SUPPLEMENTARY INFORMATION FOR

Genetically-encoded Photocrosslinkers Map the Binding Site of an Allosteric Drug on a G Protein-coupled Receptor

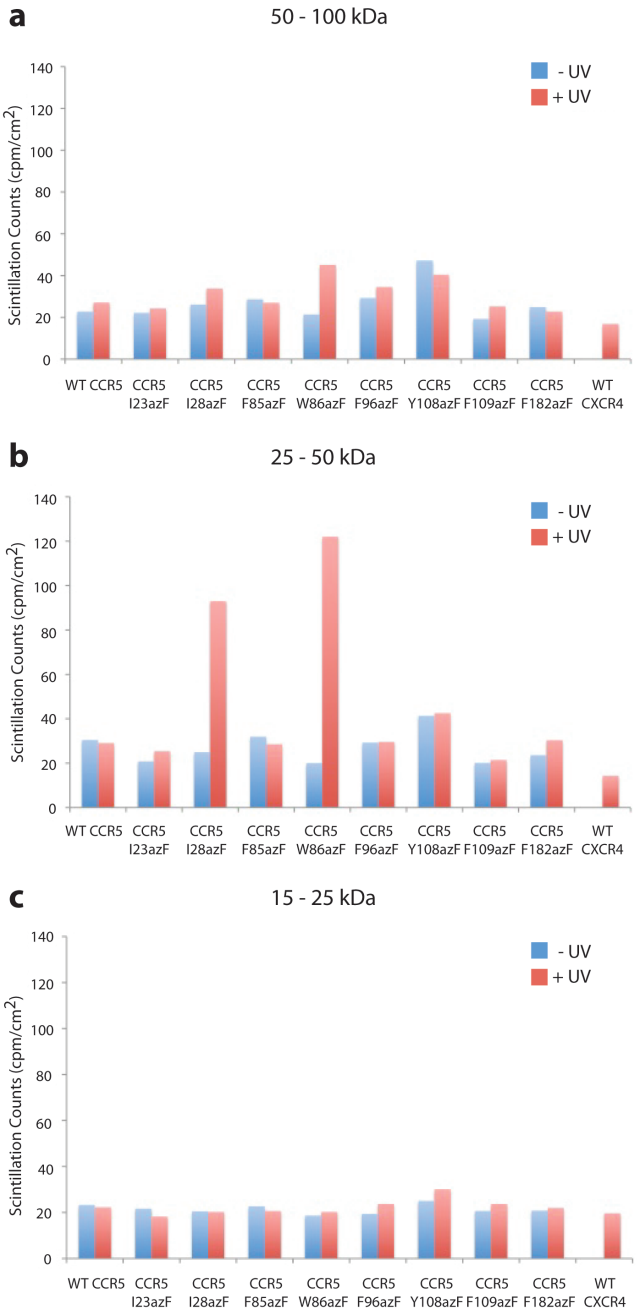
Amy Grunbeck[†], Thomas Huber[†], Ravinder Abrol[‡], Bartosz Trzaskowski[‡], William A. Goddard, III[‡], Thomas P. Sakmar[*,†]

[†]Laboratory of Molecular Biology & Biochemistry, The Rockefeller University, 1230 York Ave., New York, NY 10065. www.sakmarlab.org

[‡]Materials and Process Simulation Center (MC 139-74), California Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125, United States.

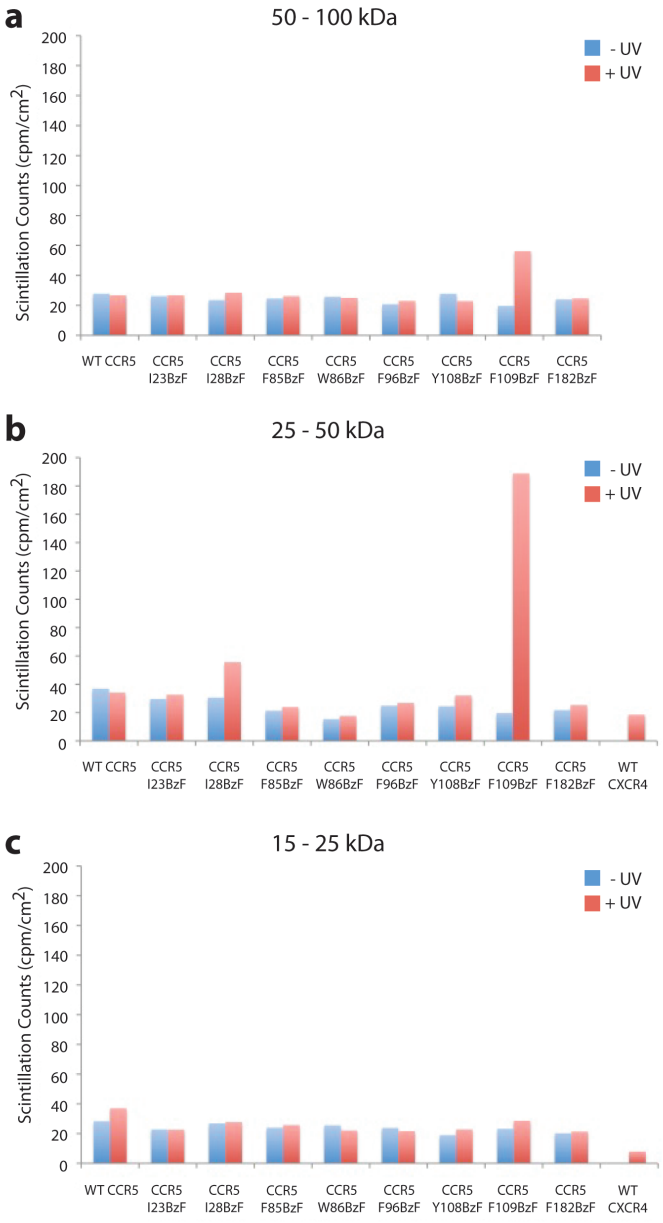
Contents:

