



Supplementary Information for

Structural basis for ligand modulation of CCR2 conformational landscape

Bryn C. Taylor, Christopher T. Lee, Rommie E. Amaro

Corresponding Author

E-mail: ramaro@ucsd.edu

This PDF file includes:

Supplementary text
Figs. S1 to S16
Tables S1 to S2
References for SI reference citations

Supporting Information Text

Methods

System Preparation. Two systems were simulated: holo CCR2, with both co-crystallized antagonist ligands bound, and apo CCR2, without the ligands bound. Each all-atom system is embedded in a biologically similar POPC bilayer and explicitly solvated with TIP3P. The initial coordinates were taken from the experimental crystal structure([1](#)) and simulated for 50ns MD simulations on local resources before simulation on Anton2. All simulations are in a POPC lipid bilayer and cubic water box with 150mM NaCl, at pH 7.4, at 310K and 1 bar.

The two small molecules CCR2-RA-[R] and BMS 681([1](#)) were deleted to build the apo system. Both systems were protonated at pH 7.4 in Maestro-integrated PROPKA. A POPC lipid bilayer was added to each system and solvated with TIP3 waters and 0.15 M NaCl using CHARMM-GUI([2](#)). The small molecules were parameterized with CGenFF([3](#)). System coordinates were parameterized with the CHARMM36([4](#)) force field. No restraints were added.

Modification of CCR2 coordinates. The coordinates for CCR2 were taken from the experimental crystal structure ([1](#)) and modified to build the apo (unbound) and holo (dual-antagonist-bound) systems. For both systems, the T4 Lysozyme was removed from the crystal structure and intracellular loop 3 and part of the ECL3 and N terminus was constructed. For ICL3, a peptide containing residues 223:243 was built ab-initio, the backbones of residues 223:231,236:243 and the side chains of residues 223:226,241:243 were tethered to their respective positions, the receptor represented as a set of potential grid maps representing vw, el, hb, and sf "potentials" and, and the peptide was sampled in these maps. For NT/ECL3, the protocol is similar except that 2 separate peptides are built (31:41 and 276:285), a disulfide bond is imposed between 21 and 277, and the entire thing is sampled as above. There are several zero-occupancy side chains whose conformations are predicted as a part of this simulation.

Best scoring conformations of the two fragments are merged with the rest of receptor coordinates and the system is minimized in its full-atom representation: first by exhaustively sampling polar rotatable hydrogens, then by minimizing the side-chain conformers, then by Monte-Carlo sampling of side-chain conformers, then by minimizing everything. During these steps, harmonic restraints of gradually decreasing strength are imposed between the model and either the X-ray coordinates or the best prediction conformations of the built regions. Towards the end of the optimization, the restraints were released almost entirely and the complex remained stable. This was done in the presence of ligands. Zn ion and water molecules were removed.

Molecular Dynamics Simulations. Both systems were minimized and equilibrated using the GPU version of AMBER12. The systems were minimized at NPT for 15,000 total steps and were equilibrated for 2 sequential 25 nanosecond runs. The systems were then simulated for 50 nanoseconds in the NPT ensemble at 310K and 1 bar with 2-fs time-step and particle mesh Ewald electrostatic approximation. The additional replicates were made from the final production output by simulating for three additional nanoseconds to scramble the input velocities.

MD simulations on Anton2 were performed on the final coordinates from the short GPU-enabled AMBER12 simulations. The Anton2 simulations were run in the NPT ensemble, using Anton's Nose-Hoover thermostat-barostat, at 310K and 1 bar with a 2.5-fs time-step and particle mesh Ewald electrostatic approximation. The two systems were simulated for an aggregate total of 260 microseconds.

- Ensemble and Constraints: After minimization, equilibration, and initial production runs, both simulations were run as NPT ensembles using Anton's Multigrator framework. No constraints were used in the simulations.
- Boundaries: The fully solvated systems are cubic with X, Y, Z unit lengths of 72 Å, 72 Å, and 103 Å respectively.
- Force Fields: CHARMM36 force field with TIP3P waters; small molecule ligands were parameterized with CGenFF.
- Atom Count and Types: Each system contains 55,000 atoms. The systems are composed of the protein, POPC lipids, water, small molecule drugs, Na⁺, and Cl⁻.

Trajectory Preparation. MD trajectories were processed using VMD([5](#)). All frames were aligned to the first frame using all residues of the protein. The frame rate for trajectories was 240ps, the standard for Anton2 simulations. The trajectories were converted into NAMD's .dcd trajectory format for analysis with PyEMMA([6](#)), MSMBuild([7](#)), and in-house scripts.

Markov State Models. A Markov State Model (MSM) is a time-dependent master equation that describes the probability of transitioning between discrete states at a fixed time interval. These models are required to have the Markovian property (i.e., the probability of transitioning between discrete states is independent of previous transitions). By clustering protein structures extracted from an MD trajectory, discrete conformational states can be identified for use in MSMs([7–10](#)). Transitions between conformational clusters observed over the course of an MD trajectory are tallied, and the MSM is then built from the transition probabilities between these distinct clusters. MSM/MD analysis provides access to the thermodynamic, kinetic, and structural characteristics of the protein conformational ensemble (i.e., a robust description of the free-energy landscape of the protein)([7–12](#)). The thermodynamics of the various conformational states can be calculated from the equilibrium distribution. It is also possible to resolve the transition kinetics between individual states, the concerted or principal protein motions, metastable states, and the transition pathways between discrete states([7, 9, 11](#)). Lastly, the source molecular dynamics simulations provide representative cluster structures for use in structure-based drug design([12](#)).

Building the Markov-State Models. PyEMMA was used for feature selection, dimensionality reduction, clustering, MSM building, and MSM analysis. Dimensionality reduction using tICA(13, 14) was performed on all trajectories.

We used time-structure independent component analysis (tICA)(13–15) starting with all pairwise inter-residue distances to perform dimensionality reduction and identify the features and collective variables (time-structure based independent components (tICs)) that best represent the dominant slow motions in the apo and holo simulations. To reduce the number of pairwise distances to a manageable number, we employed an iterative tICA feature selection approach. tICA was run on a curated starting set of hundreds of distances between transmembrane helices. Features with low tIC correlation were then removed from the set and tICA run on the resultant new basis. In this iterative fashion, we selected 22 representative features, listed below. These resultant features are the sets of distances between residues in the orthosteric and allosteric ligand-binding pockets.

The apo and holo systems were clustered separately using K-means. Two MSMs were built: on just the apo data, and on just the holo data. The MSMs were selected based on implied timescale plots, and Chapman-Kolmogorov tests. MSMs were coarse-grained with HMMs. Representative states were selected by identifying the most populated cluster in each metastable state, and using MSMBuilder(7) to find the centroid of that cluster.

Set of 22 Features:

- Ile 40, Asn 199
- Tyr 222, Arg 138
- Tyr 305, Arg 138
- Tyr 305, Tyr 222
- Tyr 49, Thr 292
- Tyr 120, Glu 291
- Tyr 120, Tyr 259
- Glu 291, Tyr 259
- Tyr 49, Trp 98
- Trp 98, Tyr 120
- Trp 98, Glu 291
- Trp 98, Thr 292
- Tyr 49, Tyr 259
- Phe 246, Leu 81
- Ile 245, Leu 134
- Ile 245, Leu 81
- Val 244, Tyr 305
- Tyr 305, Leu 81
- Tyr 305, Val 63
- Tyr 305, Leu 134
- Tyr 305, Leu 67
- Leu 67, Val 244

	Apo			Holo					
	1	2	3	1	2	3			
Simulation Number	1	2	3	1	2	3			
Simulation Time in microseconds	50	50	10	50	50	50			
Ligands Bound	None			BMS-681 and CCR2-RA-[R]					
Number of Atoms	53,097			53,077					
Membrane Lipids	POPC								
Water Model	TIP3P								
Force Field	CHARMM36 FF								
Box Dimensions	72 Å, 72 Å, 103 Å								

