Supplementary Information

Large Scale Prediction and Testing of Drug Activity on Side-Effect Targets

Eugen Lounkine, Michael J. Keiser, Steven Whitebread, Dmitri Mikhailov, Jacques Hamon, Jeremy Jenkins, Paul Lavan, Eckhard Weber, Allison Doak, Serge Côté, Brian K. Shoichet & Laszlo Urban

Supplementary Methods

In vitro assay panel

GPCR and transporter radioligand binding assays

Binding assays were performed according to standard procedures published in the literature.[1–7] Both filtration and scintillation proximity assay (SPA) technologies were used.

Preparation of Membranes from rat brain cortex

Rat brain membranes were purchased from Analytical Biological Services (ABS), Delaware, USA, and were prepared according to the following protocol. The animals were decapitated; the brains removed, the cerebral cortices dissected and homogenized in 10 volumes of ice-cold 0.32 M sucrose, containing $MgCl_2$ (1 mM) and K_2HPO_4 (1 mM), with a Teflon™/glass homogenizer. The membranes were centrifuged at 1000 x g, the pellet discarded and the centrifugation repeated. The supernatants were pooled and centrifuged at 18,000 x g for 15 min. The pellet was osmotically shocked in 10 vol. of H_2O and kept on ice for 30 min. The suspension was centrifuged at 39,000 x g, resuspended in Krebs-Henseleit buffer pH 7.4 containing 20 mM Tris, and kept for 2 days at -20°C. The membranes were thawed at 20-23°C, washed three times with Krebs-Henseleit buffer by centrifugation at 18,000 x g for 15 min, left overnight at 4°C and washed again three times. The final pellet was resuspended with a Teflon™/glass homogenizer in 20 mL of the same buffer. Aliquots were frozen and stored in liquid nitrogen. Just before use, membranes were thawed quickly in a water bath at 37°C and washed by centrifugation at 18,000 x g for 15 min with the same buffer three times.

Membranes from cells expressing recombinant or native receptors

Cell membranes were obtained commercially from PerkinElmer (Boston, USA; formerly BioSignal, Receptor Biology, Euroscreen), Millipore (Billerica, Massachusetts, USA) and Caliper (Hopkinton, Massachusetts, USA). Some membrane preparations were prepared at Novartis according to the following method.

Cells stably expressing the recombinant or native receptor were grown to confluency in 500 cm^2 dishes, washed in 100 mL phosphate buffered saline and harvested using a rubber scraper and 10 mL ice-cold PBS (without Mg^{2+} or Ca^{2+}). The cells were then homogenised in 30 mL capacity centrifuge tubes using a Polytron homogeniser (3 bursts of 10 seconds at setting 8) and centrifuged at 32,000 x g for 25 min at 4°C. The supernatant was discarded. The pellet was washed by resuspending in 10 mL ice-cold hypotonic buffer (Tris-HCl 20 mM; EDTA 5 mM; pH 7.7 at 4°C) using a Teflon™/glass homogeniser (1,100 rpm) and

recentrifuged. The final pellet was resuspended in incubation buffer using a Teflon™/glass homogeniser (1,100 rpm) (12 mL for 10 dishes, giving a protein concentration of approx. 5 mg/mL). Protein concentration was measured by the BCA method (Pierce, Rockford, IL), or a Coomassie Blue method (Bio-Rad, Hercules, CA), using bovine serum albumin as standard. The crude membrane preparation was aliquoted (1 mL) in cryo-tubes, flash-frozen in liquid nitrogen and stored at -80°C.

Typical filtration assay protocol

Incubations were performed in 96 well plates (Costar) in a total volume of 200 µL, consisting of 100 µL of radioligand prepared in assay buffer, 40 µL of various concentrations of test compound or buffer, 60 µL of membrane preparation (cells or rat brain membranes). Non-specific binding (NSB) was determined in the presence of an appropriate ligand specific for the receptor under study. The plates were incubated to steady state for the time specified. Reactions were terminated by flash filtration and inverse transfer to 96 well filter plates (Packard 96 well cell harvester, filter plates GF/C). The plates were dried for 30 min at 50°C in an oven and sealed at the bottom using an adhesive sheet. Subsequently, 40-50 µL of scintillation cocktail (Microscint-20, PerkinElmer) were added, sealed on top with an adhesive sheet (TopSeal, PerkinElmer) and the radioactivity was counted in a 96-well plate counter (TopCount, PerkinElmer).

Typical SPA assay protocol

Wheatgerm agglutinin SPA beads, lyophilized (RPNQ0001) were purchased from GE Healthcare (Piscataway, NJ, USA). One vial (500 mg) was reconstituted using 5 mL of distilled water to give a final concentration of 100 mg/mL. Reconstituted SPA beads were stored at 4°C and not frozen. 384 well white, clear bottom plates (CAT # 3706) were purchased from Corning (Lowell, MA, USA).

The incubations were performed in a final volume of 50 µL per well in a 384 well white, clear bottom polystyrene plate.

The components of the incubation were added to each well as follows:

10 µL test compound in 2.5 % DMSO in distilled water

20 µL radioligand

20 µL SPA bead / membrane mixture

The final concentration of beads ranged from 2 to 7 mg/mL and that of membranes from 0.25 - 800 µg/mL, depending on the assay. The plates were then sealed and allowed to sit at

room temperature for target-specific incubation times ranging from . Plates were counted in Perkin Elmer Microbeta readers for 90 seconds per well.

GPCR functional assays

Typical cAMP quantification assay for agonism

The assay was performed in white 384-well flat bottom assay plate (Nunc #262360) using the Homogenous Time Resolved Fluorescence Technology (HTRF) following standard procedure as described previously.[8] HTRF cAMP was purchased from Cisbio International (Cisbio #62AM4PEC - Bagnols-sur-Cèze, France). Aliquots of cryopreserved cells overexpressing the target of interest were used.

On the day of the experiment, the cells were thawed rapidly by placing the vials at room temperature. The cells were diluted in HBSS/HEPES buffer and 5000 cells per well were distributed in a volume of 12 µL. Cells were kept 10 minutes at room temperature before the compound addition. 4 µL IBMX (1 mM final concentration) diluted in HBSS/HEPES buffer was then added followed by 4 µL of compounds diluted in 0.3% water/DMSO solution. Low controls (unstimulated cells) were determined with the addition of 4 µL of assay buffer. High controls were determined with the addition of 4 µL of a known agonist at high concentration and a reference agonist compound was also included on each plate. HTRF reagents were prepared by diluting stock solutions of anti cAMP cryptate and cAMP D2 1:20 in lysis buffer supplied with the kit. After 30 min of incubation at room temperature, 10 µL of cAMP-D2 and 10 µL of anti cAMP cryptate were added to the assay plates. After one hour of incubation time at room temperature, the plates were read on the PheraStar Multimode reader (Excitation wavelength: 337 nm, Emission wavelengths: 620 and 665 nm). Results were calculated from the 665 nm / 620 nm ratio following standard protocols.[8]

Typical calcium flux assay using FLIPR

The day before the experiment, batches of cryopreserved cells (10 million of cells per mL) were thawed in a water bath at 37°C, centrifuged for 5 minutes at 1000 rpm and the cell pellet was resuspended in cell culture medium. Cells were seeded into black 384-well plates with clear bottom at a density of 10000 cells per well in 50 µL of cell culture medium. Plates were incubated overnight at 37°C in 5% CO_2. The day of the experiment, medium was removed by aspirating the plate with a cell washer (Witec-ELX405) and cells were loaded for 60 min at 37°C, 5% CO_2 with 40 µL of loading buffer (HBSS 1X, 20 mM HEPES, 2.5 mM Probenecid, Fluo4-AM 1.6 µM). Plates were washed twice with 100 µL of assay buffer (HBSS 1X, 20 mM HEPES, 2.5 mM Probenecid). After the final wash, 20 µL were left on the cells and the plates were incubated for at least 10 minutes before performing the experiment.

