Design and Validation of FRESH, a Drug Discovery Paradigm Resting on Robust Chemical Synthesis.

Qi Shi,[†] Thomas M. Kaiser,[†] Zackery W. Dentmon,[†] Mariangela Ceruso,[§] Daniela Vullo,[§] Claudiu T. Supuran[§] and James P. Snyder[†,*]

[†]Department of Chemistry, Emory University, Atlanta, Georgia 30322, United States;
[§]Neurofarba Dept., Sezione di Scienze Farmaceutiche and Laboratorio di Chimica Bioinorganica, Universita degli Studi di Firenze, Florence, Italy

FRESH Case Study ROC Curves

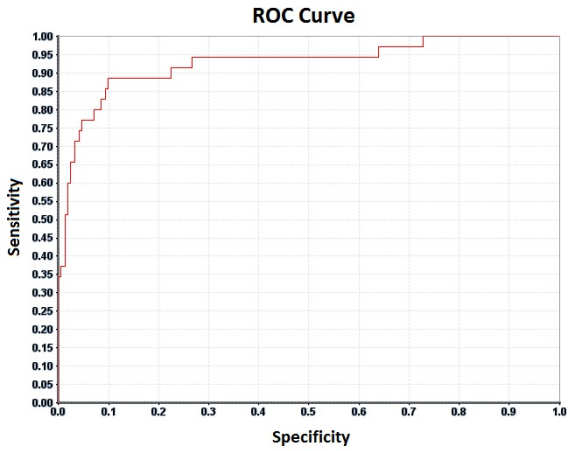


Figure S1. ROC curve for the ECFP method in Case 1(PI3K). AUC = 0.93

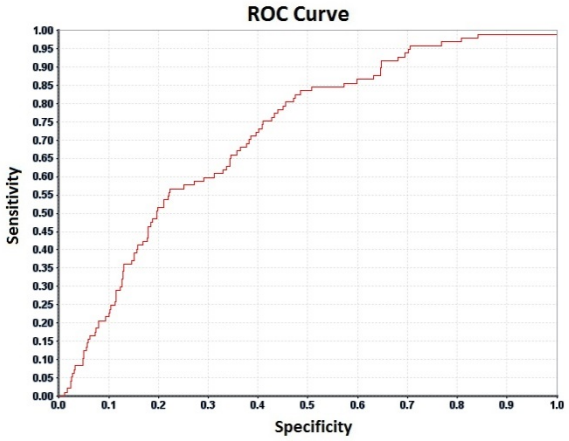


Figure S2. ROC curve for the MM-GBSA score in Case 1 (PI3K). AUC = 0.72

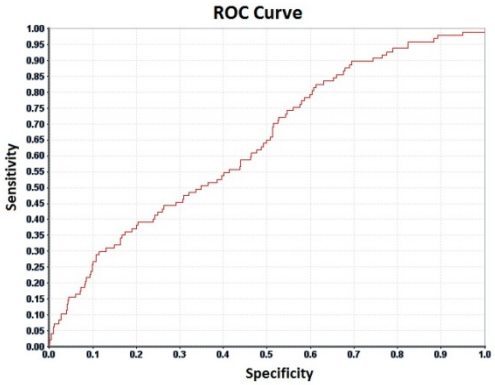


Figure S3. ROC curve for the Glide Score in Case 1 (PI3K). AUC = 0.64

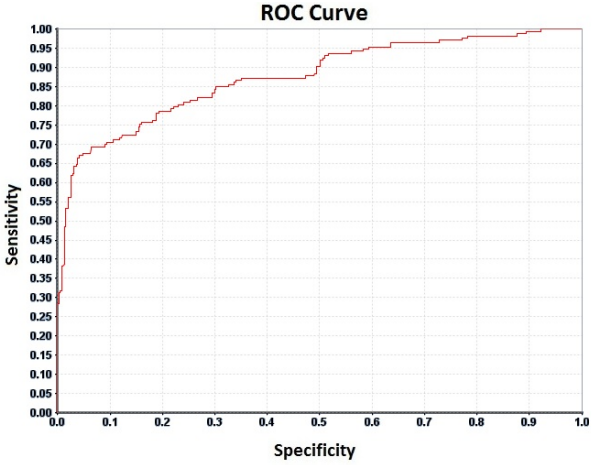


Figure S4. ROC curve for the ECFP method in Case 2 (CA II). AUC = 0.88

