Supporting Information

Screening for GPCR Ligands using Surface Plasmon Resonance

Iva Navratilova, Jérémy Besnard and Andrew L. Hopkins

Division of Biological Chemistry and Drug Discovery, College of Life Sciences, Dow Street, University of Dundee, Dundee, DD1 5EH, UK

TABLE OF CONTENTS

Materials and Methods

Databases: ChEMBL_02 was used for this study[1]; ChEMBL is a database of bioactive molecules containing around 500,000 compounds with more than 2,400,000 endpoints published in medicinal literature over the last 30 years. The data was processed for this work by keeping only compounds with an activity which standard unit is nM and is inside the group of (IC_{50}/K_i/K_d/EC_{50}) with a protein as target and some custom filter were then applied. Compounds were also standardized, and a unique tautomer form was kept, using the Pipeline Pilot (Accelrys, San Diego, USA) components.

Molecules with a molecular weight less than 60Da or superior to 650Da were discarded. Finally compounds belonging only to a target with less than 10 active compounds (activity < 10 μM) were reassigned to "no target". The final set is composed of 206,748 compounds with 1717 biological targets (one compound can have multiple targets assigned). The University of Dundee Drug Discovery Unit's compound set is composed of 90,000 internal compounds, mainly from commercial suppliers. The compounds are processed similarly as the ChEMBL data.

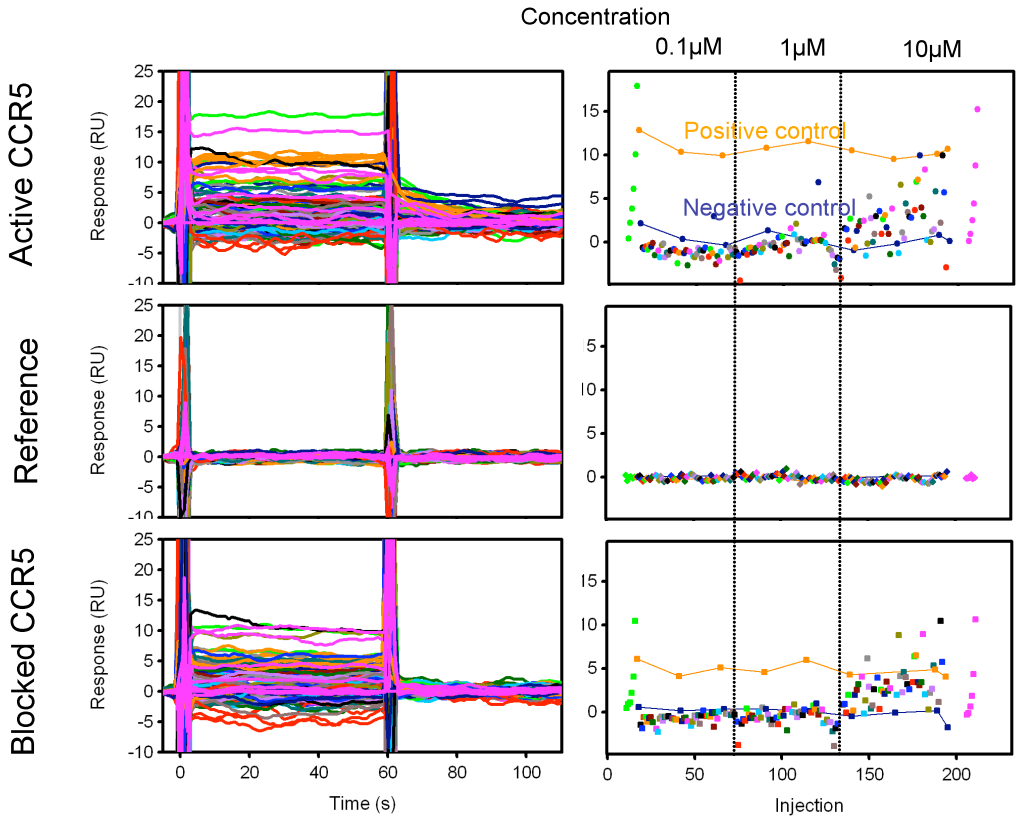
Chemical Structure Descriptors: The descriptors used are a mix of molecular properties and structural information. The properties selected are the default setting for a model: molecular weight, AlogP, Number of hydrogen bond donors, acceptors, number of rotatable bonds and Fractional TPSA (TPSA divided by number of atoms). For the structural information the Extended Connectivity Fingerprint (ECFP) were selected [2], this type of descriptors is used in other studies with very good results even with noisy data [3] [4] [5]. A neighbourhood of size 6 was selected to match the parameters of Nidhi *et al.* [6].