For Gi or Gs coupled GPCRs, cells were first sensitized by a prestimulation with a Gq agonist, and later exposed to various ligands. Using this experimental paradigm, activation of most Gs and Gi-coupled receptors becomes possible using the FLIPR technology. 20 µL ATP at 20 µM diluted in assay buffer was added by the FLIPR 384 head for cell priming. Cells were incubated at room temperature for 30 minutes. Compounds were then characterized in the agonist and/or antagonist mode. For Gq coupled GPCRs, this step is not necessary and cells can be directly stimulated by compounds.

Agonist mode

Cells were stimulated by adding compounds (5x concentrated), using the FLIPR pipette. A measure of the signal baseline was first recorded every second for 2 minutes before the injection of the compounds. Calcium measurements were performed by exciting the cells with the argon ion laser at 488 nM at 0.6 W laser power and recording the fluorescence signal with a CCD camera (opening of 0.4 sec) for 2 minutes. Low controls (unstimulated cells) were determined with assay buffer and high controls with a known agonist at high concentration (EC_{100}).

Antagonist mode

After performing the agonist experiment, plates were incubated at room temperature during 15 minutes and a known agonist, at EC_{80} concentration, was added in each well. Low controls were determined with a known antagonist at high concentration (EC_{100}), high controls with the addition of buffer.

The normalized fluorescence (dF/F) was used to evaluate compound activity as described in the literature.[9,10]

Nuclear receptor binding assays using a fluorescence polarization (FP) method

Assays were performed in 384 well black flat-bottom plates (Corning 3654) with a total volume of 50 µL: 10 µL of sample (test compound, buffer or reference compound), 20 µL of fluorescent ligand and 20 µL of receptor. Ligand, receptor and their corresponding buffers were purchased from Invitrogen (Carlsbad, CA, USA) as complete kits (ER alpha #P3029, ER beta #P3032, PR #P2962, GR #P2893 and AR #PV4293). The kits were run according to the instructions provided. 20 µL of ligand and 20 µL of receptor were added to plates already containing 10 µL of test compound using a Tecan EVO. The plates were then incubated for set times before reading on the Amersham LeadSeeker using fluorescence polarization mode with CY3 filters and epi-mirror. Total binding was determined by adding 10 µL 2.5 %

DMSO in water to 8 separate wells and non-specific binding was determined by the addition of 10 µL of the respective compound to 8 separate wells, instead of the test compound.

Pregnane X Receptor (PXR) binding assay using a lanthanide time-resolved fluorescence method (LanthaScreen, Invitrogen)

LanthaScreen™ TR-FRET PXR Competitive Binding Assay Kits were purchased from Invitrogen (CAT # PV4839). In this assay, a terbium-labeled anti-GST antibody is used to indirectly label the nuclear receptor PXR by binding to its GST tag. When a fluorescent ligand (tracer) is bound to the receptor, energy transfer from the antibody to the tracer occurs, and a high TR-FRET ratio is observed. The assay was performed in a final volume of 50 µL per well in a 384-well solid black plate (Perkin Elmer, CAT# 6007279).

The components of the wells were added as follows:

1. 0.25 µL test compounds in 100% DMSO. Aliquots were pipetted into the dry plate using a Hummingbird (Genomic Solutions, Ann Arbor MI), followed by the addition of 9.75 µL distilled water.
2. 20 µL Fluormone™ PXR Green. The final concentration of Fluormone (ligand) was 30 nM.
3. 20 µL of a PXR Receptor & Tb-anti-GST Ab solution. The final concentration of PXR receptor was 3.75 nM and the final concentration of Tb-anti-GST was 3.75 nM.
4. The plates were allowed to sit at room temperature for at least 1 hour and not longer than 4 hours.
5. Plates were read in a Perkin Elmer Envision reader. The signal measured is the fluorescence ratio 520 nm / 495 nm.

For total binding 10 µL 2.5 % DMSO in distilled water was added to wells A23 to H23 before adding the Fluormone. For non-specific binding 10 µL SR12813 at 100 µM (20 µM final concentration) was added to wells I23 to P23. The reference curve (SR12813) in wells A24 to P24 was also added in 10 µL at this point. The final DMSO concentration in each well was 0.5%.

Nuclear Receptor (NR) functional assays using a lanthanide time-resolved fluorescence (TR-FRET) method (LanthaScreen, Invitrogen)

LanthaScreen™ TR-FRET PR Coactivator Assay Kits were purchased from Invitrogen (e.g. CAT # PV4666 for progesterone). In this assay, a terbium-labeled anti-GST antibody is used

to indirectly label a nuclear receptor by binding to its GST tag. When an agonist binds to the receptor, a conformational change takes place allowing a fluorescein-labelled coactivator to bind with high affinity. When the terbium label on the anti-GST antibody is excited at 340 nm, energy is transferred to the fluorescein label on the bound coactivator peptide and detected as emission at 520 nm.

The assays were performed in a final volume of 50 µL (agonist), 55 µL (antagonist) per well in the same 384-well solid black plate (Perkin Elmer, CAT# 6007279). The agonist assay was performed first and then the antagonist assay in the same plate.

The components of the wells were added as follows:

1. 0.25 µL test compounds in 100% DMSO, plus 9.75 µL distilled water.
2. 40 µL GST-PR-LBD, Tb-antiGST antibody and coactivator peptide.
3. The plates were allowed to sit at room temperature for 1 hour and then read in a Perkin Elmer Envision. The signal measured is the fluorescence ratio 520nm/495nm.
4. For antagonist activity (immediately after the agonist read): Add 5 µL agonist control at ~EC_{80}. The plates were allowed to sit at room temperature for 1 hour and then read in a Perkin Elmer Envision.

Cyclooxygenase assays using an amplex red method

Assays were performed in 384 well black flat-bottom plates (Corning Cat #3654) with a total volume of 50 µL consisting of: 10 µL of sample (test compound, buffer or the reference diclofenac), 20 µL of COX enzyme and 20 µL arachidonic acid (AA) substrate with amplex red (AR). COX enzymes were purchased from Cayman Chemicals (Ann Arbor, MI, USA; ovine COX-1 Cat# 60100, human COX-2 Cat# 60122). Arachidonic acid (COX substrate) was purchased from Biomol (Plymouth Meeting, PA, USA; Cat # FA-003). Amplex red was purchased from Anaspec (San Jose, CA, USA; Cat # 85500). Diclofenac was purchased from Sigma (Cat# D6899).

Assay plates were prepipetted with 10 µL test compound. 20 µL of enzyme (COX) was added to all wells, except the minus COX control wells, using a Multidrop (Thermo Scientific, Waltham, MA, USA). The plates were covered and incubated for 30 minutes at room temperature. Arachidonic acid and amplex red were prepared from frozen stocks immediately before addition to plates to minimize unwanted reduction of amplex red to resorufin. 20 µL of the AA/AR solution was added to all wells using a Multidrop and plates were covered and incubated for another 30 minutes at room temperature. The final concentrations of the various reagents were: ovine COX-1 30 nM, AA 6 µM, AR 25 µM. Plates were read in a SpectraMAX Gemini EM (Molecular Devices, Sunnyvale, CA, USA).

Wells were read 6 times and the average value was used to analyze the data. The plates were read at the following wavelengths: excitation 530 nm, emission 590 nm, top read only.

The maximum signal was determined by adding 10 µL 2.5 % DMSO in water to 8 separate wells and the minimum signal was determined by the addition of 10 µL of diclofenac to 8 separate wells, instead of the test compound.

Monoamine Oxidase Assay (MAO)

The MAO luminescence assay was performed in a final volume of 60 µL per well in a 384-well plate using a kit sold by Promega (Cat No. V1402) using a luminogenic substrate (a derivative of beetle luciferin ((4S)-4.5-dihydro-2(6-hydroxybenzothiazolyl)-4-thiazole-carboxylic acid)). The components of the wells were added in two steps:

Step 1: Enzymatic step

6 µL test compound in 1.5% DMSO/distilled water.

9 µL MAO substrate. Final concentration: 40 µM.

15 µL MAO-A enzyme from Sigma (ref. M7316). Final concentration: 1 µg/well.

The plates were incubated at room temperature for 1 hour (volume: 30 µL). The maximum signal was determined by adding 6 µL water with 1.5% DMSO instead of test compound and the minimum signal was determined by the addition of 6 µL clorgyline (10 µM final concentration).

