Supporting Information

Reker et al. 10.1073/pnas.1320001111

SI Materials and Methods

All starting materials and solvents were purchased from Sigma-Aldrich, Fluka, or ABCR and were used without further purification.

Microwave syntheses were performed on a Biotage Initiator microwave synthesizer.

Melting points (mp) were recorded on a Büchi M560 apparatus and are uncorrected. Proton and carbon NMR (^{1}H and ^{13}C NMR) spectra were recorded on a Bruker Avance 400 (400 and 100 MHz, respectively). All chemical shifts are quoted on the δ scale in parts per million using residual solvent peaks as the internal standard. Coupling constants (*J*) are reported in hertz with the following splitting abbreviations: s, singlet; d, doublet; dd, doublet of doublets; td, triplet of doublets; m, multiplet.

Analytical HPLC-MS was carried out in a Shimadzu LC-MS2020 system, equipped with a Nucleodur C_{18} HTec column, under an appropriate gradient of acetonitrile: H_2O (+0.1% formic acid in each solvent), and a total flow rate of 0.5 mL/min. Preparative HPLC was carried out on a Shimadzu LC-8A system, coupled to a Nucleodur 100–5 C_{18} HTec column, and a SPD-20A UV/Vis detector. High-resolution MS (HRMS) analyses were performed on a Bruker Daltonics maXis electrospray ionization (ESI) quadrupole time-of-flight device. MS analyses were operated in positive-ion mode with ESI. Nominal and exact *m/z* values are reported in daltons. All compounds present purity ≥95% based on LC-MS analysis.

Docking. Docking was performed using GOLD (version 5.2) with HIV1 protease cocrystalized with the de novo design template amprenavir [Protein Data Bank (PDB) code 1HPV]. The binding site was defined by the position of amprenavir. **1** and **2** were preprocessed using the Molecular Operating Environment (MOE; Chemical Computing Group, Montreal) wash function, and starting structures were calculated using the MOE energy minimizer. Hydrogen atoms were added to the protein, and waters were deleted using internal GOLD functions. Atom typing for ligand and protein was activated, and conformations were scored using the ChemPLP scoring function.

Repurposing. The publicly available target prediction methods Similarity Ensemble Approach (SEA; http://sea.bkslab.org/; version ChEMBL 10), Prediction of Activity Spectra (PASS; www.pharmaexpert.ru/passonline/), Semantic Link Association Prediction (SLAP; http://cheminfov.informatics.indiana.edu:8080/slap/), and the method SuperPRED (http://bioinformatics.charite.de/superpred/) were applied using their freely available online services. Compounds **1** and **2** and the design template amprenavir were provided as simplified molecular input line entry system (SMILES). Results were manually extracted and sorted according to provided confidence scores (scores or *P* values). For the PASS method, which provides two orthogonal confidence scores, *P*(active) and *P*(inactive), the ratio of these scores *P*(active)/*P*(inactive) was used.

Library of Pharmacologically Active Compounds Off-Target Prediction. The library of pharmacologically active compound (LOPAC) collection (version 1280) was downloaded (www.sigmaaldrich.com) and preprocessed as described in the main text. Previously precomputed self-organizing maps (SOMs) clustering the whole collection of bioactive reference analogs (COBRA, inSili.com LLC, Zurich) were used to project queries onto clusters of reference compounds. Tanimoto similarities (Tc) of every query were computed from Dayllight fingerprints of coclustered drugs. When a prediction for a particular target used at least one reference with a Tc > 0.2, the prediction was discarded. This procedure left us with a list of predictions using structurally unrelated references. Filtering for high-confidence predictions ($P < 0.001$) left us with the final set of predictions.

Screening Assays. Activity tests were performed by contractors on a fee-for-service basis. EC_{50} values against the B_1 receptor were determined by measuring intracellular calcium by fluorimetry in human recombinant CHO cells at Cerep (Ref. 1109) (1). The agonist effect on vanilloid 1 (TRPV1) was determined by measuring intracellular calcium by fluorimetry in human recombinant CHO cells at Cerep (Ref. 1640) (2). The antagonist effect on neurokinin 1 (NK1) was determined by measuring intracellular calcium by fluorimetry in U373 cells at Cerep (Ref. 2192) (3). Screening against proteases was carried out at Reaction Biology Corp. using suitable assay kits.