I. Supplementary Figures



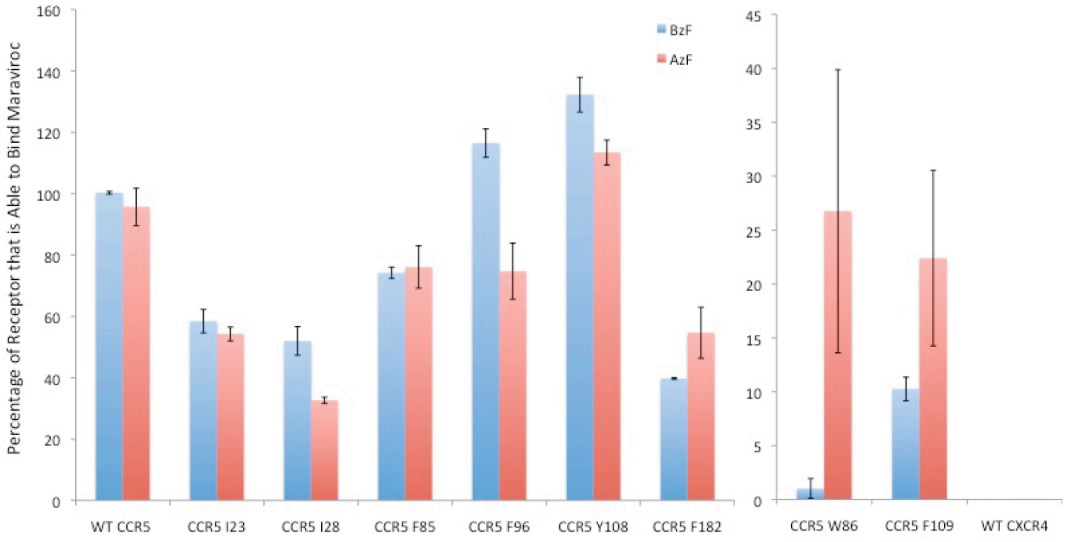
Supplementary Figure 1 Photocrosslinking results with CCR5 azF mutants. This data set is from the same experiment that is shown in Figure 2a, except here we include the control sample for each mutant that was not exposed to UV light. The three bar graphs shown here display the amount of tritium detected in the 50 – 100 kDa segment (**a**), 25 – 50 kDa segment (**b**), and 15 – 25 kDa segment (**c**). This data shows that the elevated levels of tritium detected in the 25 – 50 kDa segment for the I28 and W86 mutants is UV-specific. This supports the conclusion that these increases in tritium levels at the molecular weight of the

receptor are a result of a covalent crosslink between the azF at positions I28 and W86 in CCR5 and [^{3}H]-maraviroc.