Table S1. System Information

RMSD over Simulation Time; Apo and Holo CCR2

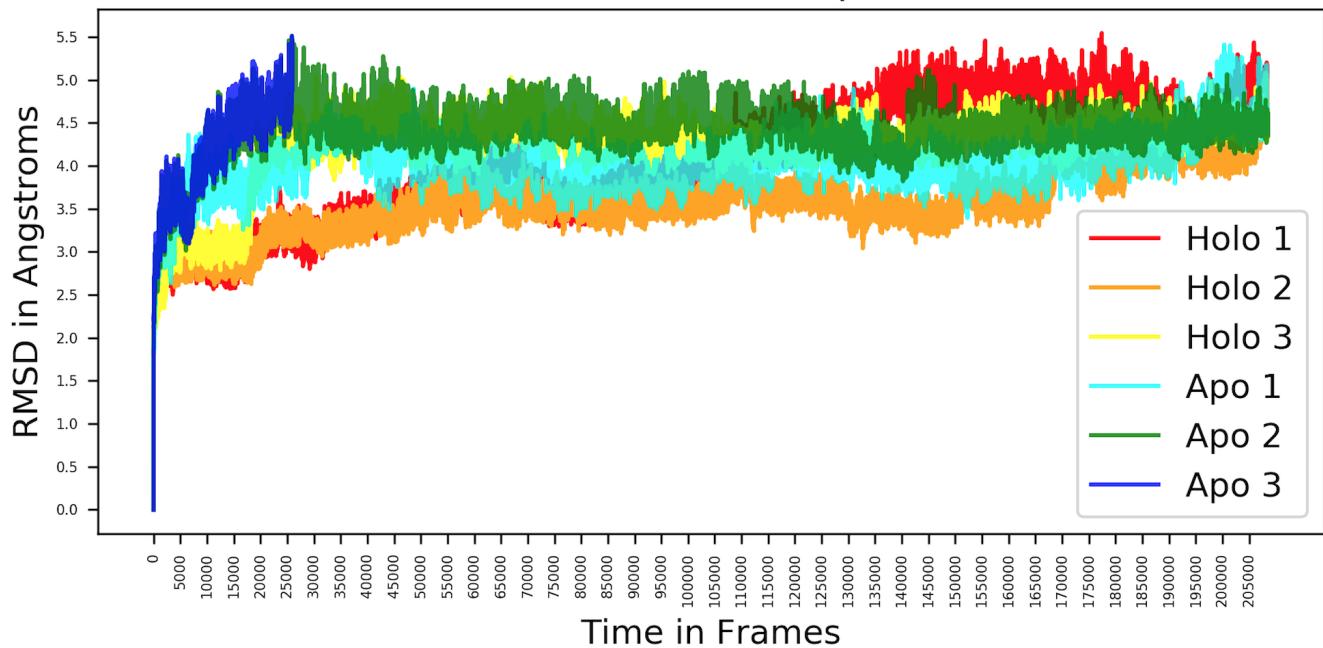


Fig. S1. RMSD plot of all trajectories over simulation time.

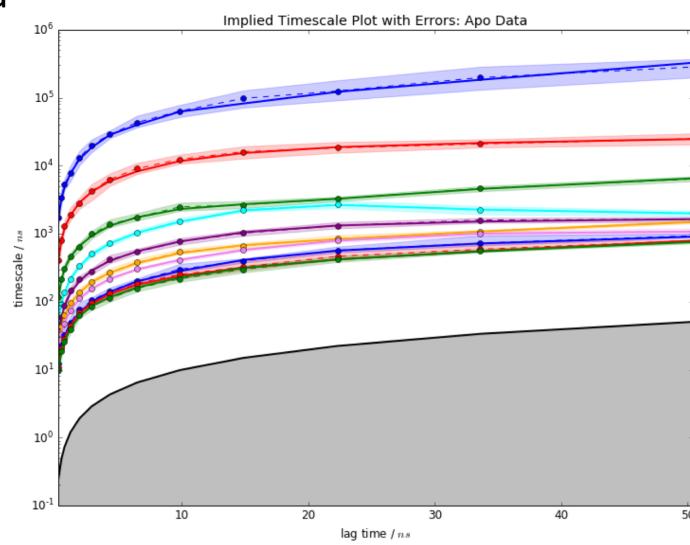
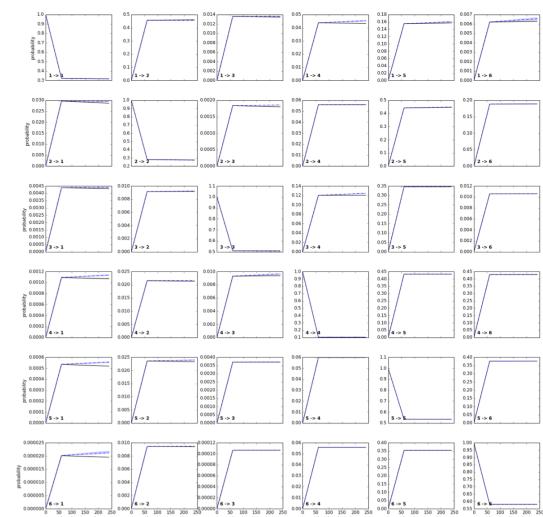
a**b**

Fig. S2. A) Implied Timescale Plot and B) CK Test for apo CCR2.