Step 2: Detection step

30 µL Luciferin detection reagent was added

The plates were then incubated at room temperature for 20 minutes and were counted using a BMG-Labtech PHERAstar reader (measurement interval time = 1s).

Phosphodiesterase (PDE) assays

The assay was performed in a final volume of 50 µL per well in a 384 well Optiplate. The reagents were added as follows: 10 µL of test compound in 2.5 % DMSO, 10 µL of a 2.5 % DMSO in water solution or an appropriate PDE inhibitor in the control wells (for the determination of the total and non specific signals, 20 µL of the [^{3}H]cAMP tracer, 20 µL of the PDE enzyme solution. The plate was then incubated at room temperature for the time specified and the reaction was ended by the addition of 30 µL of a solution containing Yttrium silicate beads (Amersham - RPNQ0150) in 30 % glycerol. The plates were incubated at room

temperature for at least 30 minutes, sealed using a TopSeal-S and counted in a Packard Topcount, each well counted for 1 min.

hERG (KCNH2) binding assay

The radioligand binding assay was performed using filtration to separate bound radioligand from free by a method similar to a previously described approach.[11] Crude cell membranes were prepared from a CHO cell line constitutively expressing functional hERG channels, which was purchased from AVIVA Biosciences (San Diego, CA) and cultured according to the recommended method.

The following components were pipetted into each well of pre-wet 96-well Millipore GF/C filter plates (#MSFCN6B50): 119 µl assay buffer (20mM HEPES, 10 mM KCl, 1 mM MgCl2, pH=7.4 at room temperature), 1 µl test compound in 100% DMSO (or 100% DMSO only for total binding), 40 µl [^{3}H]-dofetilide (12.5 nM, final concentration 2.5 nM; Novartis radioisotope laboratory, East Hanover, NJ, USA, specific activity 15-45 Ci/mmol); 40 µl crude membrane suspension (ca. 15 µg protein). The final concentration of DMSO during the incubation was 0.5%. Incubations were performed at room temperature for 90 min. Non-specific binding (NSB) was defined as the binding remaining in the presence of 25 µM terfenadine (Sigma T9652). The incubations were terminated by rapid filtration on a Millipore filtration manifold, followed by three washes of 200 µl ice-cold assay buffer. The plates were left to dry overnight before adding 40 µl scintillant (MicroScint-20). The plates were then sealed (Sealing Tape SI, Nunc 236366) and read in a Wallac MicroBeta Trilux beta-counter for 1.5 min per well.

Voltage gated Sodium channel (SCN5A) patch clamp assay

The assay has been carried out as described previously.[12]

Binding curve calculation

A standard data reduction algorithm was used to calculate percent specific binding in the presence of test compound as follows:

((B - NSB) / (T - NSB)). 100, where

B = Binding in the presence of test compound

NSB = Non-specific binding in the presence of excess inhibitor

T = Binding in the absence of test compound

Curve fitting of the normalized concentration response curves using a four parameter logistic fit was then performed using XLfit (v.2 or v.4, IDBS, Guildford, UK) or an equivalent internally developed software. The equation used was XLfit equation no. 205 for a one site sigmoidal dose response curve: Y=A+((B-A)/(1+((C/X)^D))), where A=min, B=max, C=IC50, D=slope (nH). By default, min is fixed at 0, whereas max is not fixed.

For aesthetic purposes, curves were re-plotted using GraphPad Prism 5.0.

Platelet aggregation

Platelet aggregation assessment was carried out as described previously,[13] with the following exceptions:

1. A mix of CaCl2 (3 mM) and physiological saline (0.9%) was used instead of plain saline as blood diluent (as suggested by the manufacturer, this is to partially recalcify citrate-anticoagulated whole blood);

2. Whole blood samples were incubated with acetylsalicylic acid (ASA), indomethacin and chlorotrianisenefor 10 min at room temperature before being added to the 37°C preheated CaCl2/saline diluent;

3. Collagen was used as agonistat 1 µg/mL concentration that induces platelet responses halfway between the baseline and maximum in order to detect potential hyper-aggregability. Aggregation was recorded during 15 minutes instead of 5 minutes.

In brief, human blood specimens were drawn by venipuncture into EDTA and Citrate Monovette® collecting tubes from 6 healthy male volunteer donors (nos. 1M to 6M). Evaluation of platelet count and related parameters were performed on EDTA-anticoagulated blood using an ADVIA 2120® analyzer (Siemens, Healthcare Diagnostics GmbH, Eschborn, Germany). Citrate-anticoagulated blood was incubated with test solutions or vehicle for 10 min at room temperature. Platelet aggregation was recorded for 15 min at 37°C in whole blood using a Multiplate® impedance aggregometer (Dynabyte Medical, Munich, Germany), starting with the addition of collagen (Dynabyte Medical, Munich, Germany) at a concentrations that yield half-maximal platelet aggregation (1 µg/mL). The area under the curve (AUC) recorded by the aggregometer was used to express platelet aggregation response over the recorded time (aggregation units x min).

Specimens were analyzed as follows:

- Vehicle + Collagen (1 µg/mL);

- ASA (250 µM) + Collagen (1 µg/mL);

- Indomethacin (0.5 µM, 5 µM, 50µM) + Collagen (1 µg/mL);

- Chlorotrianisene (0.5 µM, 5 µM, 50µM) + Collagen (1 µg/mL).

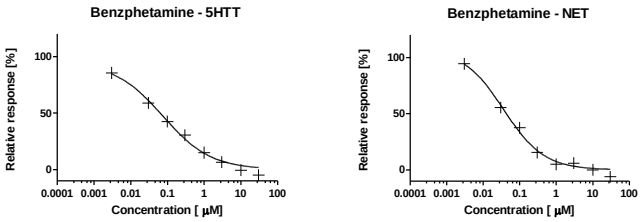
Values measured were compared with the in-house 5 to 95% centile reference normal range derived from measurements done in the same conditions on healthy men volunteer donor's blood samples (n= 81) incubated with vehicle only (AUC range for 1 µg/mL Collagen, 125-254). Two-tailed t-tests were performed and $p<0.05$ was considered statistically significant.

Supplementary Results

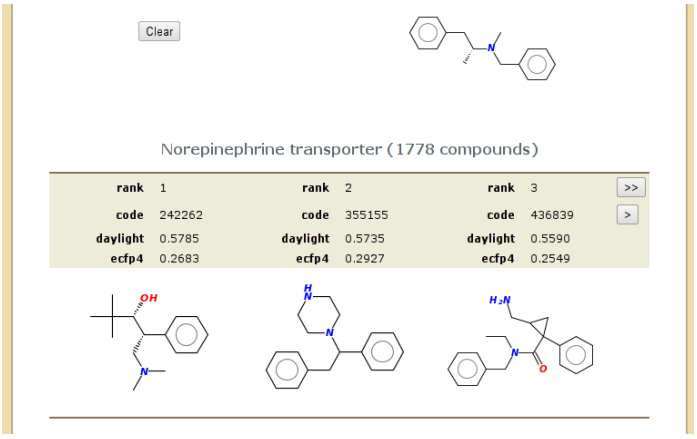
We assessed eight cases of internally confirmed, highly active (<1 µM) compounds at targets that were not predicted by SEA.