Spectral Data for Compound 1. ^{1}H NMR (CD_3OD, 400.13 MHz) d 1.36 (9H, s, $(CH_3)_3$), 1.91–2.02 (2H, m, CH_2), 2.43 (3H, s, CH_3), 2.62 (1H, dd, J = 13.6 and 9.7 Hz, CH_2), 2.80 (1H, dd, J = 13.6 and 5.2 Hz, CH_2), 3.35–3.52 (4H, m, CH_2), 3.74 (1H, d, J = 5.0 Hz, C*H*), 3.94 (1H, m, C*H*), 5.03 (1H, m, C*H*), 6.50 (1H, d, J = 9.4 Hz, N*H*), 7.16–7.27 (5H, m, Ar-*H*), 7.42 (2H, d, J = 8.0 Hz, Ar-*H*), 7.73 (2H, d, J = 8.0 Hz, Ar-*H*); ^{13}C NMR (CD_3OD, 100.61 MHz) d 21.55; 28.74; 32.28; 37.00; 47.49; 54.55; 56.12; 73.86; 75.65; 80.23; 127.37; 128.79; 129.28; 130.55; 130.64; 131.02; 134.78; 139.48; 145.38; 173.53; HRMS-ESI calc. ($C_{26}H_{34}N_2O_7S + Na^+$): 541.1979, found: 541.1976.

Spectral Data for Compound 2. ^{1}H NMR (CD_3OD, 400.13 MHz) d 1.63–1.80 (2H, m, CH_2), 2.32 (3H, s, CH_3), 2.76 (1H, dd, J = 14.0 and 7.0 Hz, CH_2), 2.85 (1H, dd, J = 14.0 and 7.0 Hz, CH_2), 3.16 (2H, dd, J = 9.0 and 5.2 Hz, CH_2), 3.24 (2H, m, CH_2), 3.73 (1H, td, J = 7.0 and 3.0 Hz, C*H*), 4.17 (1H, d, J = 3.0 Hz, C*H*), 4.74 (1H, m, C*H*), 7.11–7.26 (5H, m, Ar-*H*), 7.31 (2H, d, J = 8.2 Hz, Ar-*H*), 7.60 (2H, d, J = 8.2 Hz, Ar-*H*); ^{13}C NMR (CD_3OD, 100.61 MHz) d 21.50, 32.06, 34.51, 47.20, 54.42, 56.07, 70.56, 76.37, 128.66, 128.77, 130.02, 130.73, 130.98, 134.71, 136.50, 145.50, 171.37; HRMS-ESI calc. ($C_{21}H_{26}N_2O_5S + H^+$): 419.1635, found: 419.1634.

1. Simpson PB, Woollacott AJ, Hill RG, Seabrook GR (2000) Functional characterization of bradykinin analogues on recombinant human bradykinin B(1) and B(2) receptors. *Eur J Pharmacol* 392(1-2):1–9.
2. Phelps PT, Anthes JC, Correll CC (2005) Cloning and functional characterization of dog transient receptor potential vanilloid receptor-1 (TRPV1). *Eur J Pharmacol* 513(1-2):57–66.
3. Eistetter HR, et al. (1992) Functional characterization of neurokinin-1 receptors on human U373MG astrocytoma cells. *Glia* 6(2):89–95.