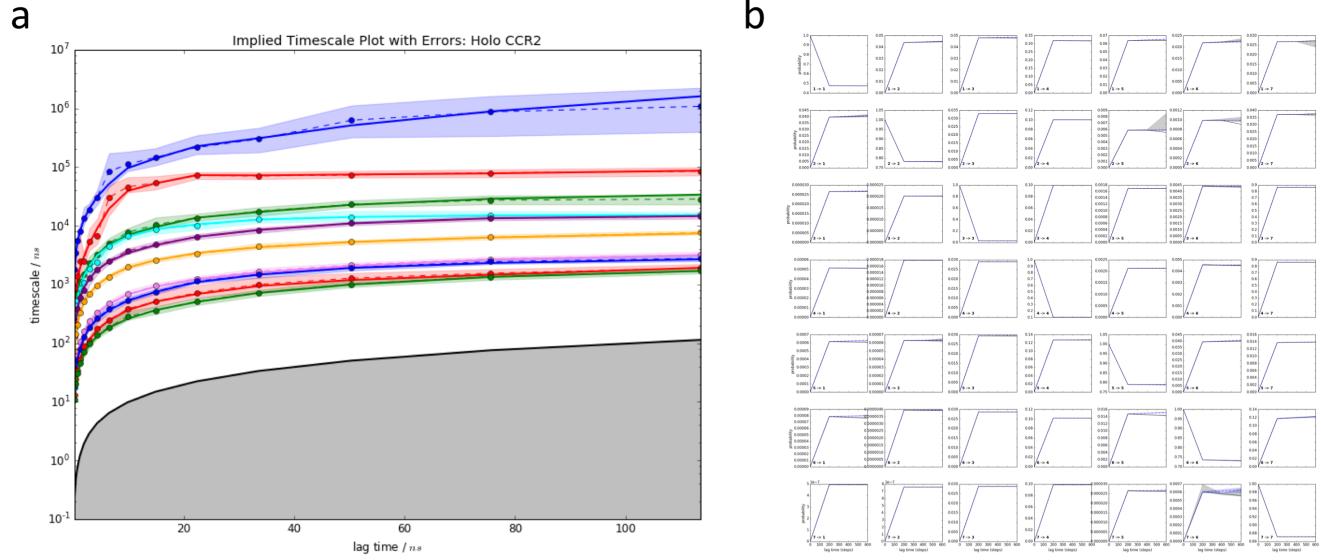


Fig. S3. A) Implied Timescale Plot and B) CK Test for holo CCR2

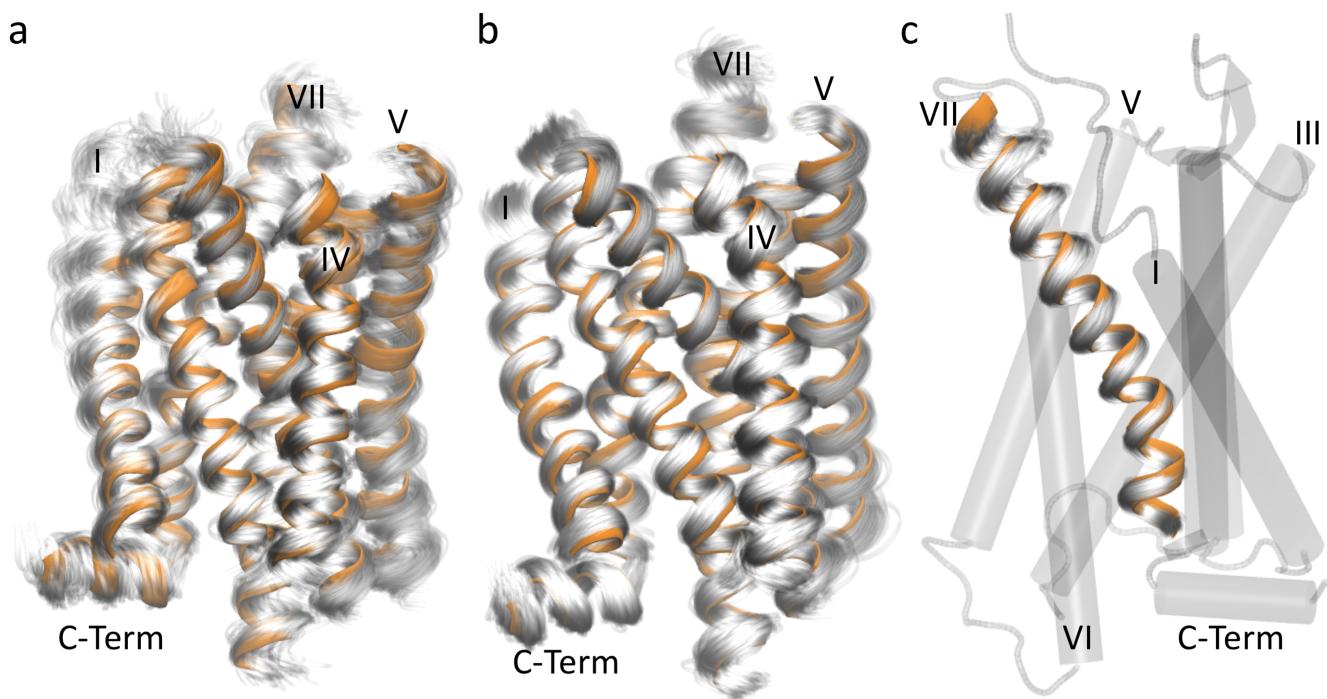


Fig. S4. The centroid of apo macrostate A, in orange, is compared to A) the 214 other microstate centroids in white (average RMSD of alpha helices from representative structure: 2.01 Å, standard deviation 0.50 Å), and B) the 1,024 frames from its microstate in white (average RMSD of alpha helices from centroid: 0.872 Å, standard deviation 0.139 Å). C) Helix VII of the same microstate (average RMSD from centroid: 1.035 Å, standard deviation 0.259 Å.)

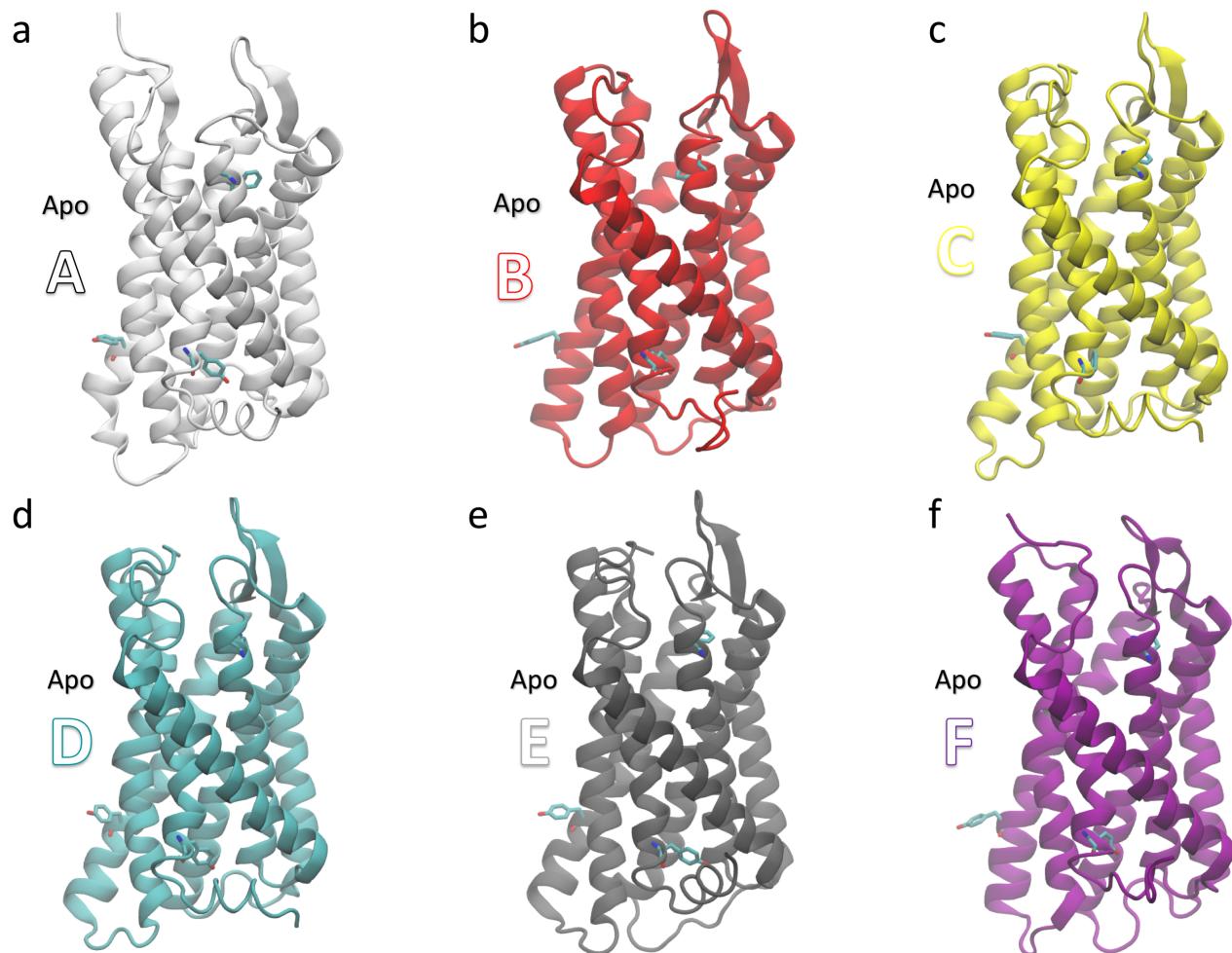


Fig. S5. The apo macrostates.

