Supporting Information

Abrol et al. 10.1073/pnas.1413216111

SI Materials and Methods

Computational Methodology. We manually aligned the C-C chemokine receptor type 5 (CCR5) transmembrane (TM) sequence to the sequences of four available G protein-coupled receptors (GPCRs): bovine rhodopsin (PDB ID code 1u19), human β2 adrenergic receptor (PDB ID code 2rh1), human adenosine A2A receptor (PDB ID code 3eml), and turkey β1 adrenergic receptor (PDB ID code 2vt4) using the conserved residues found in each TM. We predicted the lengths of the CCR5 TMs based on the PredicTM approach that predicts the hydrophobic helical regions and combines them with the output of three prediction servers—Porter (1), APSSP2 (2) and PsiPred (3)—to predict any helical extensions beyond the membrane. The starting structure of the TM bundle for each template was obtained by mutating the template residues to the CCR5 residues (as in homology modeling) but using the SideChain Rotamer Energy Analysis Method (SCREAM) side-chain optimization protocol and a 100-step minimization of TMs (with the Dreiding force field) performed independently for each helix (i.e., without the presence of other helices) (4, 5). Afterward, the mutated and optimized helices were put back together to form a seven-helix bundle. This approach yielded four CCR5 seven-transmembrane helical domains, each based on a different experimental GPCR structure. In the next step, the BiHelix step (6) was performed, followed by the SuperBiHelix step (7), as described in *Prediction of the CCR5 Structure* in the main text.

Because some loop residues (e.g., K191) were suggested to be important for ligand binding, we modeled the extracellular loops of CCR5 before ligand docking, Thus, we built the extracellular loops and N terminus using homology approaches and the sequence similarity to human β2 adrenergic receptor loops (extracellular loop 1, 12 first residues of the extracellular loop 2) and human adenosine A2A receptor loops (last 14 residues of the extracellular loop 2, extracellular loop 3). We built the N terminus using the part of the bovine rhodopsin N terminus fused with the experimental NMR structure of the middle part of the CCR5 N terminus (PDB ID code 2RLL). After mutating the residues to the respective CCR5 residues using the SCREAM protocol, the loops were connected to the proper helices, and the connected regions were minimized for 1,000 steps using a Dreiding force field.

The intracellular loops and helix 8 were modeled after docking, using an approach similar to that used for the extracellular loops, with homology modeling and sequence similarity to human β2 adrenergic receptor loops.

Before the docking of each CCR5 ligand described in this work [Maraviroc (MVC), PF-232798 (PF), Aplaviroc (APL), and TAK-779 (TAK)], we carried out a search over conformational space. Here, torsionally diverse conformations for each ligand were generated by the Maestro MacroModel program using a 0–10 kcal/mol window and requesting 10,000 poses. These conformations were clustered at 2.0 Å heavy-atom rmsd into a number of diverse families whose family heads were then docked to the selected CCR5 models. This procedure yielded 17 different conformations for Aplaviroc, 31 conformations for Maraviroc, 35 conformations for PF-232798, and 22 conformations for TAK-779. Each of these conformations was docked to each of the ∼10 ensembles of protein conformations using the DarwinDock protocol.

The DarwinDock protocol used in this study relied on sampling a complete set of ligand poses (∼50,000) in the potential binding regions for each ligand conformation. In the first step, each of the CCR5 model binding sites had the bulky, hydrophobic residues (ILE, LEU, PHE, TRP, TYR, and VAL) mutated to alanine residues to allow flexibility in the binding site. Next, ∼300 poses were selected for each CCR5 model–ligand pair using the lowest-energy criterion a Dreiding force field and three different scoring criteria: total energy, polar energy (defined as the sum of coulombic and H-bonding energy contributions to the total energy), and hydrophobic energy (defined as Van der Waals energy contribution to the total energy). Afterward, the bulky hydrophobic residues were restored to the original ones, and the entire binding had the side chains optimized using the SCREAM protocol. This approach allows a different side-chain conformation for each of the 300 poses. In the next step, a 10-step optimization of the binding site was performed using a Dreiding force field from which we selected 50% of the lowest-energy poses for further refinement. The computational charge neutralization of the whole system was then performed by mutating the charged residues of the CCR5 (ASP, ARG, GLU, and LYS) to their charge-neutral forms as well as detaching the most basic hydrogen atom from Aplaviroc, Maraviroc, or PF-232798 to change the formal charge of these ligands from +1 to 0. These neutral protein-ligand systems were subjected to a 10-step minimization run, and half of the lowest-energy poses were selected for the final step. Here, a 100-step minimization of the entire ligand–protein complex was performed and the best systems were selected based on the lowest total energy of the neutral system.