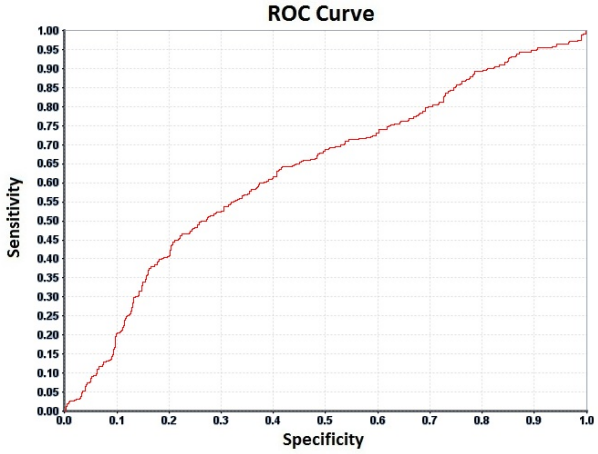


Figure S5. ROC curve for the Glide score in Case 2 (CA II). AUC = 0.63

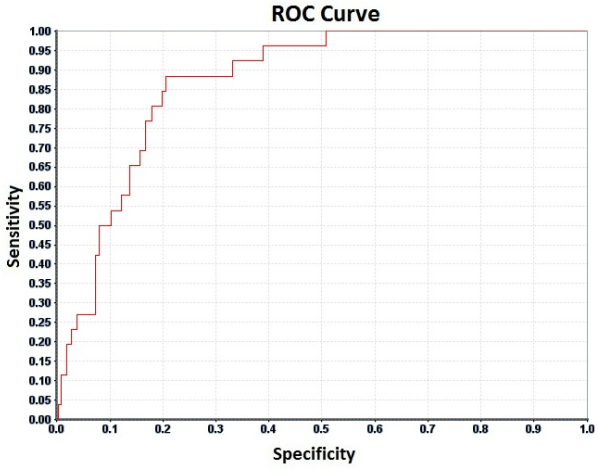


Figure S6. ROC curve for the ECFP method in Case 3 (HDAC1), AUC = 0.87

Filter Details

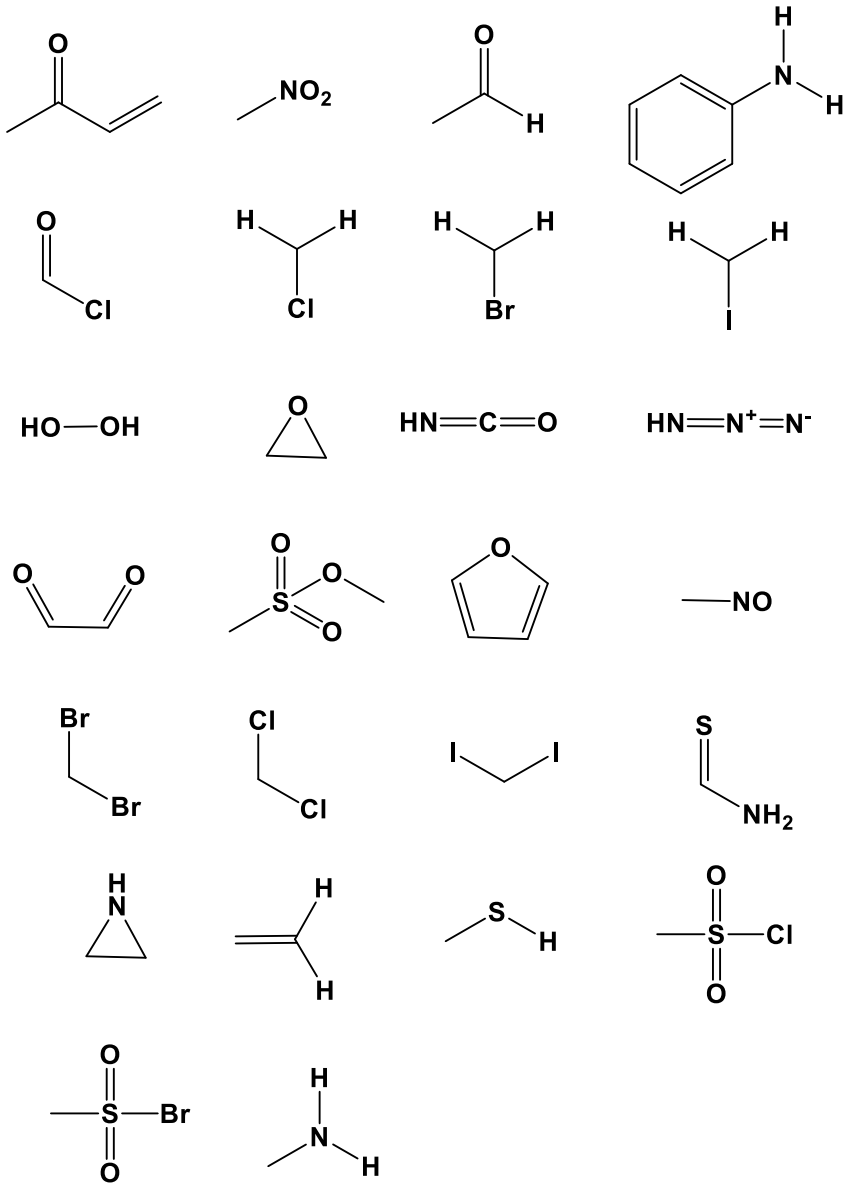


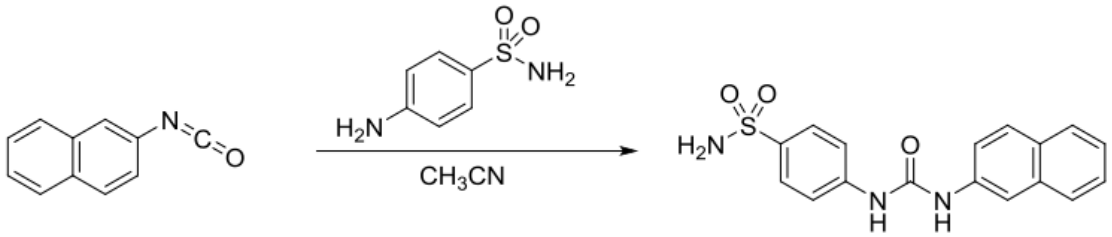
Figure S7. Substructures with potential liability, stability and reactivity concerns.

The general fragment selection filter is expressed by the following PilotScript language: "molecular_weight < 301 and (N_Count + O_Count) <= 3 and Num_H_Donors <= 3 and Num_positiveatoms == 0 and Num_negativeatoms == 0 and alogP <= 3;", which implements the Fragment Rules of Three and requires no permanent charges. The building blocks containing bridghead atoms and spiro atoms are usually costly and frequently require "made on request", so additional filters are added by "Num_BridgeHeadAtoms == 0 and Num_SpiroAtoms == 0;".