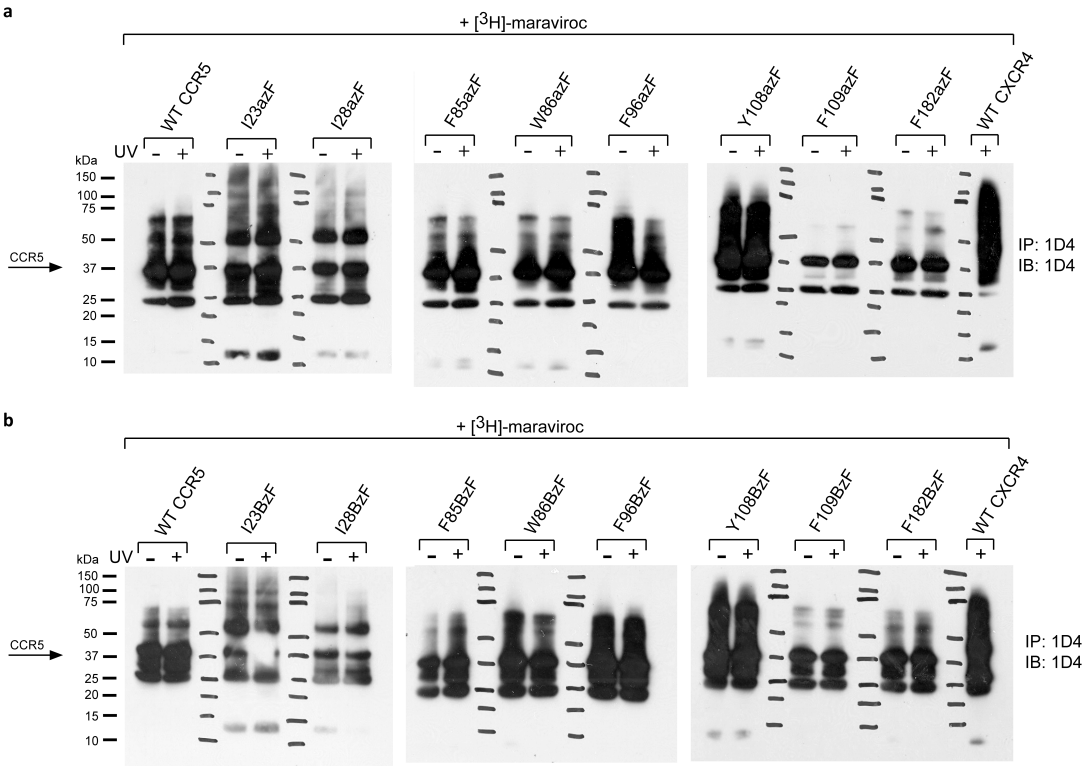


Supplementary Figure 2 Photocrosslinking results with CCR5 BzF mutants. This data set is from the same experiment that is shown in Figure 2b, except here we include the control sample for each mutant that was not exposed to UV light. The three bar graphs shown here display the amount of tritium detected in the 50 – 100 kDa segment (**a**), 25 – 50 kDa segment (**b**), and 15 – 25 kDa segment (**c**). This data shows that the elevated levels of tritium detected in the 25 – 50 kDa segment for the I28 and F109 mutants is UV-specific. This supports the conclusion that these increases in tritium levels at the molecular weight of the

receptor are a result of a covalent crosslink between the BzF at positions I28 and F109 in CCR5 and [^{3}H]-maraviroc.



Supplementary Figure 3 CCR5 UAA mutants retain binding to maraviroc. Each of the CCR5 UAA mutants was measured for their ability to bind maraviroc under the conditions of the crosslinking experiments. In brief, after the HEK293T cells expressing the CCR5 UAA mutants were incubated with [^{3}H]-maraviroc, exposed to UV light and then solubilized in detergent, the receptors were then immunopurified from the lysate using sepharose beads conjugated to the 1D4 mAb. The beads were then washed, samples were eluted using SDS, and the amount of tritium in the elution was quantified. The [^{3}H]-maraviroc detected in the elution was determined to be a result of being specifically bound to the purified receptor. This was supported by no tritium was detected in the sample with the homologous receptor, CXCR4, which is unable to bind maraviroc. Therefore the amount of tritium in the eluted samples was quantified for each CCR5 UAA mutant to determine the amount of [^{3}H]-maraviroc that was specifically bound. The values were then normalized to each other by setting the amount of tritium that was bound to WT CCR5 as an equivalent to 100% binding. The error bars are a result of averaging the values obtained from the sample that was exposed to UV light with the sample that was not. The differences in binding between the mutants are a result of varied expression levels of each of the mutants and differences in binding affinity for maraviroc. This graph shows that all of the mutants retain significant binding to maraviroc except for the CCR5 W86BzF mutant.



Supplementary Figure 4 Western Blots showing the expression level of each of the CCR5 UAA mutants. The Western Blots shown above were immunoblotted with the 1D4 mAb, which recognizes an epitope at the C-terminus of the receptor. The band corresponding to CCR5 is indicated with the arrow at 37 kDa. The other prominent band at 25 kDa is the light chain of the 1D4 antibody that was used to immunopurify these samples. The exposure of each of these blots is not normalized to each other. The samples shown in (**a**) are from the azF experiments and (**b**) shows the samples from the BzF experiments. These membranes were then cut at specific molecular weights and counted to determine the amount of tritium in each segment as previously described.

II. Methods

Structure prediction of the CCR5-maraviroc complex.