Bayesian Model: The Bayesian activity model is built using previously describe methods [6-9] using Pipeline Pilot (Accelrys) which automatically creates a Laplacian-modified Bayesian model for each of the bioactive targets. For each target a compound is or active (activity < 10 μM) or inactive (all the remaining compounds, as by default a compound is considered inactive against a target). Thus if a compound is active against 10 targets, it will be in 10 categories as an active compounds, and in 1707 an inactive compound. Once the model is built it can be used to calculate a score for a particular target, and a high score provides more confidence of binding. For CCR5, ChEMBL ID is CHEMBL274, the number of active compounds is 1,166 and the number of inactive is 205,582. The DDU set is processed and each compound gets a score for CCR5, and ranked from best to worst.

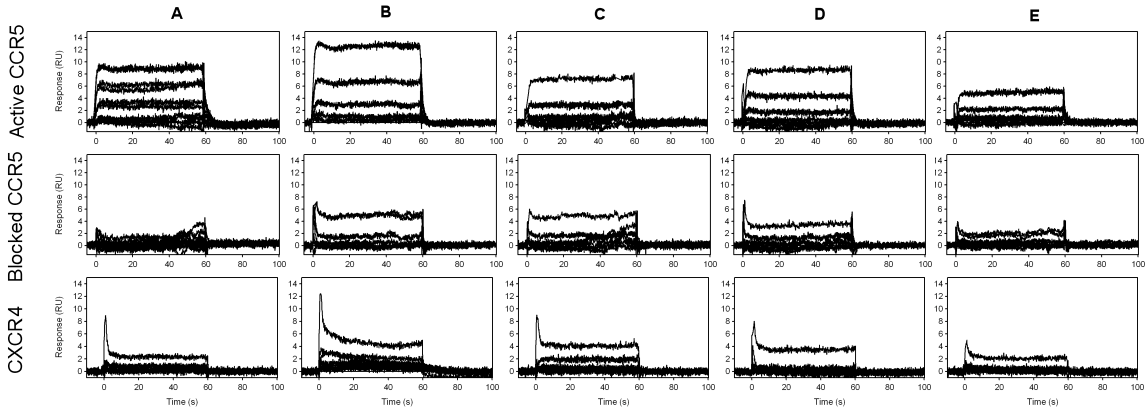
Surface Plasmon Resonance Assay: 1D4 mAb was immobilised using standard amine coupling on all spots. The monoclonal antibody 1D4 was immobilized on a CM4 sensor chip using standard amine-coupling chemistry.Hepes-buferred saline (HBS-N) buffer (10mM Hepes, 0.15M NaCl, pH 7.4) was used as the running buffer. The carboxymethyl dextran surface was activated with a 7-min injection of a 1:1 ratio of 0.4M EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride) and 0.1M NHS (*N*-hydroxysuccinimide). The antibody was coupled to the surface with a 7-min injection of 1D4 diluted in 10mM sodium acetate (pH 5.0). Remaining activated groups were blocked with a 7-min injection of 1M ethanolamine (pH 8.5). CCR5 and CXCR4 was obtained from National Cell Culture Centre (USA), solubilised as described previously [10-12] and captured in running buffer composed of 50 mM Hepes, 150 mM NaCl, 0.5% DMSO, 0.1% Dodecyl Maltoside, 0.1% CHAPS, 0.02% Cholesteryl Hemi Succinate and 5% glycerol and captured at flow rate 10 ul/min over spot 1 and 5. Approximately 4,000 - 5,000 RU of receptor was immobilised