The general compound selection filter is expressed by the following PilotScript language: “Molecular_weight < 501 and (N_Count + O_Count) <= 10 and Num_H_Donors <= 5 and AlogP <= 5 and Num_rotatablebonds <= 10 and Molecular_solubility >= -6 and i_qp_qplogPotow <= 5 and r_qp_PSA <= 120 and i_qp_nummetab <= 6 and r_qp_qplogS >= -6;”, which covers the Lipinski Rules of Five, Jorgensen Rules of Three and polar surface area. They are routinely incorporated in FRESH unless one specific term removes all the structures or uses another range required in the project.

Synthesis and Characterization.

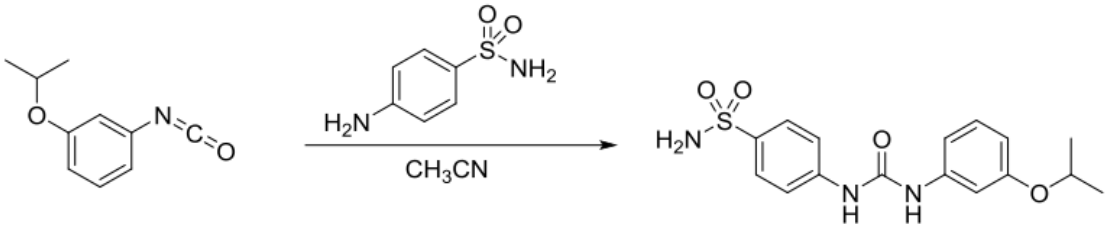
Thin-layer chromatography was done on TLC Silica gel 60 F_{254} commercial plates (Merck KGaA) ^{1}H NMR spectra were recorded on Varian Inova 400 (400 MHz) spectrometer. All spectra are referenced to the residual solvent peak (2.5 ppm for D_6-DMSO). The chemical shift (δ) of each signal is reported in parts per million (ppm) and all coupling constants (*J*) are reported in Hertz (Hz). ^{13}C NMR spectra were recorded on Varian Inova 400 (100 MHz) spectrometer. All spectra are referenced to the residual solvent peak. The chemical shift (δ) of each signal is reported in parts per million (ppm). The Emory University Mass Spectrometry Center recorded high resolution mass spectra. All compounds were analyzed by Agilent 1200 series LCMS or Agilent 1100 series HPLC and found to be greater than 95% pure.



4-(3-(naphthalen-2-yl)ureido)benzenesulfonamide

4-aminobenzenesulfonamide (0.5 g, 2.90 mmol) was dissolved in CH_3CN (25 mL) and 2-isocyanatonaphthalene (0.49 g, 2.9 mmol) was added. The reaction was stirred at r.t. for 19 hours and then the precipitate was filtered off. The precipitate was washed with

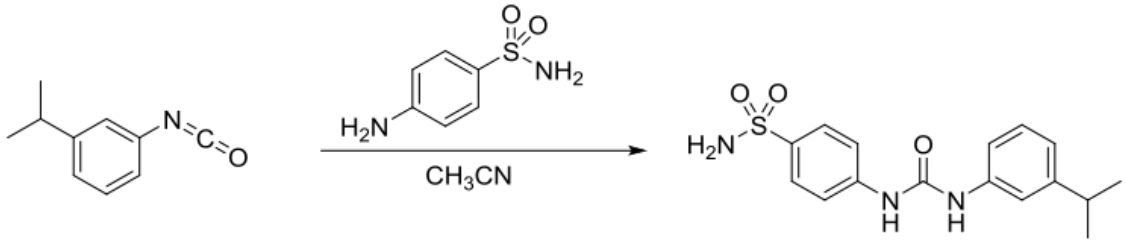
diethyl ether (100 mL) and dried in vacuo to afford 4-(3-(naphthalen-2-yl)ureido)benzenesulfonamide (282 mg, 28% yield, white solid). ^{1}H NMR (400 MHz, D_6DMSO) δ 9.16 (s, 1H), 9.03 (s, 1H), 8.13 (s, 1H), 7.87-7.80 (m, 3H), 7.75 (d, *J* = 8.8 Hz, 2H), 7.66 (d, *J* = 8.8 Hz, 2H), 7.51-7.44 (m, 2H), 7.37 (t, *J* = 7.6 Hz, 1H), 7.23 (s, 2H). ^{13}C NMR (100 MHz, D_6DMSO) δ 152.4, 142.8, 137.0, 136.9, 133.7, 129.3, 128.5, 127.5, 127.1, 126.9, 126.4, 124.2, 119.7, 117.5, 113.8, 109.8. HRMS (EI+) m/z calculated for $C_{17}H_{16}N_3O_3S$ $[M+H]^+$: 342.0907, found: 342.09154.



4-(3-(3-isopropoxyphenyl)ureido)benzenesulfonamide

Triphosgene (0.45 g, 1.52 mmol) and sat. aq. $NaHCO_3$ (20 mL) were added to a stirred solution of 3-isopropoxyaniline (0.674 mL, 4.57 mmol) in DCM (20 mL) at 0 ^{o}C. The reaction was warmed to r.t. and stirred for 2 hours. Water (20 mL) was then added. The aqueous layer was extracted with Et_2O (3x 50 mL) and the organics were dried over $MgSO_4$. The organics were filtered and concentrated and the material was used in the next step without further purification (0.7759 g, >95% yield).