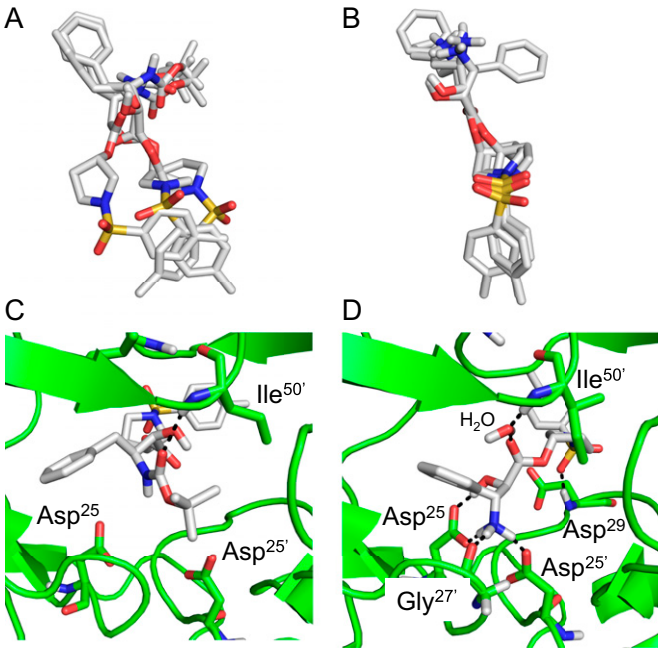

Fig. S1. Superimposition of the top three docking poses for **1** (*A*) and **2** (*B*). Docking poses of compounds **1** (*C*) and **2** (*D*) in the catalytic site of HIV-1 protease (PDB code 1HPV) (1). Hydrogen bonds with Asp^{25} residues are highlighted. Images were generated with PyMOL v1.4.1 (Schrödinger), and for clarity, only polar hydrogen atoms are shown.

1. Kim EE, et al. (1995) Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. *J Am Chem Soc* 117(3):1181–1182.

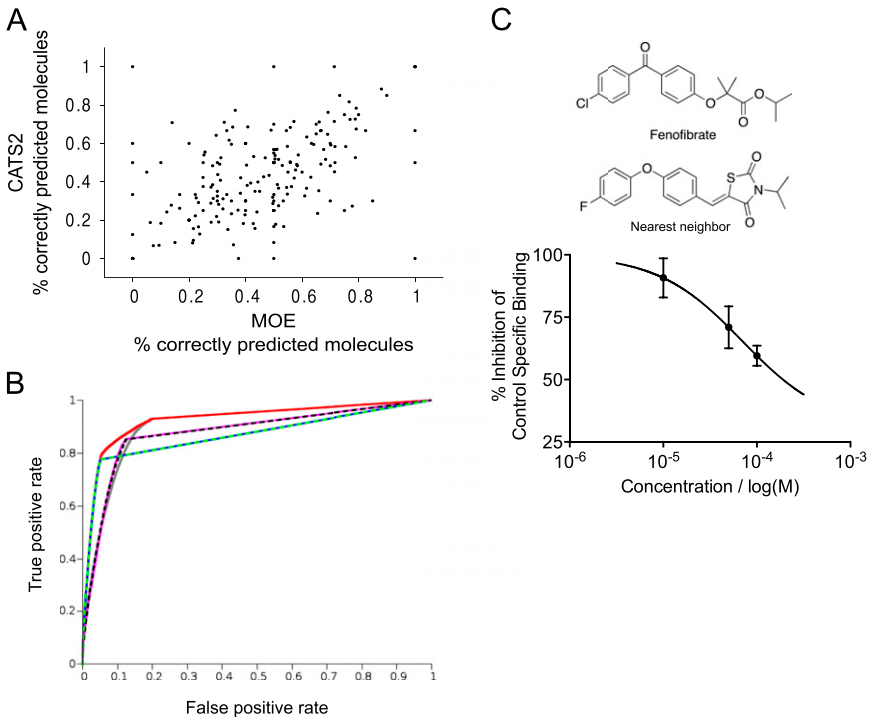

Fig. S2. (*A*) Performance of the prediction models using solely the chemically advanced template search (CATS) or the MOE representation, measured in percentage of molecules for which the correct target is predicted with significant confidence ($P < 5\%$). Each point represents the performance on molecules associated with one particular target. (*B*) ROC curves for the different combination functions (arithmetic average, red; maximum, gray; minimum, blue; geometric average, dashed green; CATS2 only, pink; MOE only, dashed black). (*C*) Structures of fenofibrate and its nearest neighbor from the training data annotated to block sodium ion channels and the concentration-response curve of fenofibrate against $Na_v1.5$ ion channels ($n = 2$, mean and SEM).