Benzphetamine

Benzphetamine was internally measured to bind to the serotonin transporter (5-HTT, 80nM) and the norepinephrine transporter (NET, 40nM):

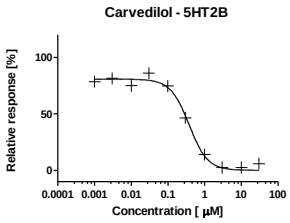


Additionally, 10 µM activity at the dopamine transporter has been measured internally. However, none of these targets were predicted by SEA. In this case, only few ligands in the reference database had a similarity score that positively contributed to the SEA score. For example, for NET, the highest ECFP4 similarity was <0.3:

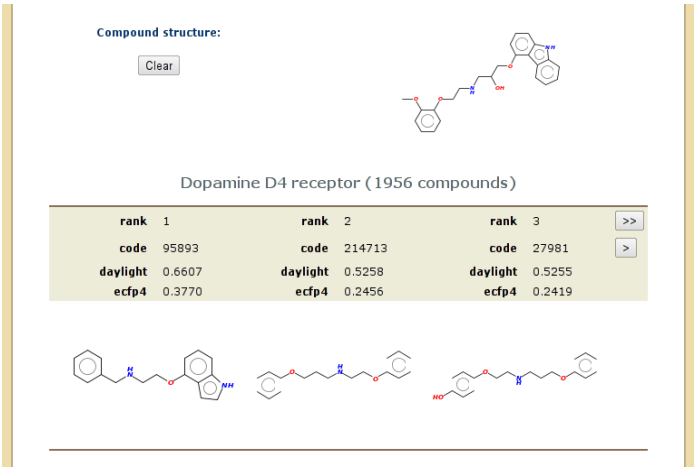


Carvedilol

This non-selective beta-blocker is promiscuous among GPCRs (hit rate 0.46, Supplementary Table S6). SEA correctly predicts the primary targets (ADRB1-3), as well as HTR1A, which is known[14] and may explain asthenia (weakness) observed with the compound. Two targets with high affinity that we measured and that were not previously known, nor predicted, are DRD4 (binding, 20nM, Tc 0.38) and HTR2B (antagonist, 500nM, Tc 0.56). Interestingly, HTR3A was wrongly predicted by SEA (Tc 0.57), perhaps reflecting the similarity to known serotonin analogues.



Similar to benzphetamine, only a very small number of reference ligands positively contributed to the SEA score. For example, only 5 out of 1956 DRD4 reference ligands had a ECFP4 Tc > 0.3. Thus, most similar compounds contributed insufficiently to a positive SEA score:



Although these compounds share some pharmacophore features, these similarity is not captured by the molecular fingerprints used in our study.

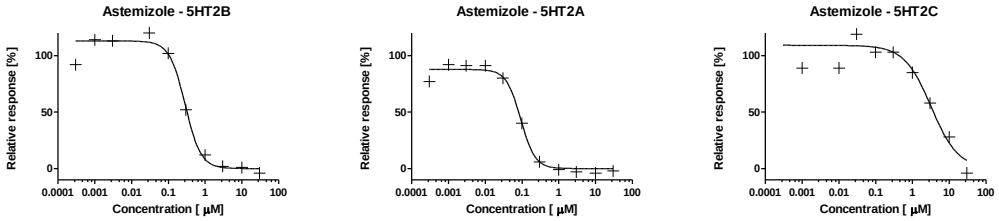
Dihydroergotamine

This compound is well known (and confirmed internally) to be promiscuous among GPCRs. Accordingly, SEA predicts activity at seven GPCRs with measured activity below 1uM, but with e-Values above the cutoff that we used for prediction calling (i.e., >1e-4). The best e-Value for any known target is for DRD2 (8.5e-4, Tc 0.4). However, SEA has correctly identified DRD1 as a target (e-Value 6e-10, Tc 0.4, 400nM agonist), which has not been

explicitly annotated in any databases that we have used for retrospective analysis. If targets had been predicted for this individual compound, the threshold could be easily adjusted, given predicted activity at the DRD1 receptor. Thus, in a practical application, the other dopamine receptors would have been tested as well, at the latest after confirmed activity at DRD1.

Astemizole

Among the top five most potent activities that we observed internally for this antihistamine, were HTR2A (100 nM) and HTR2B (330 nM). HTR2A is a known astemizole target,[15] and, furthermore, associated with adverse weight increase, which has been observed with this drug. SEA predicts both targets with highly significant e-Values (both: 6.2e-27; Tc 0.42), but the prediction was excluded using the charge filter. Similarly, SEA correctly predicted activities at HTR2C (2.2 µM, e-Value: 2e-26; Tc 0.42) and DRD2 (5.4e-27; Tc 0.42; 8.6 µM). In this case, the overall well-performing charge filter excluded four significant, true predictions, three of which were novel.



Promazine

The highest activity of this antipsychotic that we observed was at HRH1 (3 nM), which is also annotated in several databases, including Drugbank. SEA predicted this target with an e-Value of 4e-4 (Tc 0.59), close to our threshold for prediction hit calling. Similarly, novel activity at HRH2 was observed (110 nM) and predicted with a marginal e-Value of 3.1e-4. Moreover, two predictions of known off-targets that we have confirmed, but that failed the charge filter, included DRD3 (75 nM; e-Value 2.6e-6; Tc 0.75) and norepinephrine transporter (87 nM; e-Value 9.8e-6; Tc 0.59).

Trifluoperazine

Binding to DRD3 was measured internally (10 nM), representing a novel target in addition to the known DRD2 and other GPCRs. DRD3 was predicted by SEA (e-Value 8e-6; Tc 0.81), but did not pass the charge filter. This case illustrates that high chemical similarity to the closest ligand, a significant SEA prediction, as well as known activity at a closely related target may indicate activity irrespective of a single parameter such as the charge filter.

Thiothixene

For the antipsychotic thiothixene, no single target was predicted by SEA (all e-Values were >1). The compound is known and has been internally confirmed to bind multiple GPCRs with submicromolar activity. However, when ranked by increasing e-Values, among the top ten predictions, at 8 targets sub-micromolar activity was measured, and at HTR2C the compound was active at 2.8 µM. Thus, when predicting targets for individual compounds a ranking approach using the e-Values, rather than applying a cut-off, might provide additional hints toward off-target activity. However, when using for large-scale target prediction and compound prioritization, the threshold approach chosen here guarantees higher specificity.