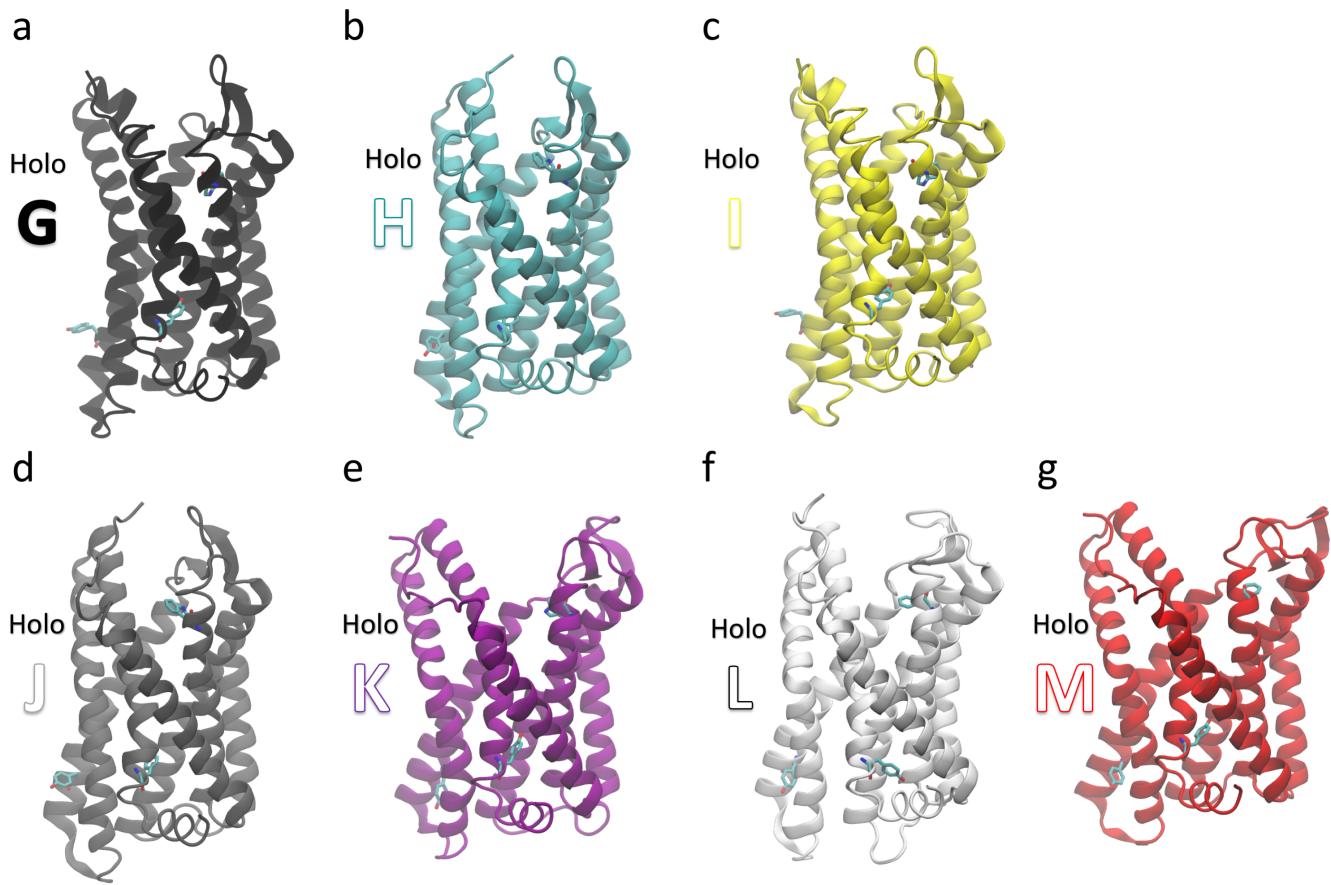


Fig. S6. The holo macrostates.

a

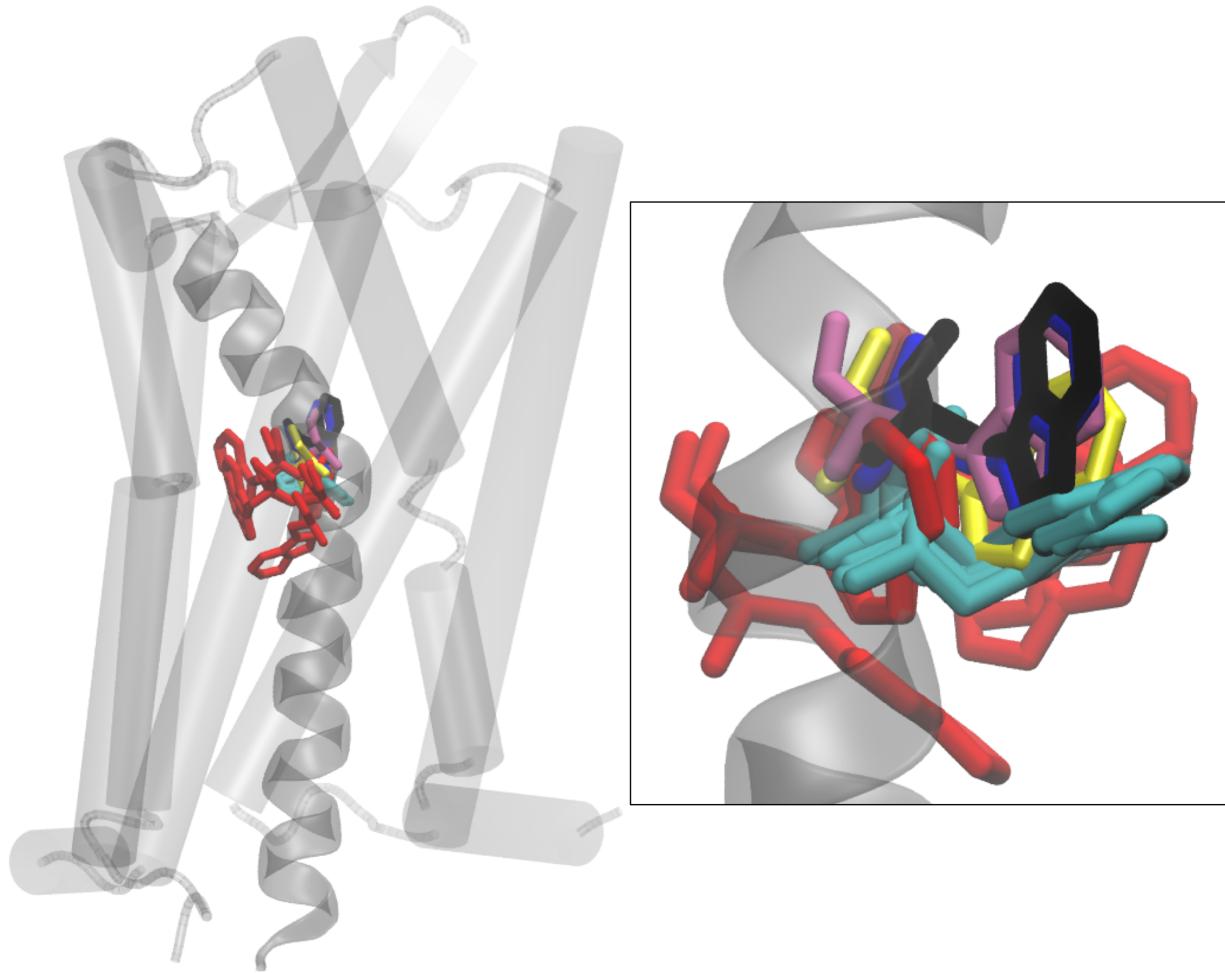
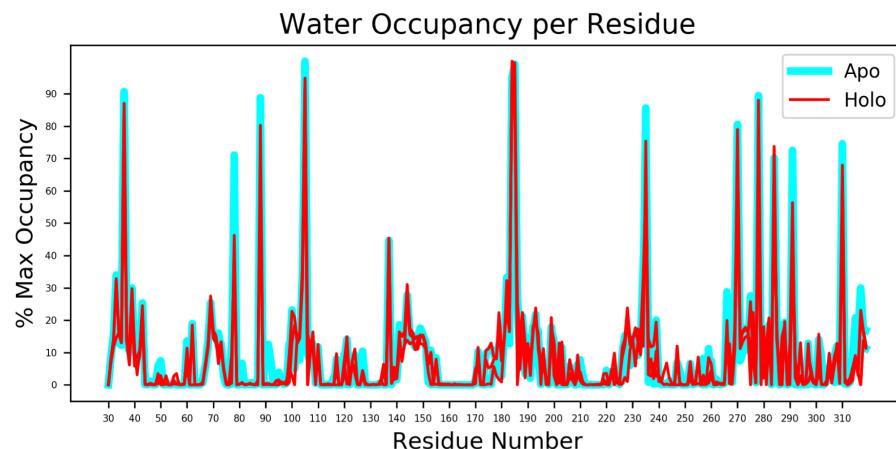