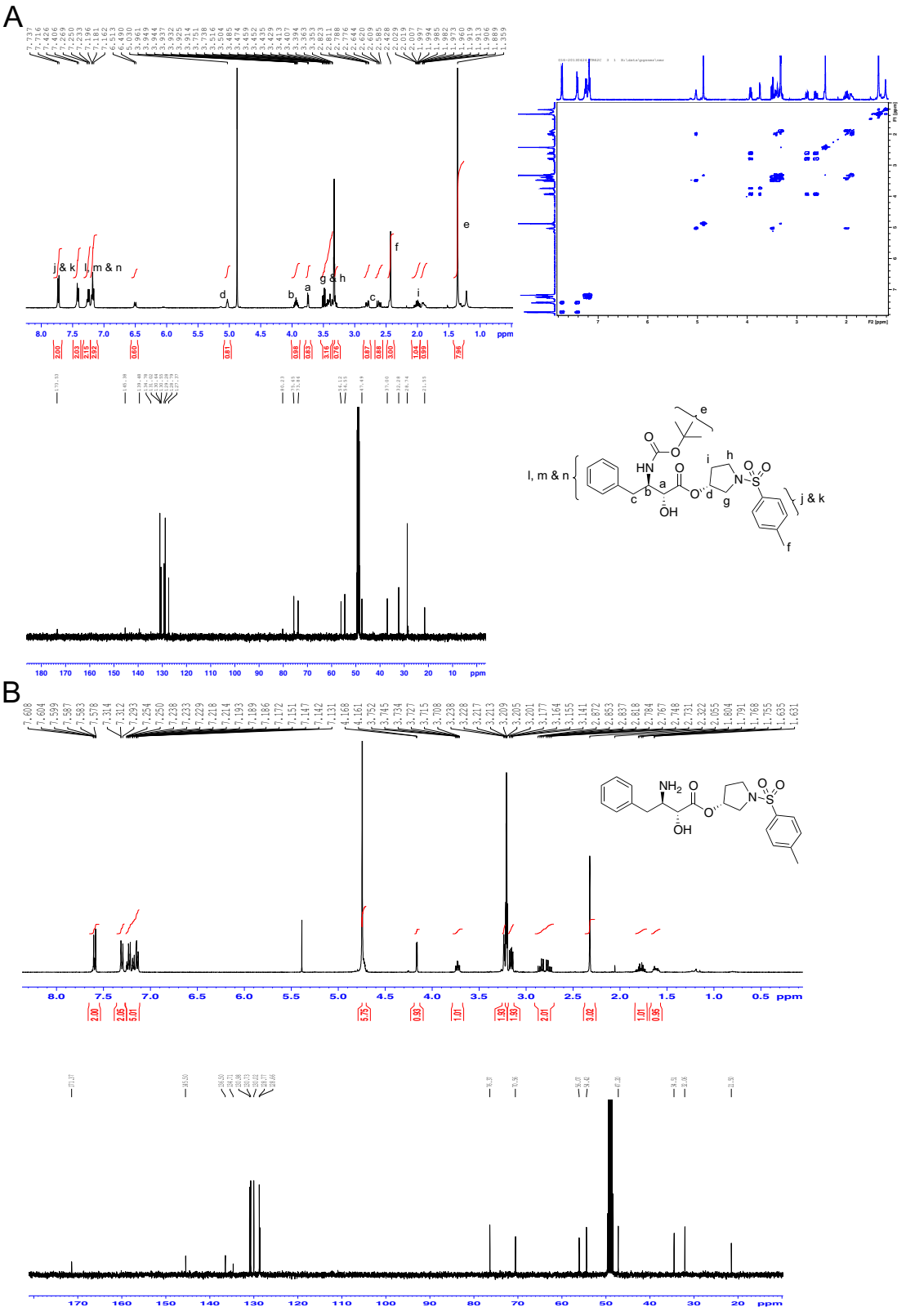

Fig. S3. ^{1}H NMR, ^{13}C NMR, and ^{1}H-^{1}H COSY data for compound **1** (*A*) and ^{1}H NMR and ^{13}C NMR data for compound **2** (*B*).

Table S1. Top 10 SEA predictions for compounds 1 and 2 and the scores for elastase and the B_1 receptor, computed with default parameters (ChEMBL v10, ECFP4)