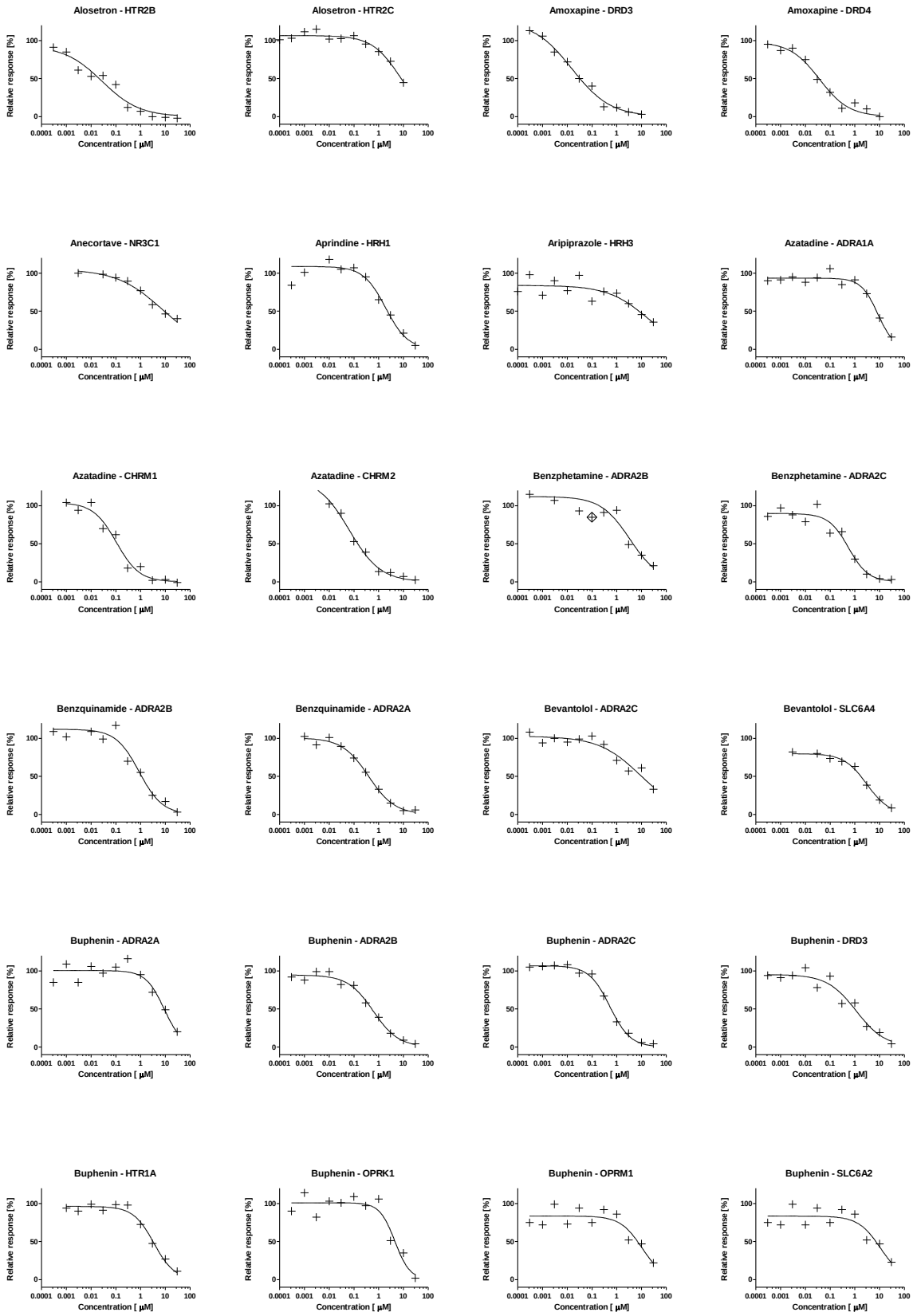
Danazol

For the steroid derivative danazol, known and internally confirmed nucleic receptor targets (AR, PGR) were predicted by SEA with marginal e-Values (23 nM; 7e-4; Tc 0.45 for AR and 20 nM; 4.4e-2; Tc 0.44 for PGR). While the molecular representations chosen here do not well capture the similarity to other steroids (presumably due to the oxazole fragment), due to the steroid backbone, these top-ranking predictions make intuitive sense to the medicinal chemist's eye.

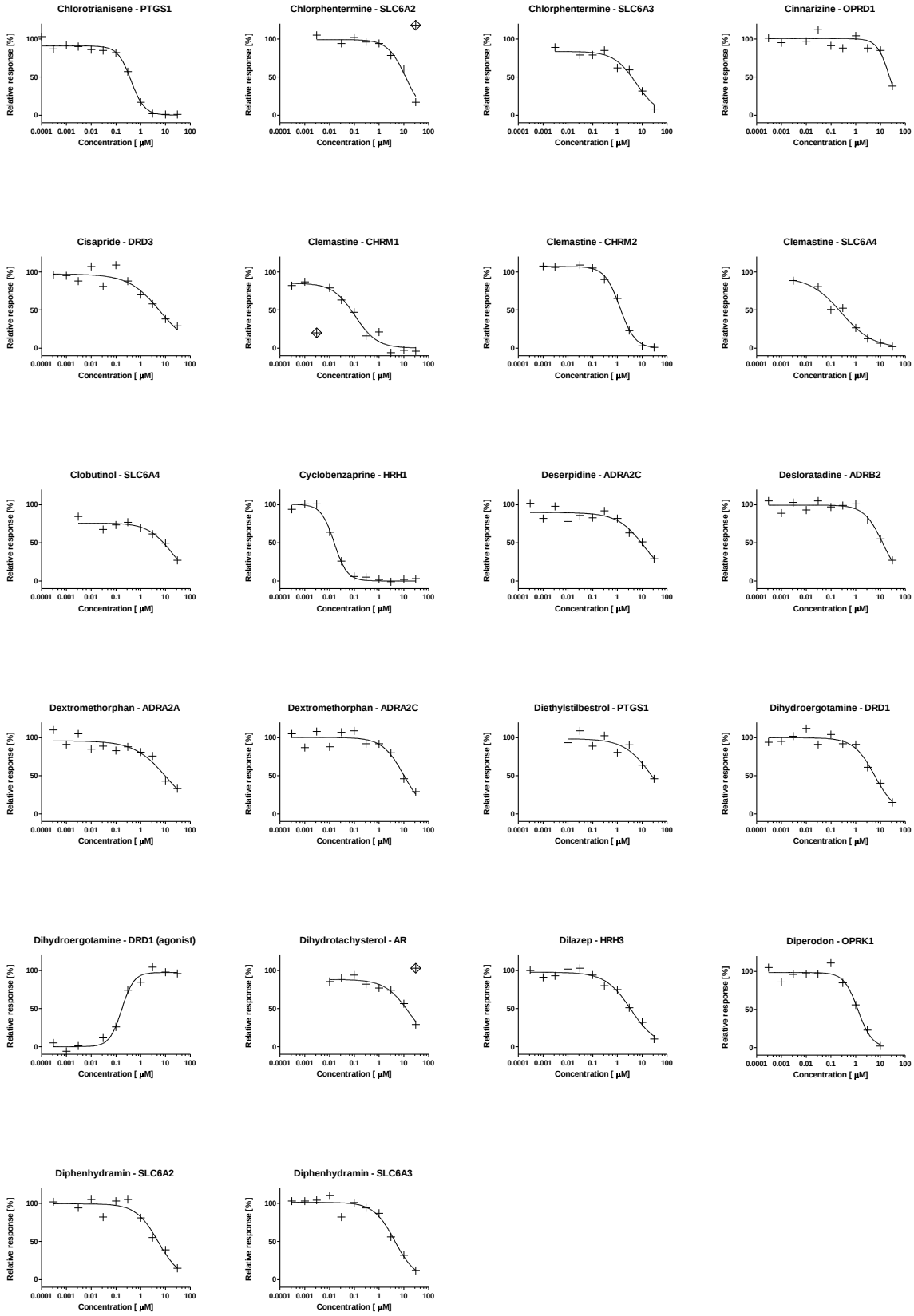
Supplementary References

1. Beaumont, K., Chilton, W. S., Yamamura, H. I. & Enna, S. J. Muscimol binding in rat brain: association with synaptic GABA receptors. *Brain Res* **148**, 153–162 (1978).
2. Engel, G. & Hoyer, D. [125I]BE 2254, a new high affinity radioligand for alpha 1-adrenoceptors. *Eur. J. Pharmacol.* **73**, 221–224 (1981).
3. Hoyer, D., Engel, G. & Kalkman, H. O. Molecular pharmacology of 5-HT1 and 5-HT2 recognition sites in rat and pig brain membranes: radioligand binding studies with [3H]5-HT, [3H]8-OH-DPAT, (-)[125I]iodocyanopindolol, [3H]mesulergine and [3H]ketanserin. *Eur. J. Pharmacol* **118**, 13–23 (1985).
4. Ransom, R. W. & Stec, N. L. Cooperative modulation of [3H]MK-801 binding to the N-methyl-D-aspartate receptor-ion channel complex by L-glutamate, glycine, and polyamines. *J. Neurochem.* **51**, 830–836 (1988).
5. Halme, M., Sjöholm, B., Savola, J. M. & Scheinin, M. Recombinant human alpha 2-adrenoceptor subtypes: comparison of [3H]rauwolscine, [3H]atipamezole and [3H]RX821002 as radioligands. *Biochim. Biophys. Acta* **1266**, 207–214 (1995).
6. Hedberg, A., Hall, S. E., Ogletree, M. L., Harris, D. N. & Liu, E. C. Characterization of [5,6-3H]SQ 29,548 as a high affinity radioligand, binding to thromboxane A2/prostaglandin H2-receptors in human platelets. *J. Pharmacol. Exp. Ther* **245**, 786–792 (1988).
7. Markstein, R. *et al.* SDZ GLC 756, a novel octahydrobenzo[g]quinoline derivative exerts opposing effects on dopamine D1 and D2 receptors. *J Neural Transm* **103**, 17–30 (1996).
8. Gabriel, D. *et al.* High throughput screening technologies for direct cyclic AMP measurement. *Assay Drug Dev Technol* **1**, 291–303 (2003).
9. Nunn, C. *et al.* Comparison of functional profiles at human recombinant somatostatin sst2 receptor: simultaneous determination of intracellular Ca2+ and luciferase expression in CHO-K1 cells. *Br. J. Pharmacol* **142**, 150–160 (2004).
10. Wigglesworth, M. J. *et al.* Use of cryopreserved cells for enabling greater flexibility in compound profiling. *J Biomol Screen* **13**, 354–362 (2008).
11. Diaz, G. J. *et al.* The [3H]dofetilide binding assay is a predictive screening tool for hERG blockade and proarrhythmia: Comparison of intact cell and membrane preparations and effects of altering [K+]o. *J Pharmacol Toxicol Methods* **50**, 187–199 (2004).
12. Cao, X. *et al.* Cardiac ion channel safety profiling on the IonWorks Quattro automated patch clamp system. *Assay Drug Dev Technol* **8**, 766–780 (2010).
13. Tóth, O., Calatzis, A., Penz, S., Losonczy, H. & Siess, W. Multiple electrode aggregometry: a new device to measure platelet aggregation in whole blood. *Thromb. Haemost.* **96**, 781–788 (2006).
14. Gaillard, P., Carrupt, P. A., Testa, B. & Schambel, P. Binding of arylpiperazines, (aryloxy)propanolamines, and tetrahydropyridylindoles to the 5-HT1A receptor: contribution of the molecular lipophilicity potential to three-dimensional quantitative structure-affinity relationship models. *J. Med. Chem.* **39**, 126–134 (1996).
15. Leysen, J. E., Niemegeers, C. J., Van Nueten, J. M. & Laduron, P. M. [3H]Ketanserin (R 41 468), a selective 3H-ligand for serotonin2 receptor binding sites. Binding properties, brain distribution, and functional role. *Mol. Pharmacol.* **21**, 301–314 (1982).