The procedure of obtaining models of CCR5 mutants was the following for each mutation listed in Table 1. For each of the top 100 CCR5 models obtained in the SuperBihelix step, the specific residue was mutated, the respective side chain was optimized using the SCREAM protocol, and the entire protein was minimized for 100 steps. Then, the CCR5 mutant models were rescored using an average of total and interhelical energies with the Dreiding force field. This procedure allowed for the evaluation of the structural changes in the apo form of the protein upon mutations.

The top 10 mutant receptor structures were selected and used for ligand docking. In this step, a simpler docking protocol was used, in which, for each ligand, the final pose from the DarwinDock step was taken and its coordinates were matched to each of the 10 low-energy CCR5 model mutants, followed by a binding-site optimization. The complexes were evaluated, and one ligand–mutant complex with the lowest total energy was selected. This procedure allowed the ligand to "choose" a different conformation for each CCR5 mutation studied.

Experimental Generation of CCR5 Mutants. Experimental mutations in the ORF of the human CCR5 gene were carried out using the quick change lightning site-directed mutagenesis kit (Agilent Technologies) following the kit's recommendations. Briefly, forward and reverse DNA primers (10 pmol each) encompassing the amino acid to be mutated were mixed with kit components and 40 ng of the mammalian expression vector pcDNA6.2 /V5/GW/D-Topo (Invitrogen) containing the human CCR5 ORF as a template.

After 18 rounds of PCR amplification, the template in the reaction was digested with the DpnI restriction enzyme for 30 min at 37 °C, and then 2 μL was transformed into XL-10 gold ultracompetent bacteria (Agilent Technologies) following the kit's instructions. Bacteria from the transformation reaction was plated into LB-Ampicillin agar plates at three different densities and incubated overnight at 37 °C. Next day, for each CCR5 mutation made, five bacterial colonies were picked and grown in LB media overnight. From these cultures, plasmid mini preps were performed using QIAprep spin miniprep kit (Qiagen), and the specific

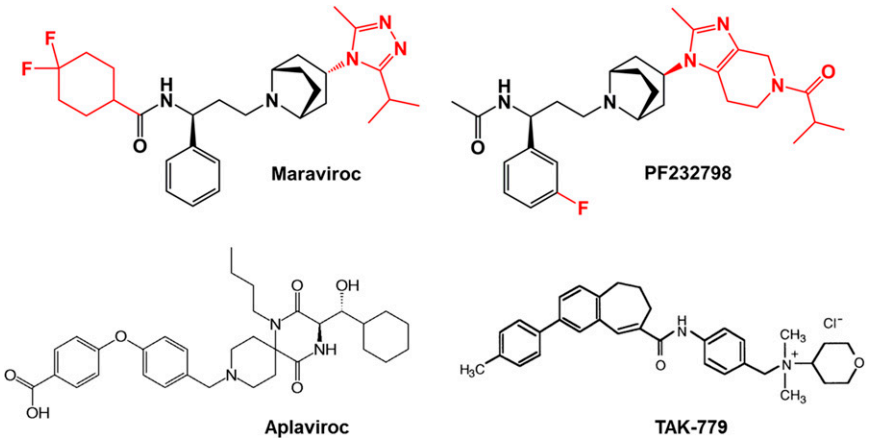
mutations within the CCR5 gene were verified by sequencing the complete ORF using specific primers.

Once bacterial clones containing sequence-verified CCR5 mutants were obtained, endotoxin-free plasmid maxiprep was done (Qiagen) to obtain specific mutant CCR5 expression vectors. These vectors were later used in Maraviroc ligand-binding assays for structure–activity relationship (SAR) studies.