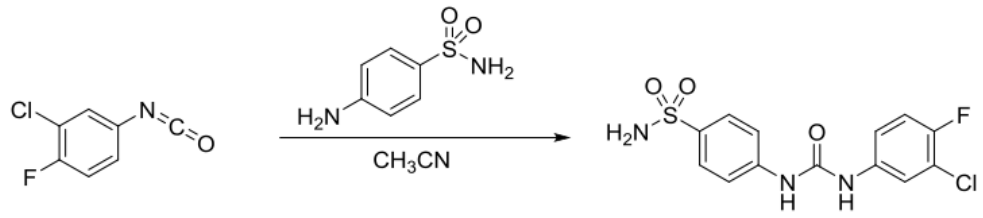
4-aminobenzenesulfonamide (0.716 g, 4.16 mmol) was added to a solution of CH_3CN (36 mL) and 5-isocyanato-2,3-dihydro-1H-indene (0.7366 mL, 4.16 mmol) was added. The reaction was stirred at r.t. overnight and then the precipitate was filtered off. The precipitate was washed with diethyl ether (100 mL) and dried in vacuo to afford the desired compound (97.1 mg, 7% yield, white solid). ^{1}H NMR (400 MHz, D_6DMSO) δ 9.06 (s, 1H), 8.77 (s, 1H), 7.74 (d, *J* = 8.8, 2H), 7.61 (d, *J* = 8.4 Hz, 2H), 7.23 (s, 2H), 7.18-7.14 (m, 2H), 6.90 (d, *J* = 8.0 Hz, 1H), 6.55 (d, *J* = 8.0 Hz, 1H), 4.57-4.50 (m, 1H), 1.27 (d, *J* = 6.0 Hz, 6H). ^{13}C NMR (100 MHz, D_6DMSO) δ 157.9, 152.2, 142.9, 140.6, 136.9, 129.6, 126.9, 117.5, 110.5, 109.4, 105.8, 69.1, 21.9. HRMS (EI+) m/z calculated for $C_{32}H_{39}N_6O_8S_2$ $[M+H+M]^+$: 699.2265, found: 699.22870.



4-(3-(3-isopropylphenyl)ureido)benzenesulfonamide

Triphosgene (0.45 g, 1.52 mmol) and sat. aq. $NaHCO_3$ (20 mL) were added to a stirred solution of 3-isopropylaniline (0.64 mL, 4.57 mmol) in DCM (20 mL) at 0 ^oC. The reaction was warmed to r.t. and stirred for 2 hours. Water (20 mL) was added then added and the aqueous layer was extracted with Et_2O (3x 50 mL) and the organics were dried over $MgSO_4$. The organics were filtered, concentrated and used in the next step without further purification.

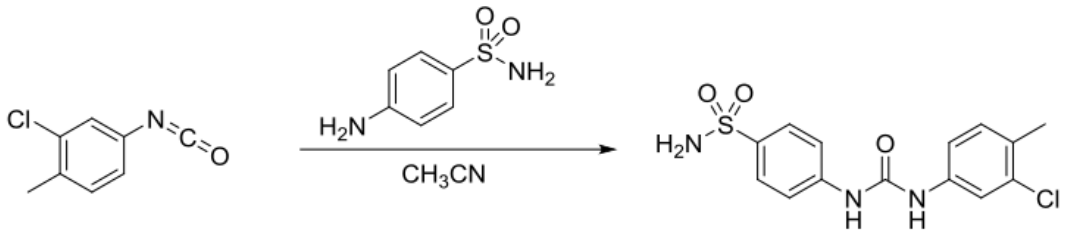
4-aminobenzenesulfonamide (0.716 g, 4.16 mmol) was added to a solution of CH_3CN (36 mL) and 5-isocyanato-2,3-dihydro-1H-indene (0.6516 mL, 4.16 mmol) was added. The reaction was stirred at r.t. overnight and then the precipitate was filtered off. The resulting solid was dried in vacuo to afford the desired compound (0.365 g, 27% yield, white solid). ^1H NMR (400 MHz, D_6DMSO) δ 9.03(s, 1H0), 8.74 (s, 1H), 7.72 (d, J = 8.8 Hz, 2H), 7.61 (d, J = 9.2 Hz, 2H), 7.37 (t, J = 2.0 Hz, 1H), 7.26-7.18 (m, 4H), 6.88 (d, J = 7.2 Hz, 1H), 2.90-2.80 (m, 1H), 1.20 (d, J = 6.8 Hz, 6H). ^{13}C NMR (100 MHz, D_6DMSO) δ 152.7, 149.5, 143.3, 139.7, 137.2, 129.2, 127.2, 120.7, 117.8, 116.8, 116.4, 33.9, 24.3. HRMS (EI+) m/z calculated for $C_{32}H_{39}N_6O_6S_2$ [M+H+M]+: 667.2367, found: 667.24086.



4-(3-(3-chloro-4-fluorophenyl)ureido)benzenesulfonamide (ZD-1-073)

Solid 4-aminobenzenesulfonamide (0.5 g, 2.90 mmol) was dissolved in acetonitrile (25 ml) to give a colorless solution. 2-chloro-1-fluoro-4-isocyanatobenzene (0.362 ml, 2.90 mmol) was drop-wise added with stirring before the solution was heated to 50 °C and

allowed to stir for 96 hours. A white precipitate formed, and the mixture was filtered through a fritted funnel. The white solid was washed with diethyl ether before being dried *in vacuo* to afford the desired compound (647 mg, 65% yield, white solid). ^{1}H NMR (400 MHz, D_6DMSO) δ 9.16 (s, 1H), 9.00 (s, 1H), 7.81 (dd, J = 2.4, 7.2 Hz, 1H), 7.74 (d, J = 8.8 Hz, 2H), 7.61 (d, J = 8.8 Hz, 2H), 7.35-7.30 (m, 2H), 7.22 (s, 2H). ^{13}C NMR (100 MHz, D_6DMSO) d 152.3, 142.6, 137.1, 126.8, 119.8, 118.9, 118.8, 117.7, 117.1, 109.9. HRMS (EI+) m/z calculated for $C_{13}H_{11}ClFN_3O_3S$ $[M+Na]^+$: 366.0094, found: 366.0079.