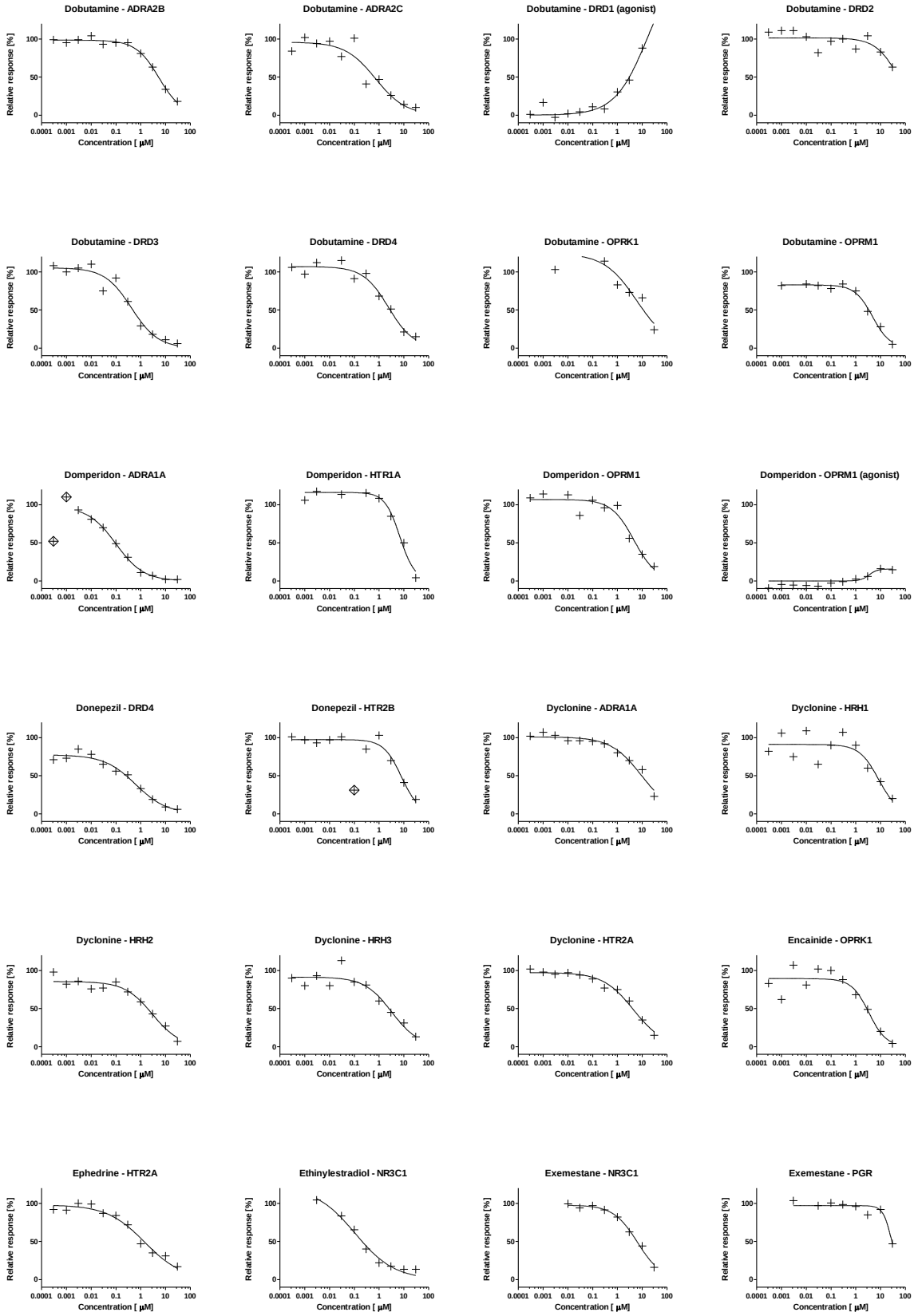
Supplementary Figure S1A



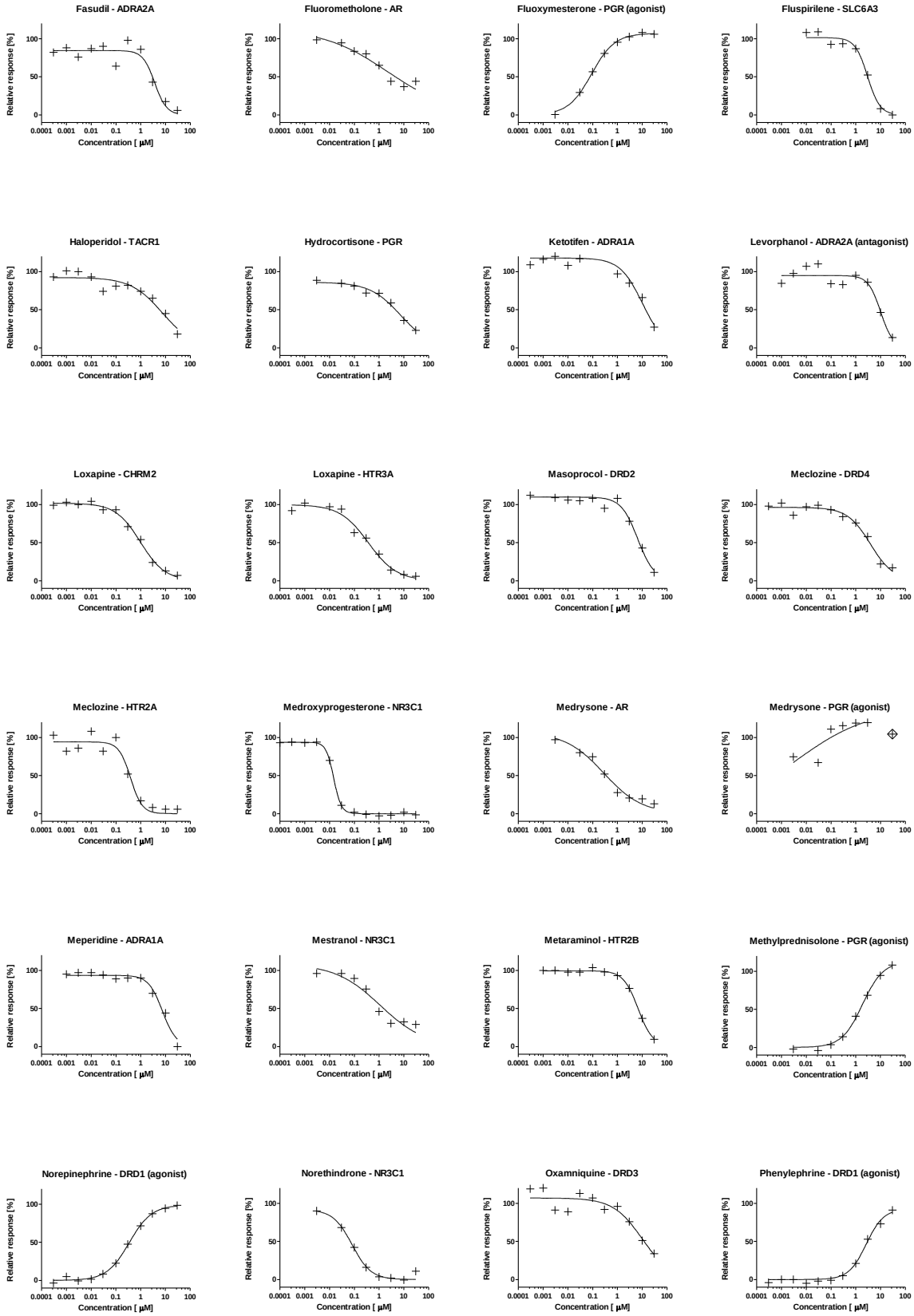
Supplementary Figure S1B



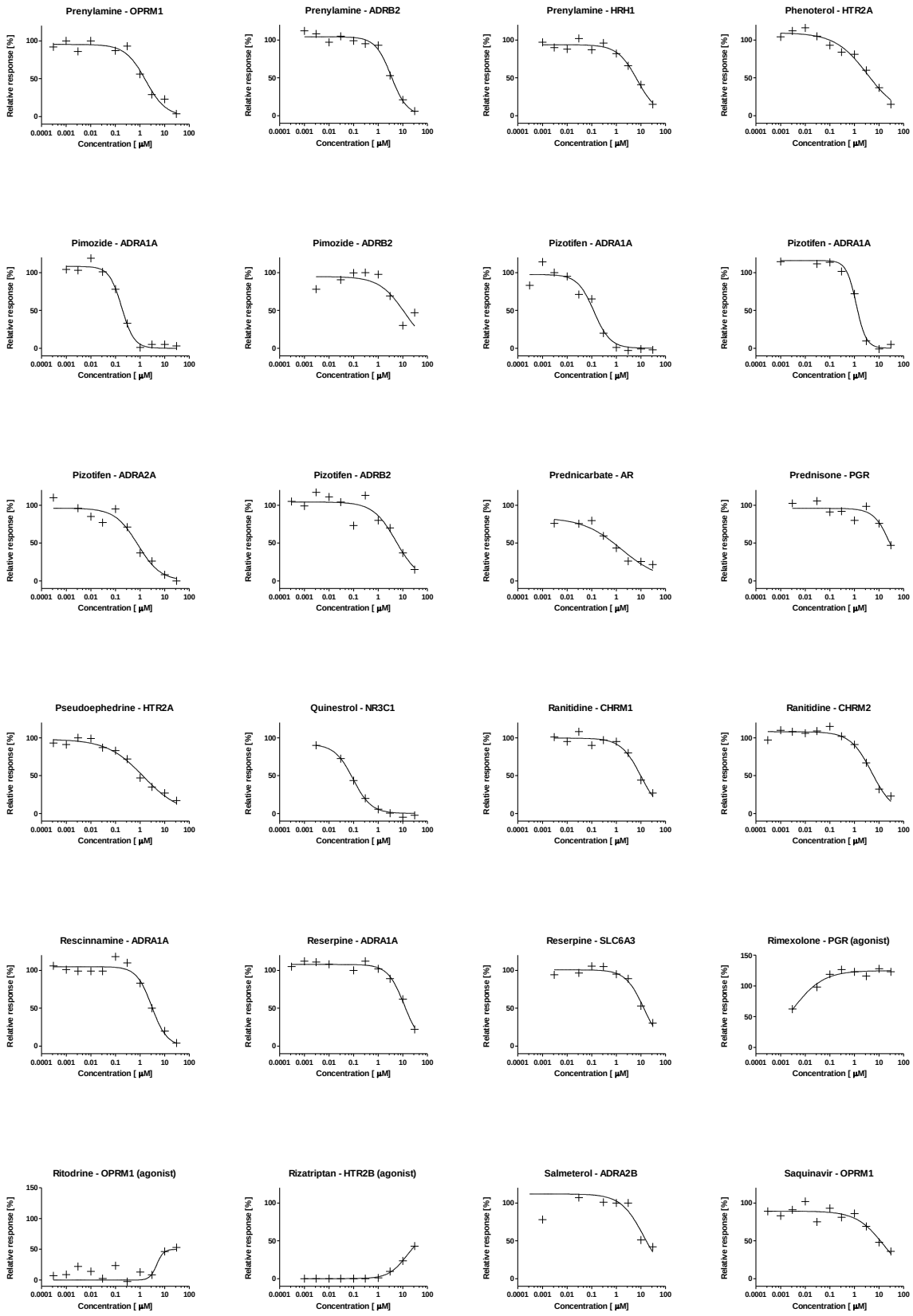
Supplementary Figure S1C



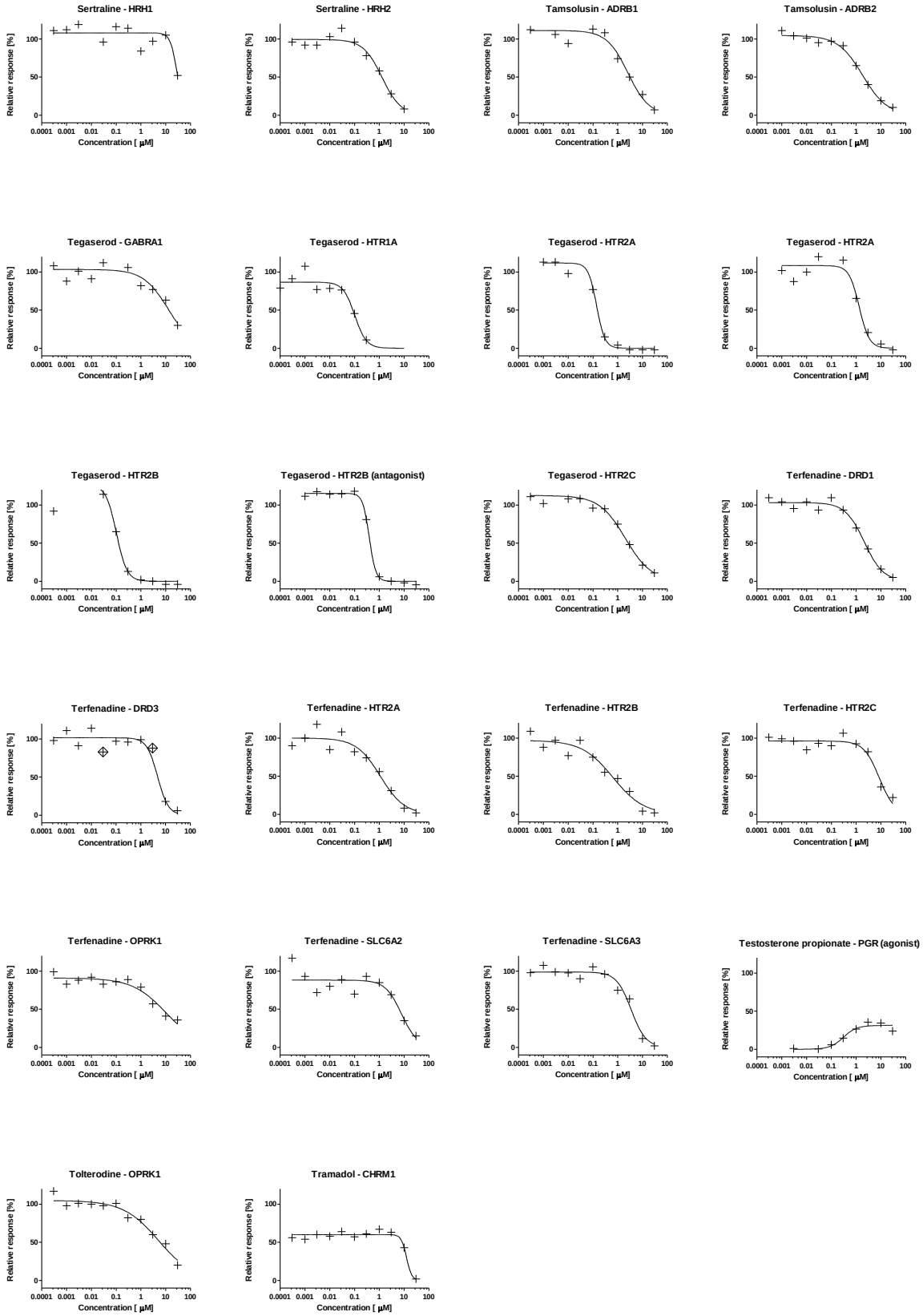
Supplementary Figure S1D



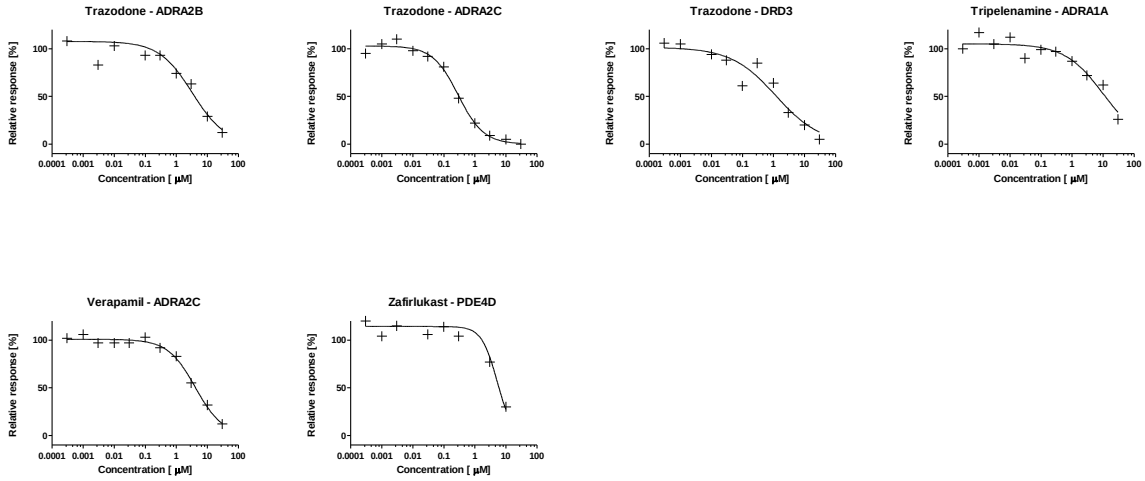
Supplementary Figure S1E



Supplementary Figure S1F



Supplementary Figure S1G



(A-G) Binding and, where specified, functional agonist / antagonist concentration-response curves for novel predicted confirmed off-targets. Targets identifiers correspond to Supplementary Table S2. Boxed markers are masked values.

Supplementary Tables

Supplementary tables are available in Excel format (Supplementary_Tables.xls). Each worksheet corresponds to one Supplementary Table. The table content and columns are described below.

Supplementary Table S1 Drugs used in study lists the 656 drugs assessed in the study together with their known human (off-)targets.
Drug name: Generic name of drug used in the study;
SMILES: Simplified molecular input line entry specification string encoding the drug structure;
Known human targets (Drugbank, ChEMBL): Entrez gene symbols of human (primary and off-) targets annotated in Drugbank (version 3.0) or ChEMBL (version 9)
Target classes: ChEMBL target classes (levels 1-3) of known human targets.

Supplementary Table S2: Safety targets lists the 73 targets for which predictions have been made.
Target: Representative Entrez gene symbol designating target;
Target description: Entrez target description;
ChEMBL targets: Uniprot identifiers of ChEMBL targets used to assemble reference ligand sets for SEA;
Human Entrez gene IDs: Entrez gene IDs that were mapped to each target across all databases;
Assay technology: Assay technolog(ies) used for binding and functional assays;
Ligand / Substrate: Radioligand or substrate (enzymatic assays) used;
Membrane source: Cell line, organ, or animal source of membranes.
Target class level 1-3: ChEMBL Target taxonomy term from levels 1-3;

Supplementary Table S3: Confirmed predictions lists predictions that were not found in literature, but were confirmed in internal assays.
Drug name: Generic name of drug;