Ligand-Binding Experiments. Tritium-labeled Maraviroc (99 Ci/mmol) was custom-made at Amersham-GE Healthcare Life Sciences. To prepare membranes for ligand-binding studies, CCR5 expression constructs were transiently transfected into HEK-293 cells (ATCC no. CRL-1573) using the Effectene Transfection Reagent (Qiagen) according to the manufacturer's instructions. The cells were harvested 3 d posttransfection as follows. Transfected cells grown on T-225 flasks (two flasks per expression construct), rinsed with PBS containing 5 mM EGTA (PBS/EGTA), and incubated in PBS/EGTA for 30 min at room temperature. Detached cells were pelleted at 300 × *g* for 10 min and resuspended in 8 mL of ice-cold 20 mM Hepes (pH 7.3) containing 1% (wt/vol) protease inhibitors mixture (Halt; Thermo Fisher). After 10 min on ice, the cells were homogenized by sonication (7× 20-s bursts). Homogenates were centrifuged at 300 × *g* for 10 min, and supernatants were collected and centrifuged for 2 h at 50,000 × *g*. Membranes were resuspended in 1 mL of 20 mM Hepes (pH 7.3) and stored at −70 °C. Protein content was determined using the BCA Protein Assay Reagent (Thermo Fisher).

For receptor-binding studies, 10 μg of membranes were incubated in 96-well polypropylene plates with tritium-labeled Maraviroc for 2 h at room temperature in 300 μL of binding buffer containing 50 mM Hepes (pH 7.3), 1 mM $CaCl_2$, 5 mM $MgCl_2$, 0.1% DMSO, and 0.5% BSA. Nonspecific binding was measured in the presence of 1,000-fold excess of unlabeled Maraviroc. The samples were filtered through UniFilter-96 GF/C plates preblocked with 0.3% polyethylene imine using a Unifilter-96 Harvester (PerkinElmer). The filters were washed four times with PBS, and radioactivity was measured using a TopCount instrument (PerkinElmer) in 40 μL of MicroScint 20 liquid scintillant. K_d and K_i values were obtained from saturation and competition binding experiments, respectively, and calculated using GraphPad Prism (GraphPad Softaware).

1. Pollastri G, McLysaght A (2005) Porter: A new, accurate server for protein secondary structure prediction. *Bioinformatics* 21(8):1719–1720.
2. Raghava GPS (2002) APSSP2: A combination method for protein secondary structure prediction based on neural network and example based learning. Available at www.imtech.res.in/raghava/apssp2. Accessed January 2, 2014.
3. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16(4):404–405.
4. Kam VWT, Goddard WA, III (2008) Flat-bottom strategy for improved accuracy in protein side-chain placements. *J Chem Theory Comput* 4(12):2160–2169.
5. Mayo SL, Olafson BD, Goddard WA, III (1990) Dreiding: A generic force-field for molecular simulations. *J Phys Chem* 94(26):8897–8909.
6. Abrol R, Bray JK, Goddard WA, III (2011) Bihelix: Towards de novo structure prediction of an ensemble of G-protein coupled receptor conformations. *Proteins* 80(2):505–518.
7. Bray JK, Abrol R, Goddard WA, III, Trzaskowski B, Scott CE (2014) SuperBiHelix method for predicting the pleiotropic ensemble of G-protein-coupled receptor conformations. *Proc Natl Acad Sci USA* 111(1):E72–E78.



Scheme S1. Structures of Maraviroc, PF-232798, Aplaviroc, and TAK-779.