Fig. S7. Trp 256^{6,48} compared to the apo and holo macrostates and the crystal structures of CCR2 (PDB ID: 5T1A; grey, black) CCR5 (PDB ID: 4MBS; blue) and CXCR4 (PDB ID: 4RWS; mauve). A) Each apo macrostate shows Trp 256^{6,48} in a single conformation pointing toward helix III. In the holo macrostates, Trp 256^{6,48} access three distinct conformations: one resembling the crystal structure but with the helix shifted slightly outward from the helical core, another that laterally twists toward helix V, and one conformation that points down into the helical core toward the G-protein binding site.

a



b

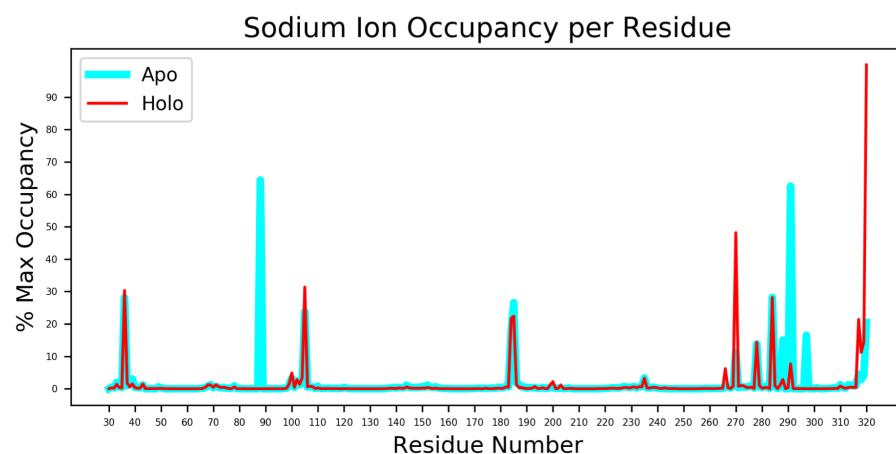


Fig. S8. Ligands disrupt a continuous internal water and sodium ion pathway. A) Water occupancy per residue in all apo (teal) and holo (red) simulations. B) Sodium ion occupancy per residue in all apo and holo simulations.

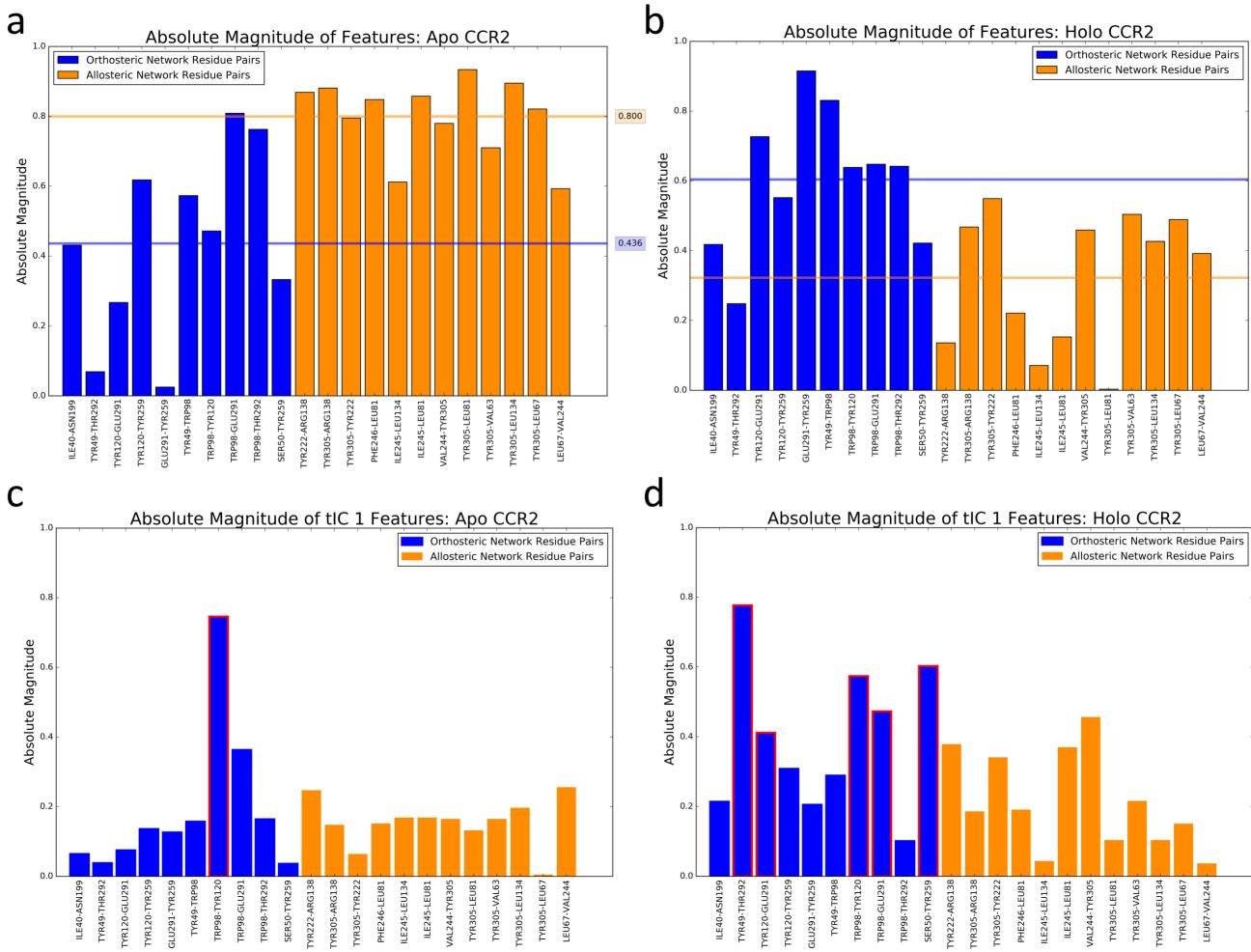


Fig. S9. The absolute magnitude of each input feature in A) apo tIC 0, B) holo tIC 0, C) apo tIC 1, and D) holo tIC 1. Each bar represents the absolute magnitude of one inter-residue distance. Blue bars are distances between residue pairs in the orthosteric pocket, and orange bars are distances between residue pairs in the allosteric pocket.