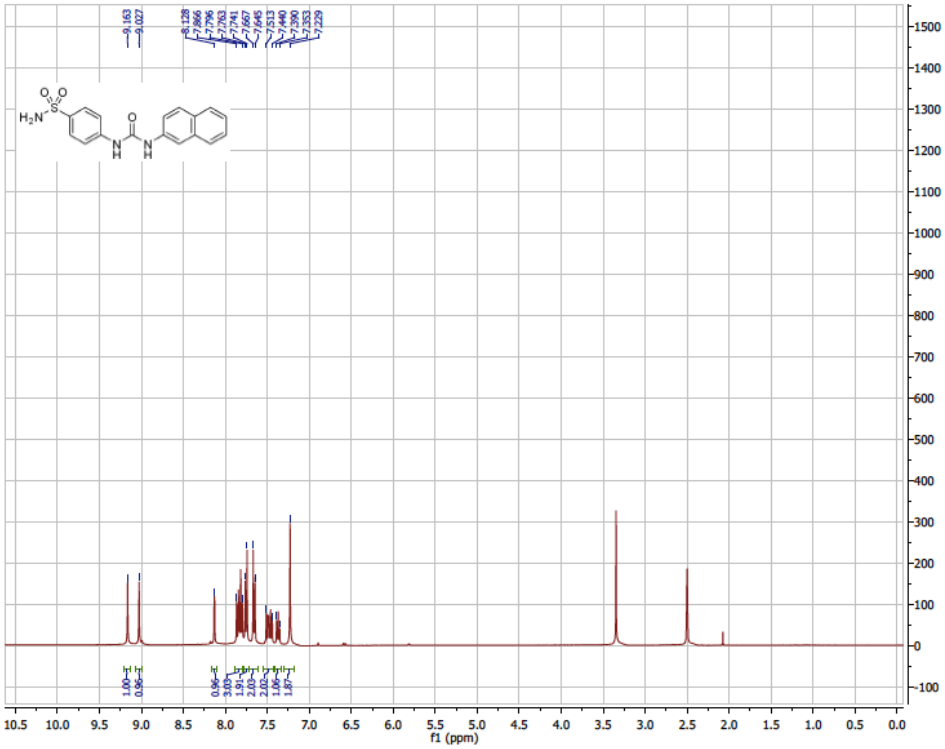
4-(3-(3-chloro-4-methylphenyl)ureido)benzenesulfonamide (ZD-1-078)

Solid 4-aminobenzenesulfonamide (0.5 g, 2.90 mmol) was dissolved in acetonitrile (25 ml) to give a colorless solution. 2-chloro-4-isocyanato-1-methylbenzene (0.398 ml, 2.90 mmol) was drop-wise added with stirring before the solution was heated to 50 °C. After

~1.5 hours, a white precipitate formed. The white precipitate was filtered with a fritted funnel, washed with diethyl ether, and dried in vacuo to afford the desired product (103 mg, 10% yield, white solid). ^{1}H NMR (400 MHz, D_6DMSO) δ 9.10 (s, 1H), 8.89 (s, 1H), 7.73 (d, J = 8.4 Hz, 2H), 7.69 (s, 1H), 7.60, (d, J = 8.4 Hz, 2H), 7.26-7.19 (m, 4H), 2.26 (s, 3H). ^{13}C NMR (100 MHz, D_6DMSO) d 152.3, 142.7, 138.5, 137.0, 133.2, 131.3, 128.8, 126.9, 118.4, 117.7, 117.3, 18.9. HRMS (EI+) m/z calculated for $C_{14}H_{14}ClN_3O_3S$ $[M+Na]^+$: 362.0344, found: 362.0338.