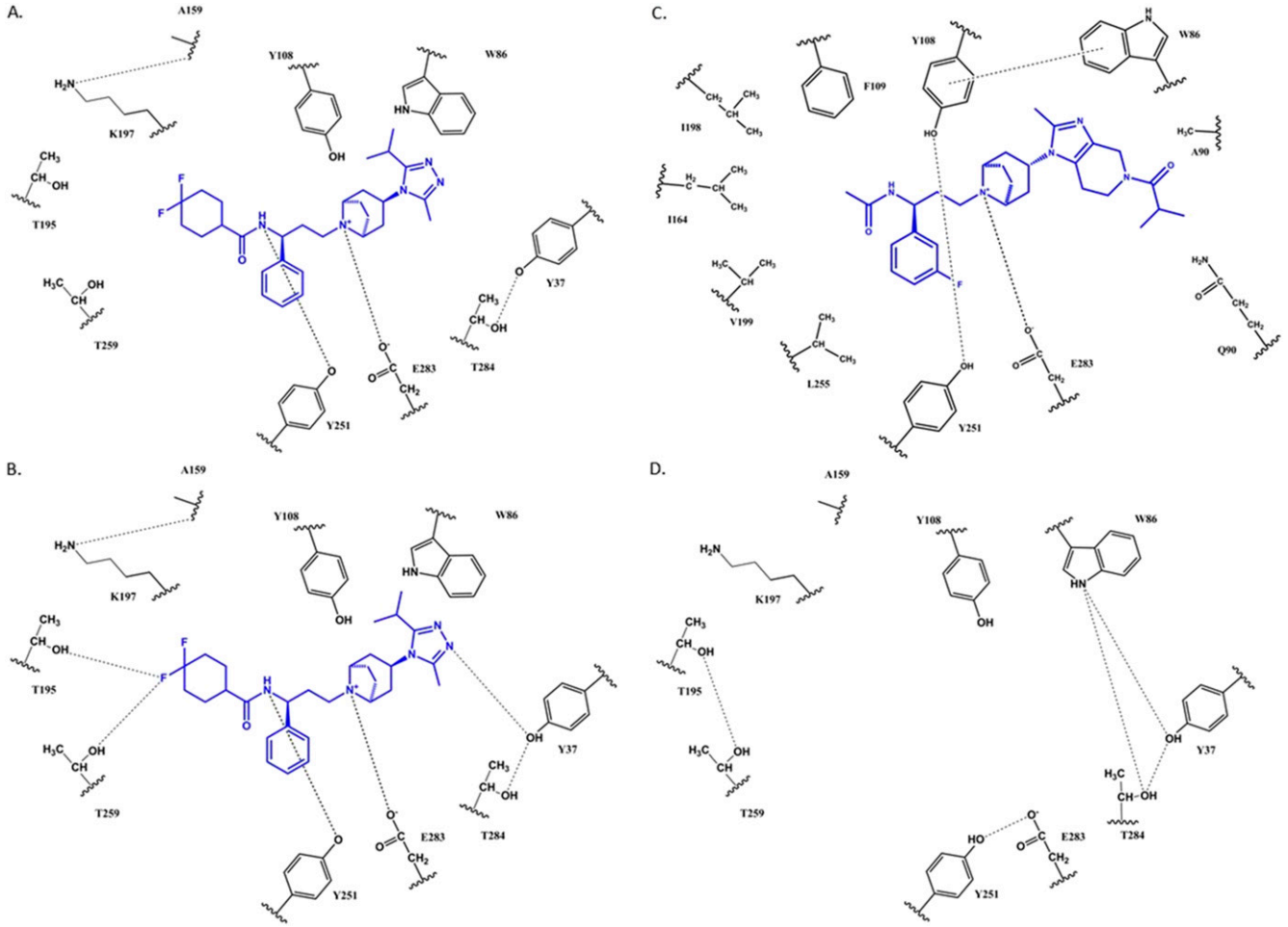


Fig. S1. Schematic representation of the CCR5 residues in the MVC binding site for (*A*) WT7 (MVC binding CCR5 conformation) with MVC, (*B*) X-ray CCR5 conformation with MVC, (*C*) WT7 (PF-232798 binding CCR5 conformation) with PF, and (*D*) WT1 (lowest energy ranked CCR5 apo WT conformation) no ligand.