on each detection spot. Spot 5 was blocked by injection of 5 μM maraviroc. Maraviroc injections were repeated during the run to maintain receptor blocked. Two hundred selected compounds selected for the screen were injected at three concentrations 0.1 μM, 1 μM and 10 μM at flow rate 30 ml/min. Association was measured for 1 minute and dissociation for 2 minutes. The control compound UK 107,543 was injected at the beginning and end of the screen in concentration series 0.0045 μM – 10 μM. Both positive (UK-107,543) and negative (sulpiride) controls were injected during the run at 5 mM concentration. Confirmation of selected hits was run at 5 concentrations in three-fold dilution series on Biacore T100 at concentrations 0.3 μM – 25 μM. Association and dissociation was measured for 1 and 2 minutes respectively at flow rate 30 ul/min. All data were processed and analysed using Scrubber 2 (BioLogic software) and responses normalised for MW of compounds.

Supplementary Figure 1: Overlay of sensorgrams for one flow cell and 50 compounds binding to active CCR5, reference spot and blocked CCR5. Report points for the injection of each compound and three concentrations 0.1 µM, 1 µM and 10 µM are showed in the right column. Green and pink report points at the beginning and end of the screen represent injection of positive control at concentration series 0.041 µM – 10 µM. Orange and blue report points represent repeated injections of positive and negative controls during the screen.



Supplementary Figure 2: Sensorgrams for compounds A-E identified as hits binding to active and blocked CCR5 and CXCR4. Each compound is injected in duplicates at 3-fold concentration series 0.3 µM – 25 µM.



References

1. ChEMBL http://www.ebi.ac.uk/chembldb/, 2010.
2. Rogers, D.; Hahn, M., Extended-connectivity fingerprints. . *J. Chem. Info. Model.* **2010**, 50, 742-754.
3. Hert, J.; Wilet, P.; Wilton, D.; Acklin, P.; K., A.; E., J.; Schuffenhauer, A., Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures. . *Org. Biomol. Chem.* **2004** 2, (22), 3256-3266.
4. Klon, A. E.; Glick, M.; Thoma, M.; Acklin, P.; Davies, J. W., Finding More Needles in the Haystack: A Simple and Efficient Method for Improving High-Throughput Docking Results. *J. Med. Chem.* **2004**, 47, (11), 2743-2749.
5. Glick, M.; Jenkins, J.L.; Nettles, J. H.; Hitchings, H.; Davies, J. W., Enrichment of high-throughput screening data with increasing levels of noise using support vector machines, recursive partitioning, and laplacian-modified naive bayesian classifiers. Journal of chemical information and modeling. . *J. Chem. Inf. Model.* **2006**, 46, (1), 193-200.
6. Nidhi; Glick, M.; Davies, J. W.; Jenkins, J. L., Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J. Chem. Inf. Model.* **2006**, 46, (3), 1124-1133.
7. Xia, X.; Maliski, E. G.; Gallant, P.; Rogers, D., Classification of kinase inhibitors using a Bayesian model. *J. Med. Chem.* **2004**, 47, (18), 4463-70.
8. Paolini, G. V.; Shapland, R. H. B.; van Hoorn, W. P.; Mason, J. S.; Hopkins, A. L., Global Mapping of Pharmacological Space. *Nat. Biotechnol.* **2006**, 24, (7), 806-815.

9. Bender, A., Bayesian methods in virtual screening and chemical biology. . In *Methods in molecular biology*, Clifton, N. J., Ed. **2011**; Vol. 672, pp 175-196.
10. Navratilova, I.; Sodroski, J.; Myszka, D. G., Solubilization, stabilization, and purification of chemokine receptors using biosensor technology. *Anal. Biochem.* **2005**, 339, (2), 271-281.
11. Navratilova, I.; Pancera, M.; Wyatt, R. T.; Myszka, D. G., A biosensor-based approach toward purification and crystallization of G protein-coupled receptors. *Anal. Biochem.* **2006**, 353, (2), 278-283.
12. Navratilova, I.; Dioszegi, M.; Myszka, D. G., Analyzing ligand and small molecule binding activity of solubilized GPCRs using biosensor technology. *Anal. Biochem.* **2006**, 355, (1), 132-139.