^{1}H and ^{13}C NMR Spectra

*Table 1 main text, Compound **7***



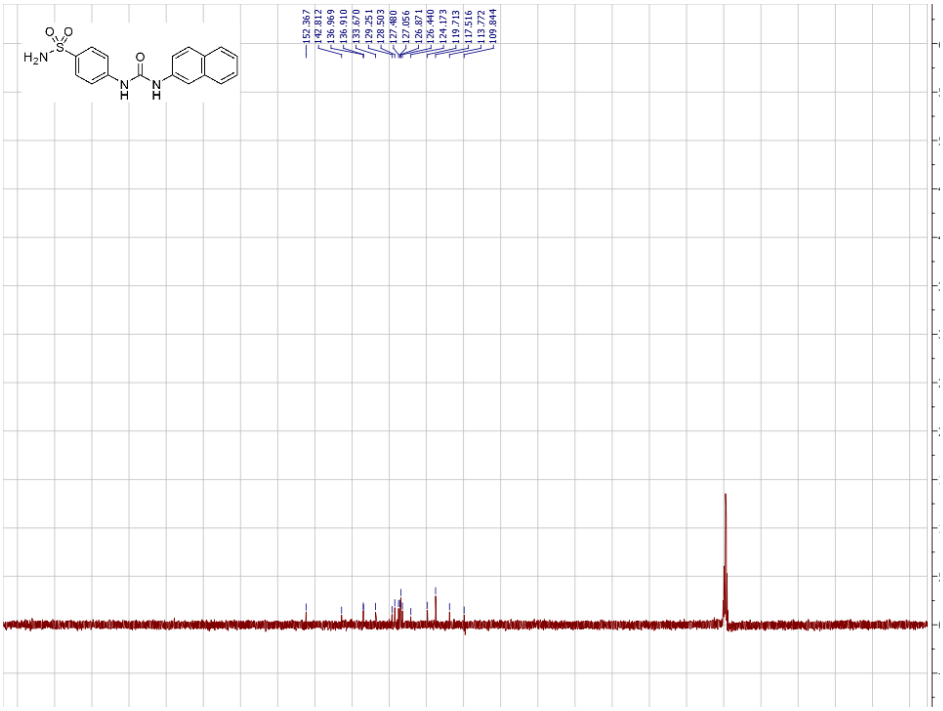
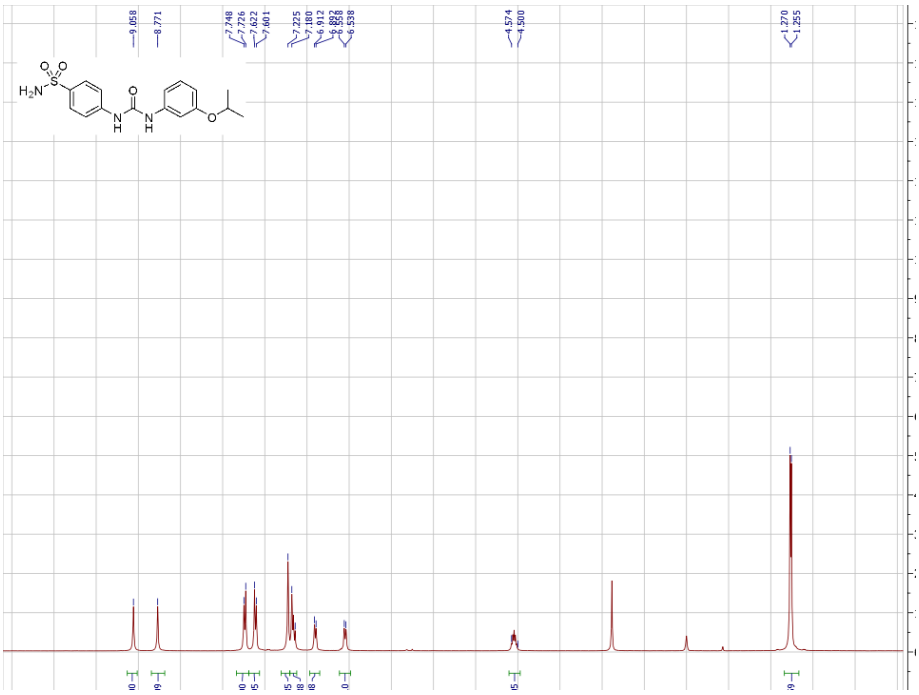


*Table 1 main text, Compound **8***



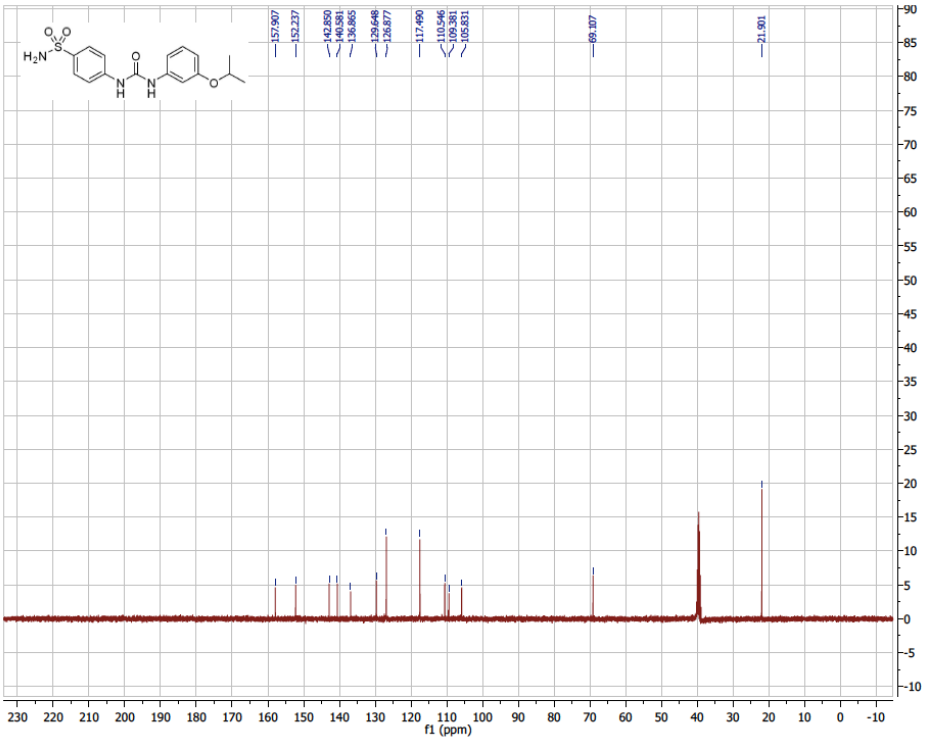
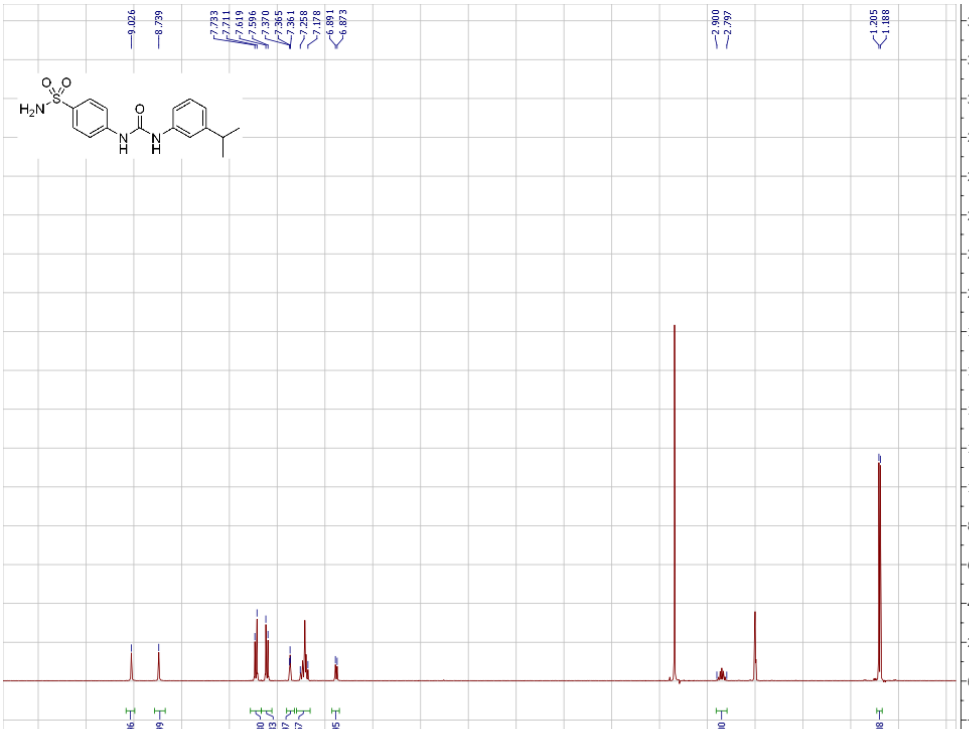