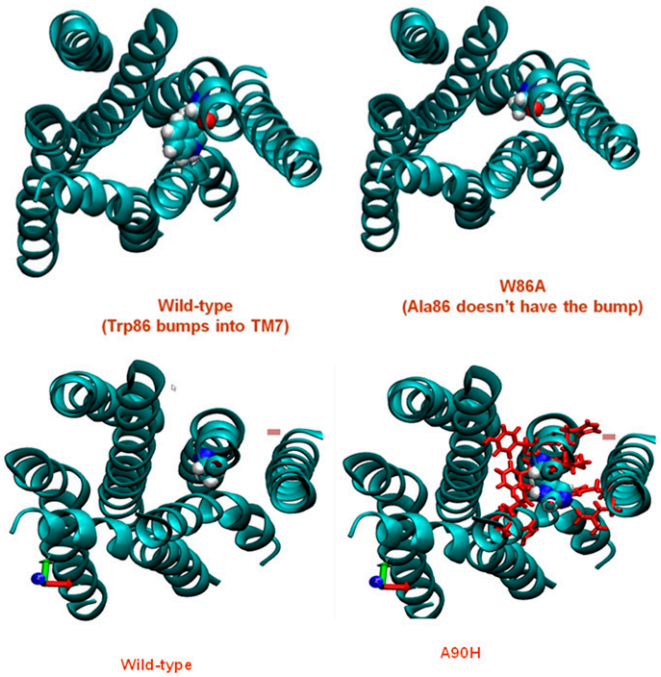


Fig. S2. Structural basis for the effect of the W86A mutation on the wt30 conformation and for the A90H mutation on the wt1 conformation.

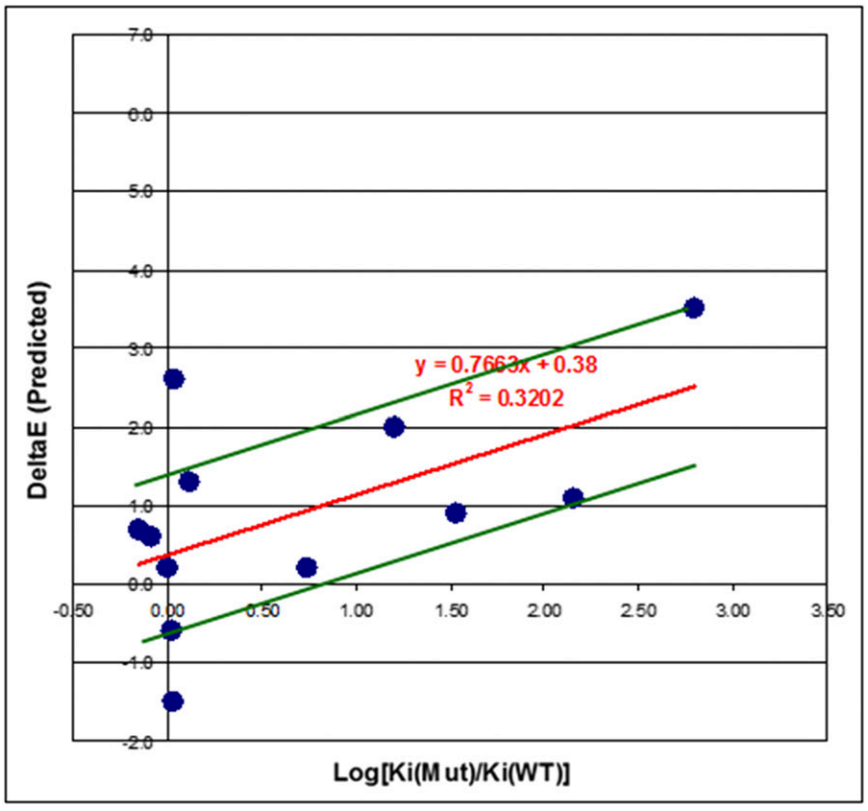


Fig. S3. Comparison of predicted effect of PF binding to mutants vs. experimental data. The ±1 kcal/mol lines are shown in green.

