calculated with the Arg121$^{3.50}$–Thr386$^{6.34}$ distance and rmsd of the Nb9-8 relative to the 4MQS X-ray conformation. (*D*) The 2D PMF calculated with the Tyr104$^{3.33}$–Tyr403$^{6.51}$–Tyr426$^{7.39}$ triangle perimeter and rmsd of the Nb9-8 relative to the 4MQS X-ray conformation.



Movie S1. Binding of the agonist IXO and G-protein mimetic nanobody Nb9-8 to the M_2 muscarinic GPCR in a 4,500-ns GaMD simulation. Although the agonist reached only the receptor extracellular vestibule, the nanobody was observed to bind the receptor intracellular G-protein-coupling site with a minimum rmsd of 2.48 Å in the core domain compared with the X-ray structure. The nanobody approached the M_2 receptor on the intracellular side and formed initial contacts with the ICL2. The TM6 cytoplasmic end became disordered in the C-terminal region of ICL3 and facilitated binding of the nanobody. Then, the nanobody rearranged its conformation and fit into the receptor intracellular pocket. The nanobody interacted dynamically with the receptor flexible ICLs and the TM7–H8 hinge. The nanobody rmsd dropped <5 Å during ~3,880–4,000 ns. In the absence of the agonist IXO at the orthosteric site, the nanobody deviated from the target G-protein-coupling site during ~4,000–4,500 ns while maintaining interactions with the receptor intracellular domains.

Movie S1