Target	E value	maxTC
Compound **1**		
HIV type 1 protease	3.65e−127	0.51
BACE-1	1.47e−85	0.40
Renin	3.05e−82	0.43
Cathepsin D	7.11e−47	0.36
Cathepsin D	2.05e−42	0.37
Cholecystokinin A receptor	2.01e−30	0.35
Cholecystokinin B receptor	3.44e−27	0.35
Nicastrin	1.87e−26	0.39
Presenilin 2, 1	1.87e−26	0.39
Gamma-secretase subunit APH-1A	4.42e−24	0.39
Leukocyte elastase	1.42e−2	0.35
Bradykinin B_1 receptor	2.41e−1	0.32
Compound **2**		
Oligopeptide transporter small intestine isoform	4.39e−55	0.33
HIV type 1 protease	8.96e−31	0.37
Angiotensin-converting enzyme	1.51e−26	0.32
Matrix metalloproteinase 7	4.32e−18	0.36
Angiotensin-converting enzyme	6.66e−17	0.32
Matrix metalloproteinase 3	4.06e−16	0.36
Neprilysin	7.40e−16	0.33
Matrix metalloproteinase 13	7.41e−16	0.36
Matrix metalloproteinase-1	1.73e−14	0.36
Integrin alpha-4	9.86e−14	0.4
BACE-1	1.17e−04	0.36
Bradykinin B_1 receptor	6.10e+00	0.32

Overall, 104 targets were predicted for compound **1** and 75 targets for compound **2**.

Table S2. List of drug targets predicted for amprenavir by SLAP

Target	*P* value	Score
BACE-1	0.0001	876
Renin	0.0009	381
ABC transporter C1	0.0026	232
ABC transporter G2	0.0066	142
Endothelin-converting enzyme 1	0.0068	139
fMet-Leu-Phe receptor	0.0074	134
Cyp 3A4	0.0091	119

Table S3. Top 10 PASS predictions for compounds 1 (of 142 predictions) and 2 (of 277 predictions) according to the ratio *P*(active)/*P*(inactive)

Target	*P*(active)	*P*(inactive)	*P*(active)/*P*(inactive)
Compound **1**			
Integrin α2β1 antagonist	0.215	0.002	107.50
Integrin α2 antagonist	0.271	0.004	67.75
Antiviral (HIV)	0.271	0.012	22.58
Centromere associated protein inhibitor	0.491	0.032	15.34
Chymotrypsin inhibitor	0.183	0.012	15.25
Secretase inhibitor	0.111	0.008	13.88
Antiviral	0.338	0.025	13.52
HIV-1 protease inhibitor	0.066	0.005	13.20
Cell adhesion inhibitor	0.18	0.015	12.00
CYP2C19 inducer	0.395	0.043	9.19
Compound **2**			
Omptin inhibitor	0.745	0.014	53.21
Integrin α_2 antagonist	0.155	0.005	31.00
Nardilysin inhibitor	0.419	0.014	29.93
PfA-M1 aminopeptidase inhibitor	0.432	0.018	24.00
Tripeptide aminopeptidase inhibitor	0.366	0.016	22.88
Glutamyl endopeptidase II inhibitor	0.674	0.033	20.42
Integrin $\alpha_2\beta_1$ antagonist	0.104	0.007	14.86
Lactase inhibitor	0.462	0.037	12.49
Viral entry inhibitor	0.263	0.022	11.95
Phosphatase inhibitor	0.595	0.05	11.90

Table S4. Top 5 (of 14) SuperPred target predictions for compound 1 and the top 3 (of 15) SuperPred predictions for compound 2

Target	Reference compound	Score
Compound **1**		
HIV-1 protease	Amprenavir	65.9
HIV-1 protease	AIDS-031900	64.1
HIV-1 protease	AIDS-073036	63.4
No target found	Fosamprenavir	59.6
HIV-1 protease	Fosamprenavir sodium	59.3
Compound **2**		
HIV-1 protease	Amprenavir	56.4
HIV-1 protease	AIDS-031900	54.8
HIV-1 protease	AIDS-073036	54.6

Table S5. Cross-validation results for different combination functions of the SOM-based prediction of drug equivalence relationships (SPiDER) model: Percentage of molecules that had their complete target profile predicted with significant confidence, area under the ROC curve, and average number of significantly predicted compounds

Combination	Percentage correctly predicted molecules	ROC AUC	Number of significant predictions
CATS-only	41.0 ± 0.7	0.879	10.9
MOE-only	41.3 ± 0.5	0.879	10.8
Minimum	65.2 ± 0.4	0.868	10.9
Maximum	41.0 ± 0.6	0.905	10.8
Arithmetic mean	65.3 ± 0.4	0.927	10.9
Geometric mean	65.3 ± 0.4	0.868	10.9