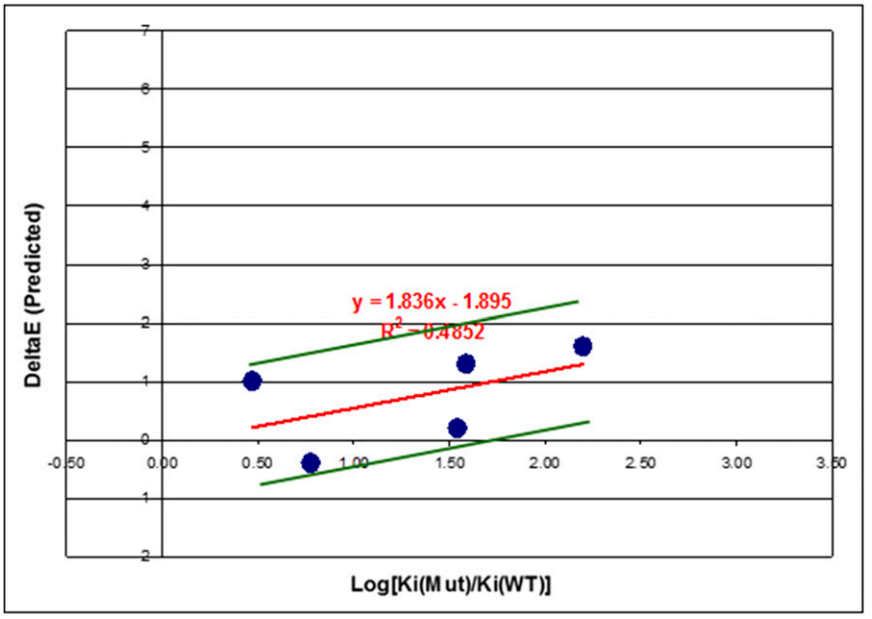


Fig. S4. Comparison of predicted effect of Aplaviroc binding to mutants vs. experimental data. The ±1 kcal/mol lines are shown in green.

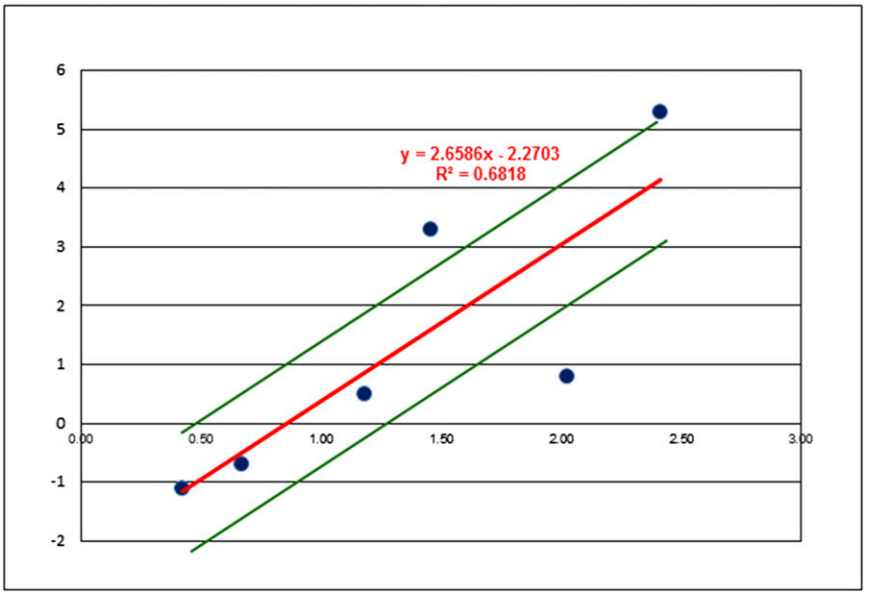


Fig. S5. Comparison of predicted effect of Tak-779 binding to mutants vs. experimental data. The ±1 kcal/mol lines are shown in green.

Table S1. The top 20 predicted conformations for human CCR5 chemokine receptor (from GEnSeMBLE)