*Table 1 main text, Compound **9***



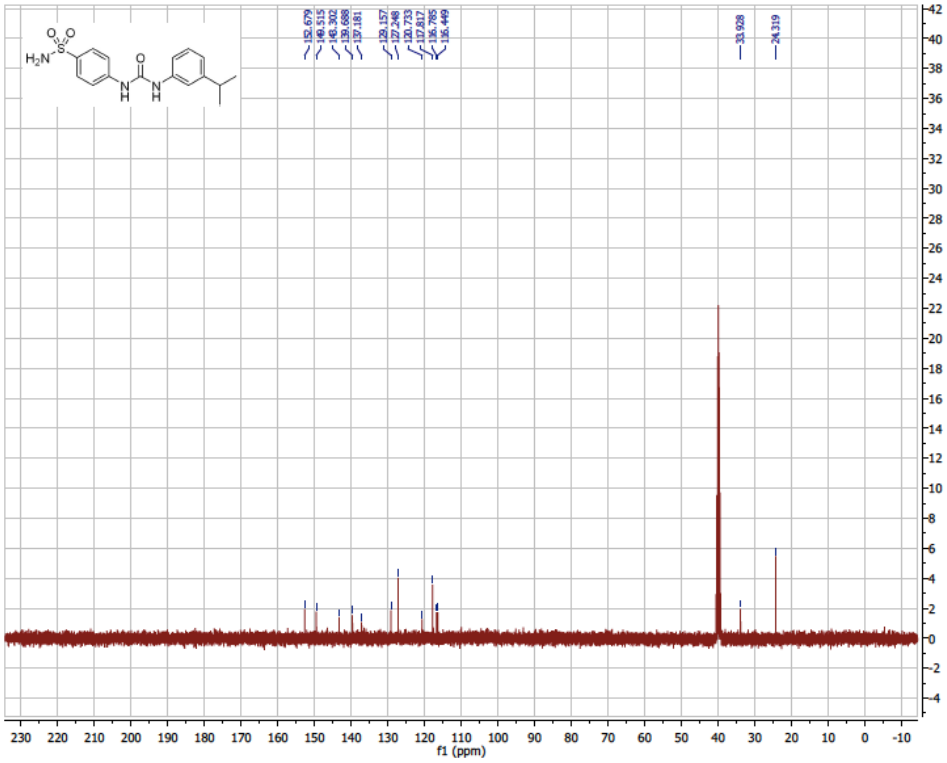
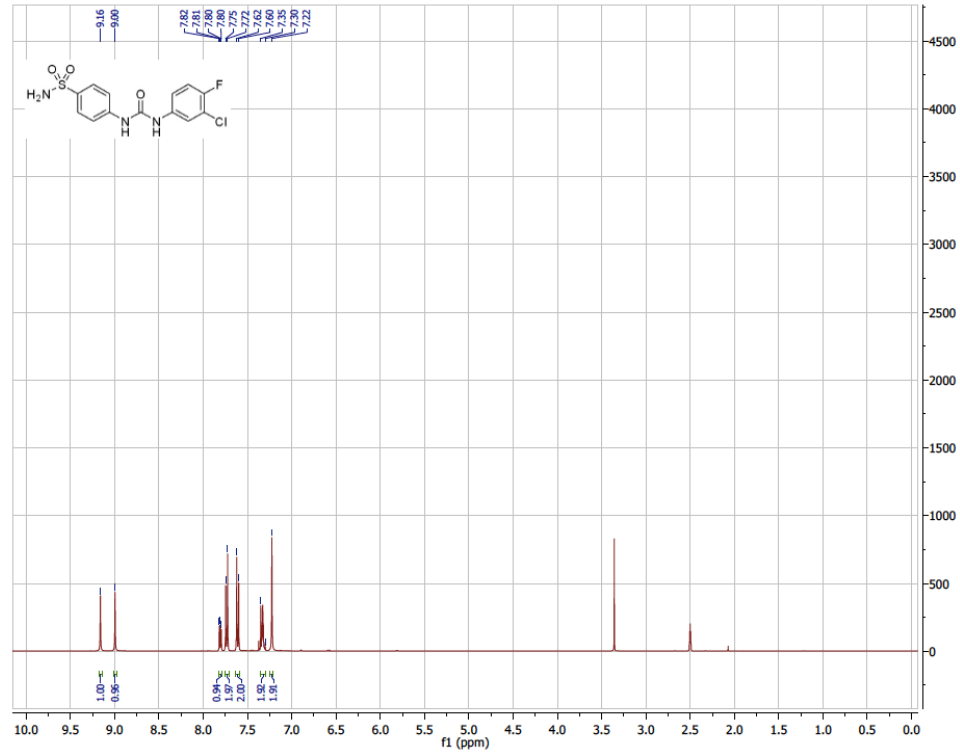