1. GEnSeMBLE(*1*) procedure for generating an ensemble of GPCR conformations:
   A. Predict transmembrane (TM) helix regions of the receptor:
      Use PredicTM to predict TM regions based on the octanol hydrophobicity scale(*2*) and extend those regions based on helix predictions from the secondary structure servers.
   B. Generate template for protein bundle:
      i. Mutate TM helix bundles of multiple crystal templates [bovine rhodopsin (PDB ID: 1u19), human β2 adrenergic receptor (2rh1), human adenosine A2A receptor (3eml), and turkey β1 adrenergic

receptor (2vt4)] to CCR5 sequence to generate 4 starting structures for the following conformational sampling.

C. Sample completely the ensemble of seven-helix bundle conformations:
   i. BiHelix:(*3*) Sample helix rotation angle (η) for all pairs of interacting helices. Use SCREAM (SideChain Rotamer Excitation Analysis Method)(*4*) to optimize sidechain interactions for each sampled conformation and minimize the sidechains for 10 steps. Combine all BiHelix energies as described here(*3*) and order by energy.
   ii. CombiHelix: Build the top 2000 conformations from previous step, and use SCREAM with 10-step minimization to optimize sidechain interactions.
   iii. Comparing the interhelical energies for these conformations for each of the 4 starting structures, showed that the β2 crystal template resulted in the lowest energy helix packing conformations for CCR5.
   iv. SuperBiHelix: Take the 16 diverse lowest energy conformations from previous step to sample, for each conformation, helix tilt angle (θ), sweep angle (φ), and rotation angle (η) for all pairs of interacting helices, and use SCREAM with 10-step minimization to optimize sidechain interactions. Combine all SuperBiHelix energies as described here(*3*) and order by energy.
   v. SuperCombiHelix: Take the top 2000 conformations from previous step, and use SCREAM with 10-step minimization to optimize sidechain interactions.
   vi. Choose the lowest energy structures from SuperCombiHelix according to the average of the charged total energy rank, neutralized total energy rank, charged interhelical energy rank, and neutralized interhelical energy rank (Average Energy Rank).

2. Docking of Maraviroc to 8 diverse lowest energy conformations of CCR5
   A. Ligand preparation
      i. The ligand maraviroc was built and minimized in Maestro/Macromodel program. Quantum mechanics calculations (B3LYP/6-311G** using Jaguar 7 software, Jaguar, version 7.6, Schrödinger, LLC, New York, NY, 2009.) were performed to obtain partial charges as well as calculate the pKa values of selected Nitrogen atoms.
      ii. Maestro/Macromodel was used to perform a long conformational search, which after additional clustering yielded ~25 diverse Maraviroc conformations.
   B. Ligand docking
      i. In our docking approach we used DOCK 6(*5*) to generate a large number (~50000) of ligand poses in the potential binding regions. The resulting ligand poses were clustered using 1.2 Å diversity until the % of new families (containing only poses from the current generation) became less than 5% to ensure completeness.
      ii. This step was followed by the energy scoring of family heads using polar, phobic and total energies. Top 100 poses from each energy scoring and additionally top 10 poses from each ligand conformation

were chosen for further optimization, which included dealanization of protein, neutralization of the protein-ligand complex and complex minimization. All optimization steps were performed independently for each of the eight CCR5 models using the DREIDING force field(*6*). The lowest total energy complex was chosen as the final CCR5-maraviroc complex model. This corresponded to the 7th lowest energy predicted apo CCR5 conformation.

References

1. Abrol, R., Kim, S. K., Bray, J. K., Griffith, A. R., and Goddard, W. A., 3rd. (2011) Characterizing and predicting the functional and conformational diversity of seven-transmembrane proteins, *Methods*.
2. Wimley, W. C., Creamer, T. P., and White, S. H. (1996) Experimentally determined hydrophobicity scales for membrane proteins, *Biophys. J. 70*, TUAM1-TUAM1.
3. Abrol, R., Bray, J. K., and Goddard, W. A., 3rd. (2011) BiHelix: Towards de novo Structure Prediction of an Ensemble of G protein-Coupled Receptor Conformations, *Proteins: Structure, Function, and Bioinformatics In press*.
4. Kam, V. W. T., and Goddard, W. A. (2008) Flat-Bottom Strategy for Improved Accuracy in Protein Side-Chain Placements, *J. Chem. Theory Comput. 4*, 2160-2169.
5. Moustakas, D. T., Lang, P. T., Pegg, S., Pettersen, E., Kuntz, I. D., Brooijmans, N., and Rizzo, R. C. (2006) Development and validation of a modular, extensible docking program: DOCK 5, *J. Comput.-Aided Mol. Des. 20*, 601-619.
6. Mayo, S. L., Olafson, B. D., and Goddard, W. A. (1990) DREIDING - A GENERIC FORCE-FIELD FOR MOLECULAR SIMULATIONS, *J. Phys. Chem. 94*, 8897-8909.