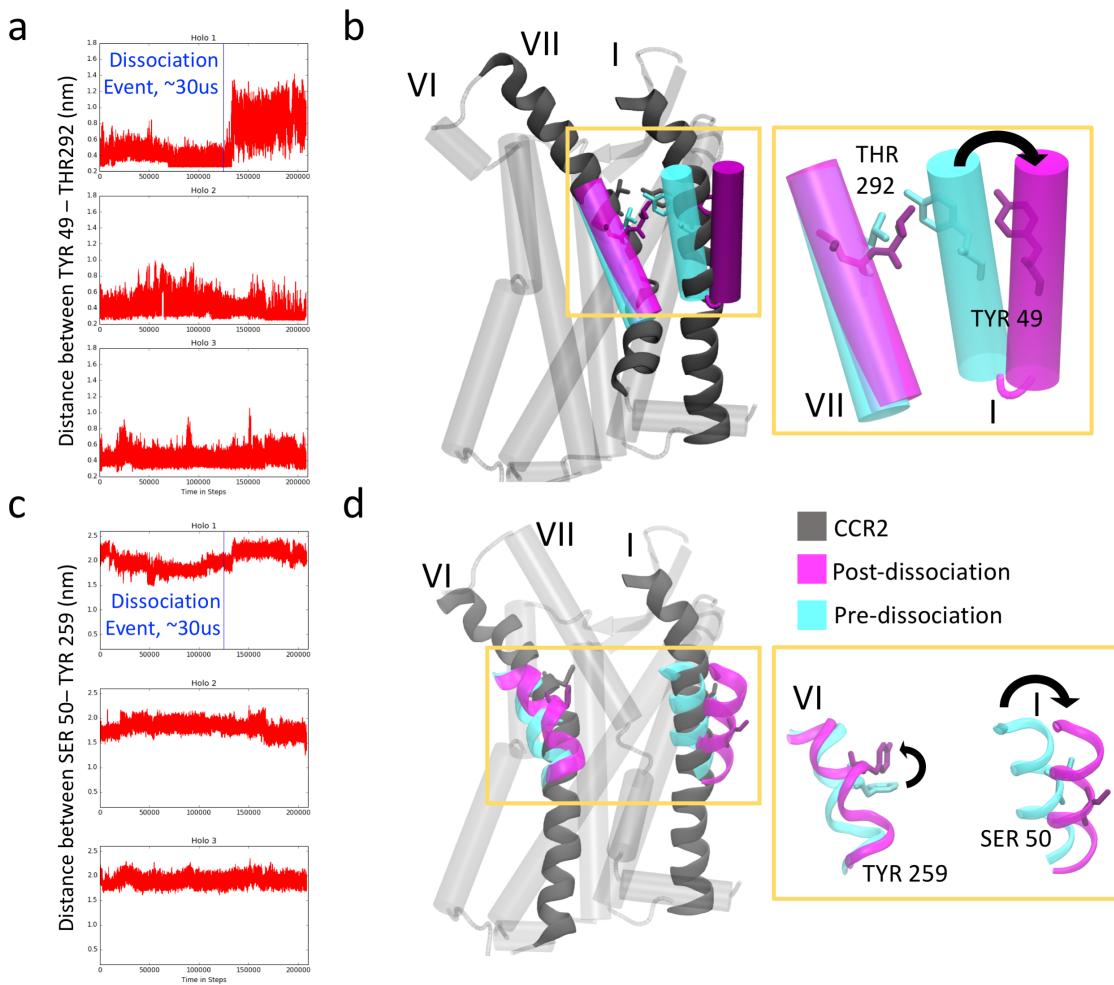


Fig. S10. A) Distance between residues TYR 49^{1,39} – THR 292 over simulation time. Dissociation event noted by blue line. B) Conformation of CCR2 before (teal) and after (purple) ligand dissociation. C) Distance between residues SER 50^{1,40} and TYR 259 over simulation time. D) Conformation of CCR2 before (teal) and after (purple) ligand dissociation.

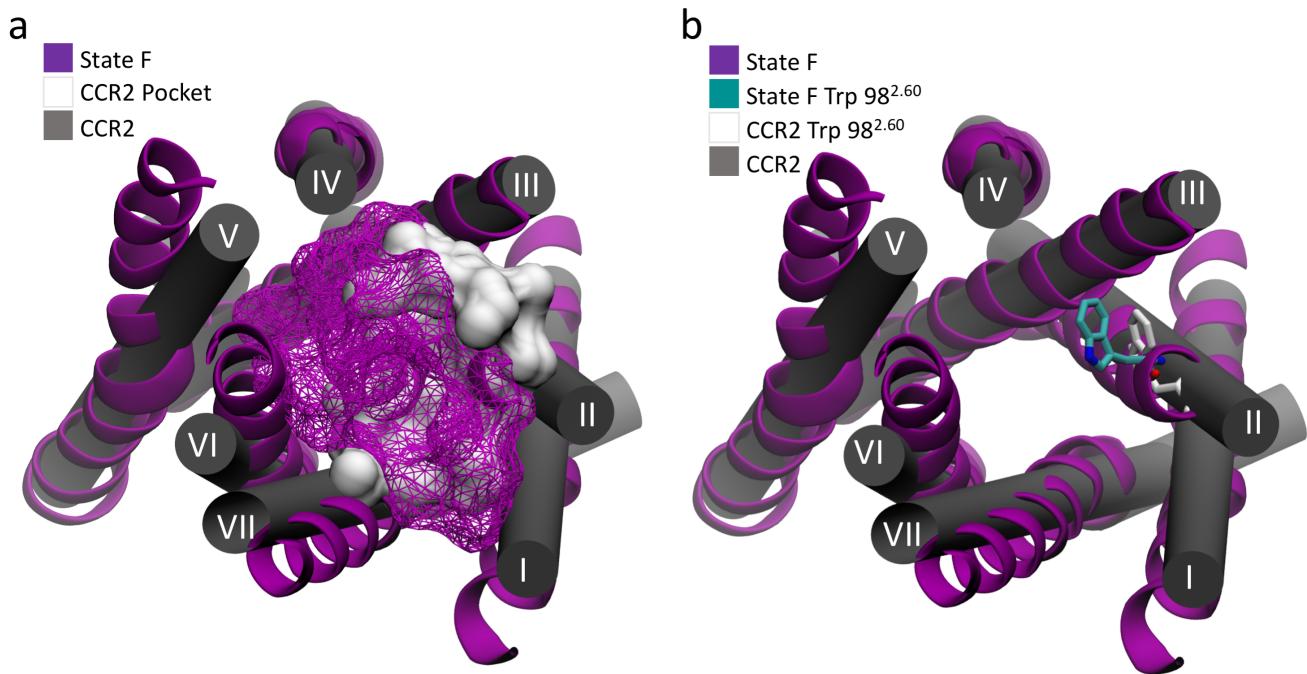


Fig. S11. A) The shape of the chemokine binding site of apo state F in comparison to the crystal structure. Without the ligand, the binding site expands and rotates toward helices IV, V, and VI, and extends between helices I and VII. B) The conformations of Trp98^{2,60} in apo state F and the crystal structure. Trp98^{2,60} protrudes into the pocket in the absence of ligands.