Drug SMILES: Simplified molecular input line entry specification string encoding the drug structure;
Closest ChEMBL neighbor SMILES: SMILES string encoding the structure of the ChEMBL reference compound structurally most similar to the drug based on ECFP_4 fingerprint comparison;
ECFP_4 Tc: The Tanimoto coefficient similarity between the drug and its closest ChEMBL neighbor;
Combined Tc: The higher of the two ECFP_4 or Daylight fingerprint Tc values;

SEA E-Value: SEA scores indicating the overall similarity of the drug to the reference set;
Target: Representative Entrez gene symbol of the predicted target;
Closest known target: Entrez gene symbol of known human (off-)target of drug that had highest sequence similarity to the predicted target;
BLAST E-value: Score indicting the sequence similarity between predicted and closest known target. Small values correspond to increasing similarity.
Confirmed activity [uM]: IC_{50} (or EC_{50} for agonist assays) values in micromolar measured in binding or functional assays for the drug and predicted target;
Functional annotation: "Binding" for binding assays, and "Agonist" / "Antagonist" for functional assays.

Supplementary Table S4: Comparison to 1NN reports false positive rates for SEA versus an alternative 1-nearest neighbor (1NN) approach.
Drug: Generic drug name;
Combined model: SEA model using both ECFP_4 and Daylight fingerprints;
ECFP4: SEA model using only ECFP_4 fingerprints;
Daylight: SEA model using only Daylight fingerprints;
Tc: Lowest Tanimoto coefficient between drug and any predicted target reference set compound;
SEA: False positive rate of SEA (targets confirmed / targets predicted);
1NN: False positive rate of 1NN at the same similarity threshold (targets confirmed / targets predicted).

Supplementary Table S5: Target-ADR associations lists pairs of significantly associated targets and adverse events.
Target: Entrez gene id of human target;
Adverse drug reaction: MedDRA preferred term coding for adverse reaction;
EF: Enrichment factor;
p-Value: Chi-square test p-Value of association;
q-Value: p-Value corrected for multiple hypothesis testing;
Chi square statistic: Value of statistic calculated from contingency table;
Functional annotation: Indicates whether all drug-target links from databases were used ("Binding") or if agonist and antagonist action were distinguished.