Theta	H1	H2	H3	H4	H5	H6	H7	Phi	H1	H2	H3	H4	H5	H6	H7	Eta	H1	H2	H3	H4	H5	H6	H7	AvgRank	CCR5	Energy
Theta	−10	−10	−10	0	0	0	0	Phi	−15	0	0	−15	−15	−15	−15	Eta	0	240	0	0	15	0	0	12.8	wt1	0 ± 3
Theta	0	−10	−10	0	0	0	0	Phi	−15	0	−15	0	0	0	0	Eta	345	225	0	0	345	0	0	14.5	wt2	
Theta	0	−10	−10	0	0	0	0	Phi	−15	0	−15	−15	0	0	0	Eta	345	225	0	0	345	0	0	21.3	wt3	
Theta	−10	−10	−10	0	0	0	0	Phi	−15	15	−15	−15	0	−15	−15	Eta	345	225	0	15	345	0	0	71.3	wt4	15 ± 3
Theta	0	−10	−10	−10	0	0	0	Phi	−15	0	0	15	0	0	0	Eta	345	225	0	15	345	0	0	89.8	wt5	
Theta	−10	−10	−10	10	0	0	0	Phi	−15	−15	15	−15	−15	0	0	Eta	15	0	0	15	30	0	0	98.5	wt6	
Theta	−10	−10	0	0	10	0	0	Phi	−15	−15	15	−15	−15	−15	0	Eta	15	0	0	105	15	0	0	100.3	wt7	
Theta	−10	−10	−10	0	0	0	0	Phi	−15	0	−15	−15	0	−15	−15	Eta	0	225	0	15	345	0	0	108.5	wt8	
Theta	0	−10	−10	−10	0	0	0	Phi	−15	0	0	15	−15	0	0	Eta	345	225	0	15	15	0	0	109.8	wt9	18 ± 4
Theta	−10	−10	−10	10	0	0	0	Phi	−15	−15	0	0	−15	15	0	Eta	15	15	0	15	15	0	0	111.0	wt10	
Theta	0	−10	−10	0	0	0	0	Phi	−15	0	−15	−15	0	−15	0	Eta	330	225	0	15	345	0	0	122.0	wt11	
Theta	−10	−10	−10	10	0	0	0	Phi	−15	−15	0	−15	0	0	0	Eta	15	0	0	105	345	0	0	126.3	wt12	
Theta	0	−10	−10	0	0	0	0	Phi	−15	0	−15	−15	0	−15	−15	Eta	345	225	0	0	345	0	0	143.5	wt13	
Theta	0	−10	−10	−10	0	0	0	Phi	−15	0	0	0	−15	0	0	Eta	345	225	0	15	0	0	0	145.5	wt14	20 ± 4
Theta	0	0	−10	10	0	0	0	Phi	−15	−15	−15	−15	0	0	0	Eta	0	0	0	120	345	0	90	145.8	wt15	
Theta	−10	0	−10	10	0	0	0	Phi	−15	−15	−15	−15	0	0	15	Eta	0	15	0	105	345	0	0	146.5	wt16	
Theta	−10	−10	−10	10	0	0	0	Phi	−15	−15	0	−15	0	15	0	Eta	15	15	0	120	75	0	0	147.0	wt17	
Theta	−10	−10	−10	−10	0	0	0	Phi	−15	−15	0	0	0	15	0	Eta	15	15	0	120	225	0	0		wt18	23 ± 2
Theta	−10	−10	−10	−10	0	0	0	Phi	−15	−15	0	15	−15	−15	0	Eta	15	0	0	120	15	0	0		wt19	
Theta	−10	−10	−10	−10	0	0	0	Phi	−15	−15	0	−15	0	15	0	Eta	15	15	0	120	225	0	0		wt20	
																									wt50	27 ± 3

For docking to ligands, the conformations in the eight shaded rows were selected based on diversity and energy. After wt20, wt50 conformation's energy is listed to show the flattening of the conformational energy landscape.

Table S2. Reordering of CCR5 conformations for W86A, A90H, and T105A mutants

W86A	Avg rank	A90H	Avg rank	T105A	Avg rank
wt1	8.0	wt3	3.0	wt1	4.0
wt3	11.8	wt2	6.0	wt3	7.8
wt2	15.0	wt4	15.3	wt2	10.0
wt30	16.0	wt5	16.3	wt11	28.3
wt6	20.8	wt13	18.5	wt9	28.8
wt17	26.0	wt10	21.0	wt12	29.3
wt7	28.3	wt9	22.5	wt18	30.5
wt8	29.3	wt27	22.8	wt4	30.5
wt4	30.0	wt14	23.8	wt27	30.8
wt18	30.3	wt32	24.8	wt6	31.0
wt11	30.3	wt64	25.0	wt7	31.3
wt48	31.0	wt15	25.0	wt20	31.5
wt20	31.3	wt16	25.8	wt8	31.8
wt12	31.3	wt26	27.0	wt14	31.8
wt19	34.8	wt28	27.5	wt13	34.0
wt14	34.8	wt31	28.5	wt15	36.0
wt9	35.8	wt17	29.3	wt5	36.0
wt27	36.0	wt33	30.3	wt64	37.3
wt34	37.0	wt11	32.0	wt32	38.0
wt13	38.0	wt74	32.5	wt30	42.5