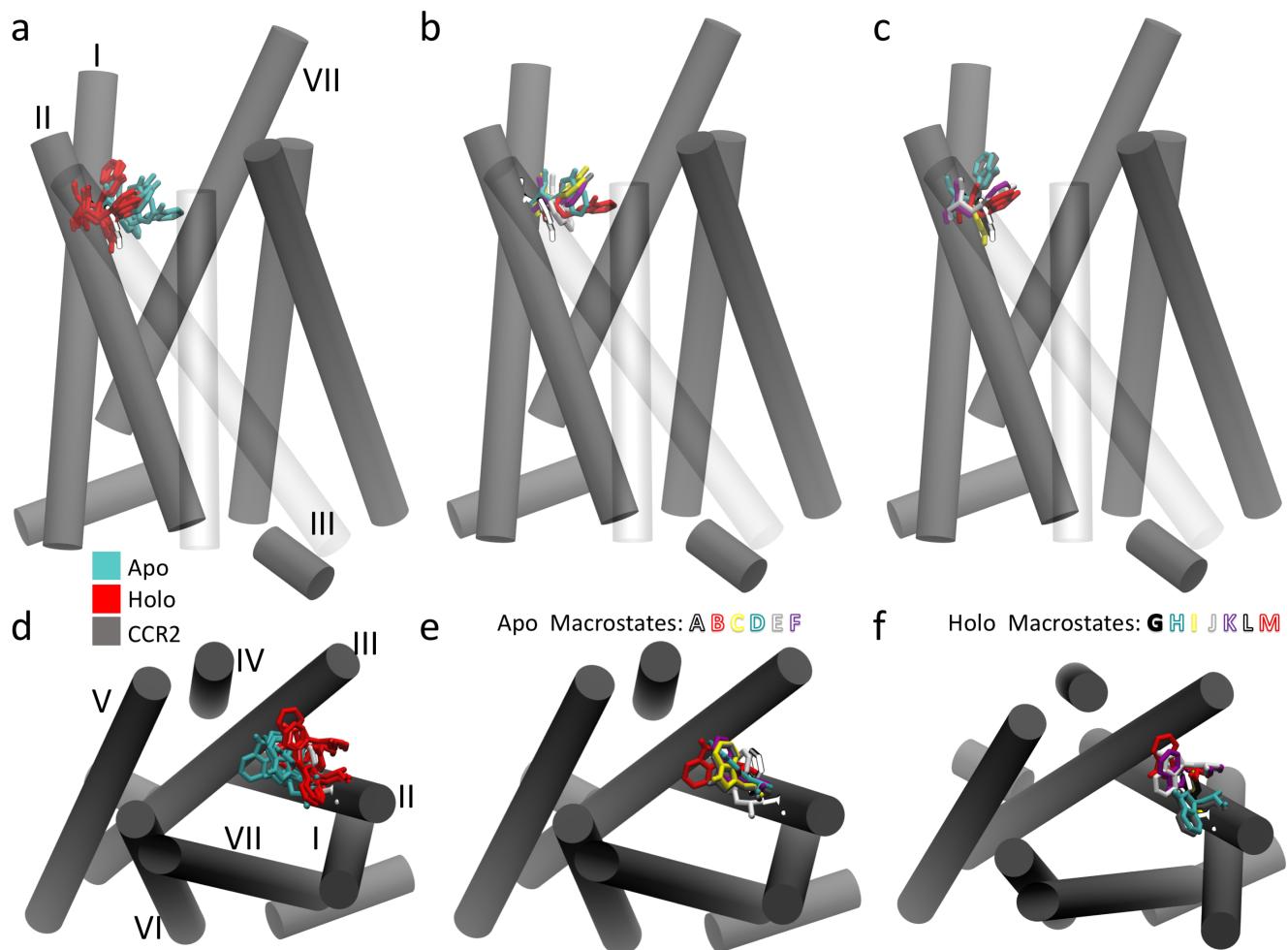


Fig. S12. Trp 98^{2,60} extends farther into the chemokine binding site in apo CCR2 than in the holo crystal structure and holo macrostates. A side view of the CCR2 crystal structure (grey) and Trp 98^{2,60} (bright white) compared to A) the apo (cyan) and holo (red) macrostate conformations of Trp 98^{2,60} B) the holo conformations, and C) the apo conformations. The extracellular-to-intracellular view of A-C).

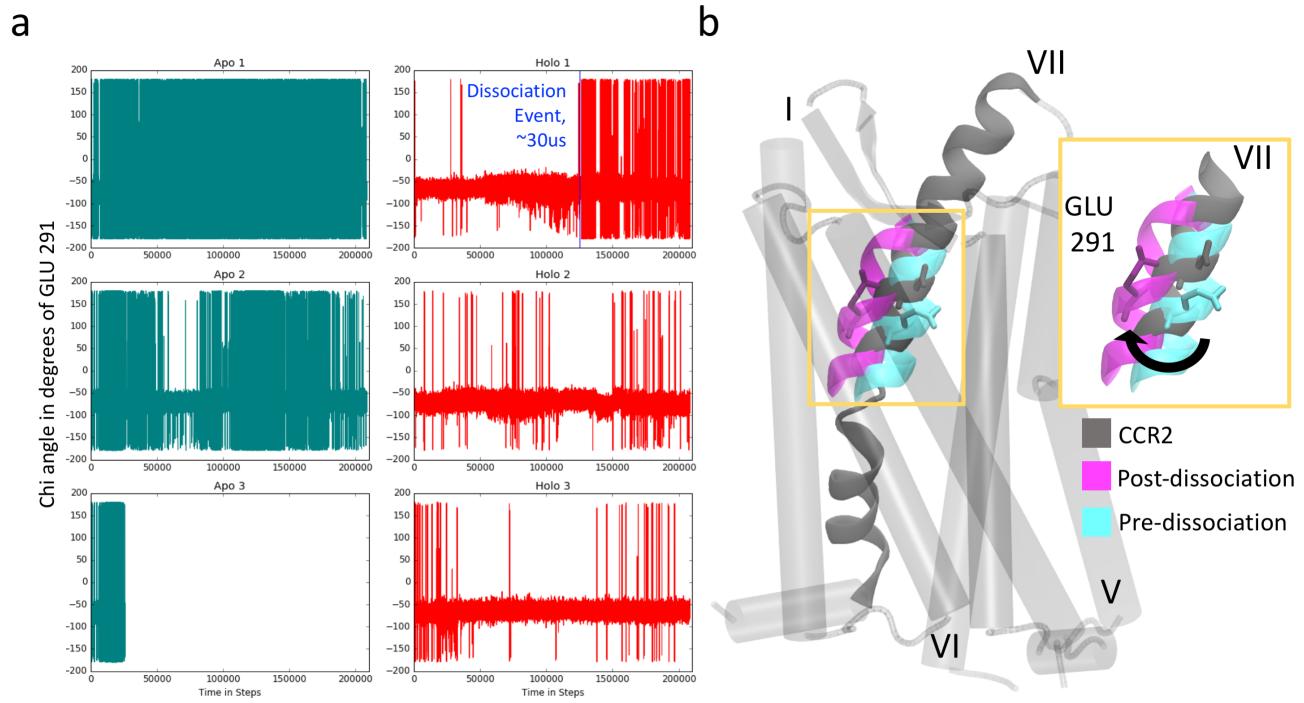


Fig. S13. A) Chi angle of GLU 291 over simulation time. Dissociation event noted by blue line. B) Conformation of CCR2 before (teal) and after (purple) ligand dissociation.

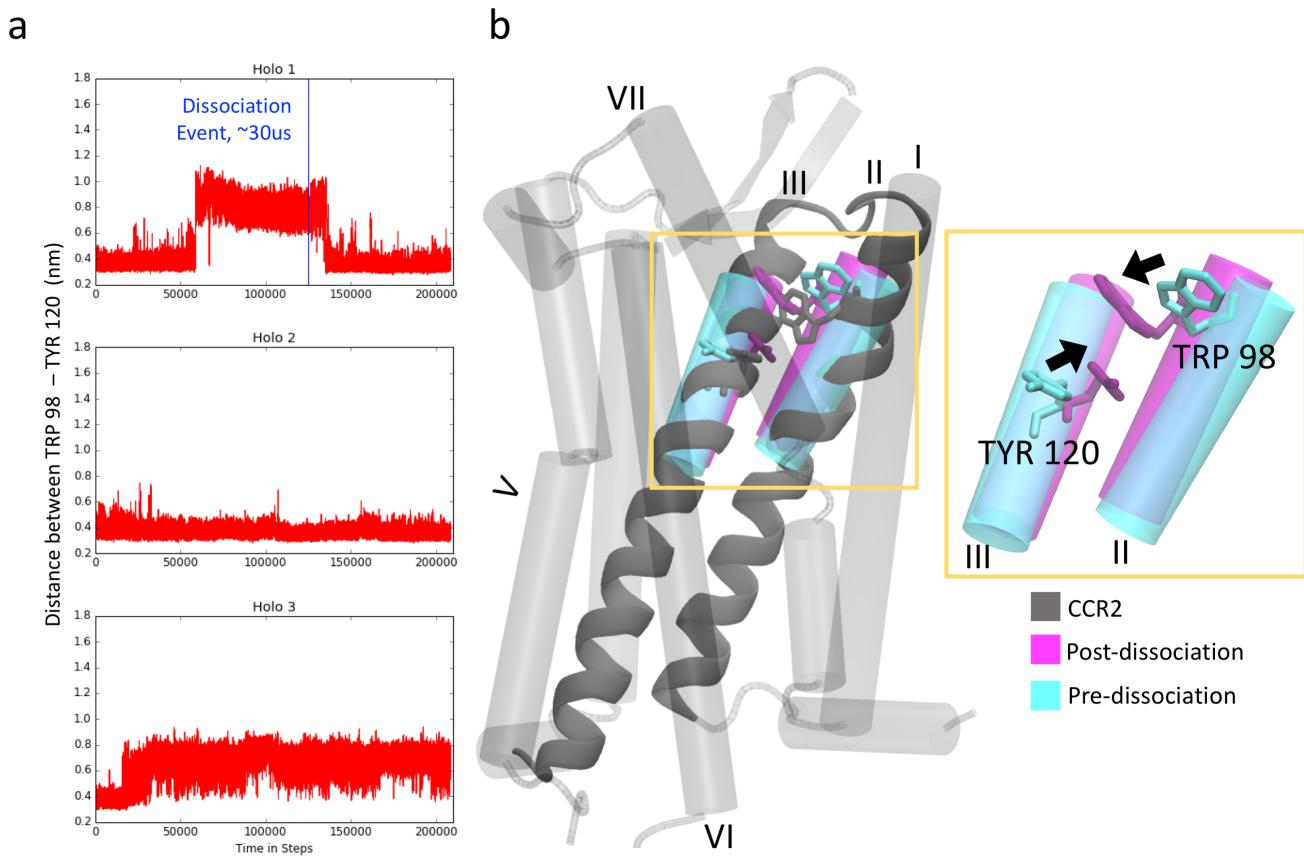


Fig. S14. A) Distance between residues TRP 98^{2,60} and TYR 120^{3,32} over simulation time. Dissociation event noted by blue line. B) Conformation of CCR2 before (teal) and after (purple) ligand dissociation.