Supplementary Table S6: Novel off-targets and ADRs reports associations between novel off-targets and side effects of drugs for cases where the novel target was more strongly associated with the adverse event than any known target.
Drug: Generic drug name;
Target: Representative Entrez gene symbol;

Confirmed activity [uM]: Micromolar activity measured in internal assays;
AUC [uM * h] range (median): Area under concentration curve in µM*h. The range and median of reported values for human plasma concentrations are reported;
Cmax [uM] range (median): Cmax in plasma measured in humans in µM. The range and median of reported values for human plasma concentrations are reported;
Adverse drug reaction: MedDRA preferred term coding for adverse reaction;
EF: Enrichment factor of adverse drug reaction for predicted target, and, if a significant association with a known target was found, the highest enrichment factor of adverse drug reaction for any known target;
Alternative target: A known target most highly associated with the adverse reaction;
Comparable drugs: Other drugs that cause the adverse reaction, and have a similar in vitro potency at target, as well as a similar median AUC or Cmax value (i.e. within one order of magnitude). Asterisks indicate that the target is the primary target or a very well-known off-target of the drug.

Supplementary Table S7: Target promiscuity reports how many of the tested drugs were active at each safety target.
Target: Representative Entrez gene symbol of safety target;
Drugs active: Number of drugs binding target with $IC_{50} < 30$ µM;
Drugs tested: Number of drugs tested in binding and/or functional assay or annotated in databases;
Hit rate: Fraction of tested drugs active at the target

Supplementary Table S8: Drug promiscuity reports how many of the tested targets each drug was active against.
Drug: Generic drug name;
Targets hit: Total number of safety targets the drug bound with an $IC_{50} < 30$ µM;
Targets tested: Number of targets the drug was tested against;
Hit rate: Fraction of tested targets the drug was active against
AlogP: Calculated lipophilicity of the drug;
Charge: Calculated formal charge at pH 7.4;