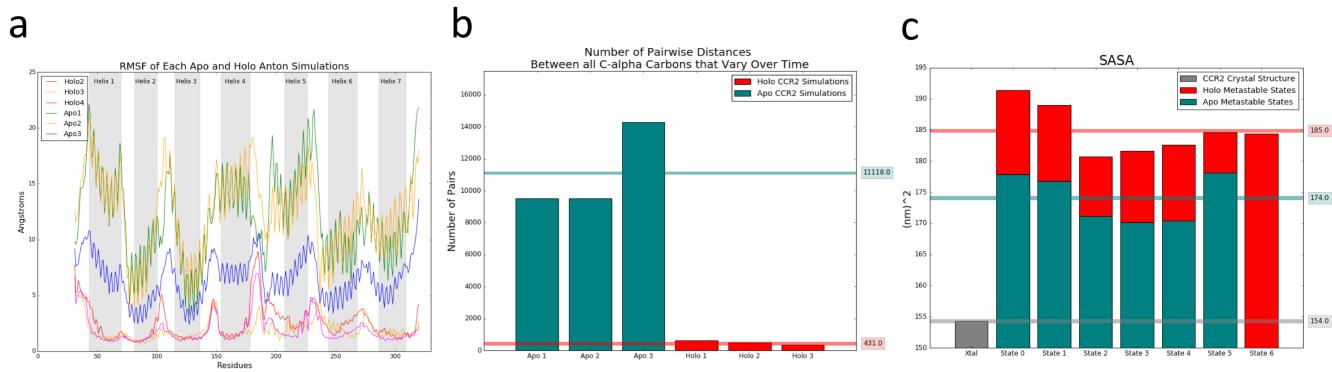


Fig. S15. A) RMSF of each residue for individual simulations of apo and holo CCR2. B) Number of pairwise distances between C-alpha carbons that vary over time for each simulation. C) SASA for each metastable macrostate of apo and holo, compared to the CCR2 crystal structure.

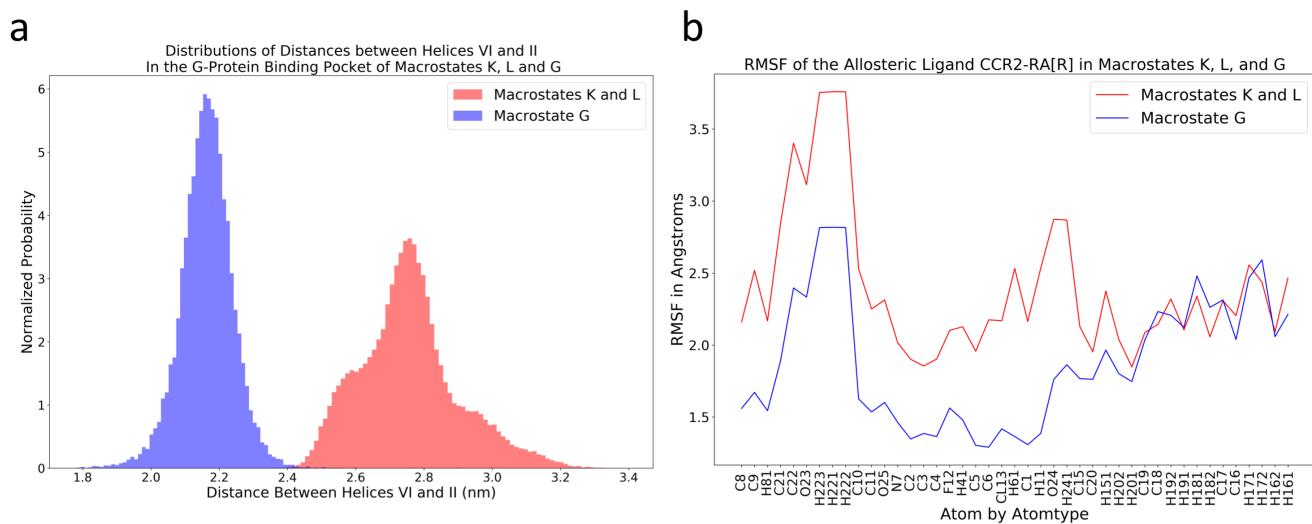


Fig. S16. The RMSF of the allosteric ligand is larger in macrostates with more open G-protein binding sites. A) Distribution of distances in macrostates K, L and G between residues GLY 224 and Asp 78 in the intracellular regions of helices VI and II. B) RMSF of the allosteric ligand in macrostates K, L, and G.

Apo	321.6 μ s	57.7 μ s	11.1 μ s	9.0 μ s	4.2 μ s	2.8 μ s	2.4 μ s	1.7 μ s	1.3 μ s	1.3 μ s
Holo	2246.9 μ s	286.8 μ s	84.9 μ s	59.1 μ s	44.3 μ s	21.6 μ s	8.2 μ s	7.7 μ s	4.4 μ s	3.9 μ s

Table S2. Relaxation timescales for the apo and holo MSMs. Units are in microseconds.

References

1. Zheng Y, et al. (2016) Structure of CC chemokine receptor 2 with orthosteric and allosteric antagonists. *Nature* 540(7633):458–461.
2. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* 29(11):1859–1865.
3. Vanommeslaeghe K, et al. (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* 31(4):671–690.
4. Huang J, Mackerell AD (2013) CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry* 34(25):2135–2145.
5. Humphrey, W. Dalke A, Schulten K (1996) VMD - Visual Molecular Dynamics. *J. Molec. Graphics* 14(6226):33–38.
6. Scherer MK, et al. (2015) PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *Journal of Chemical Theory and Computation* 11(11):5525–5542.
7. Beauchamp KA, et al. (2011) MSMBuilder2: Modeling conformational dynamics on the picosecond to millisecond scale. *Journal of Chemical Theory and Computation* 7(10):3412–3419.
8. Pande VS, Beauchamp K, Bowman GR (2010) Everything you wanted to know about Markov State Models but were afraid to ask.
9. Prinz JH, Keller B, Noe F (2011) Probing molecular kinetics with Markov models: metastable states, transition pathways and spectroscopic observables. *Phys. Chem. Chem. Phys.* 13:16912–16927.
10. Prinz JH, et al. (2011) Markov models of molecular kinetics: Generation and validation. *Journal of Chemical Physics* 134(17).
11. Senne M, Trendelkamp-Schroer B, Mey ASJS, Schütte C, Noé F (2012) EMMA: A software package for markov model building and analysis. *Journal of Chemical Theory and Computation* 8(7):2223–2238.
12. Cronkite-Ratcliff B, Pande V (2013) MSMExplorer: Visualizing Markov state models for biomolecule folding simulations. *Bioinformatics* 29(7):950–952.
13. Schwantes CR, Pande VS (2013) Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* 9:2000 – 2009.
14. Perez-Hernandez G, Paul F, Giorgino T, De Fabritiis G, Noe F (2013) Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* 139:015102.
15. Nüske F, Keller BG, Pérez-Hernández G, Mey ASJS, Noé F (2014) Variational approach to molecular kinetics. *Journal of Chemical Theory and Computation* 10(4):1739–1752. PMID: 26580382.