*Table 1 main text, Compound **10***



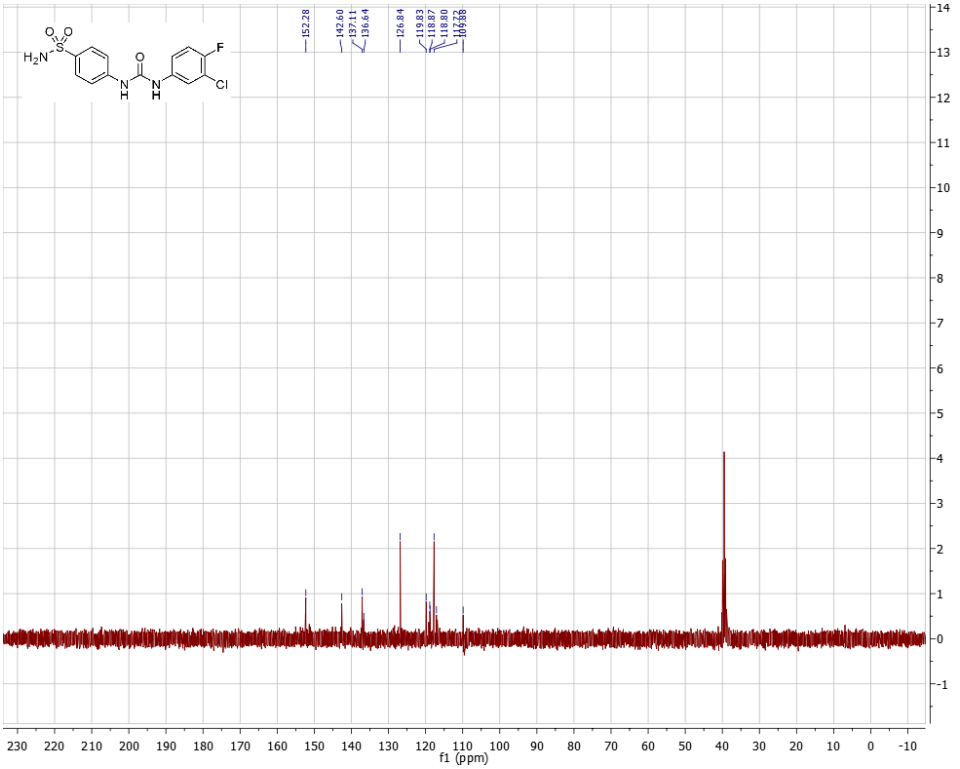
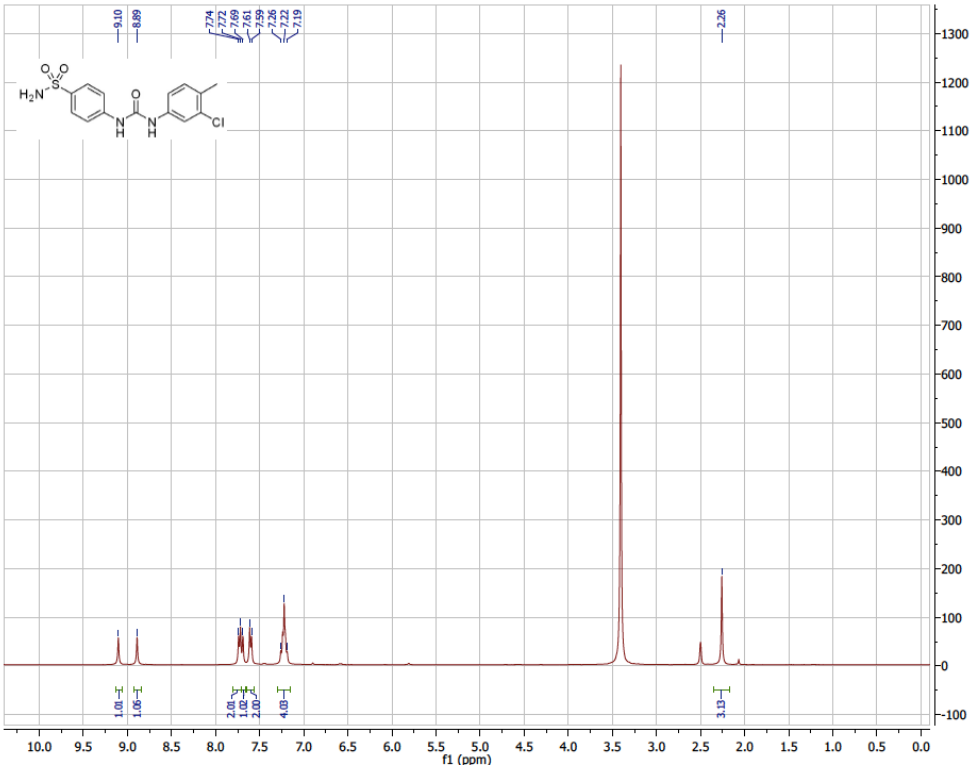
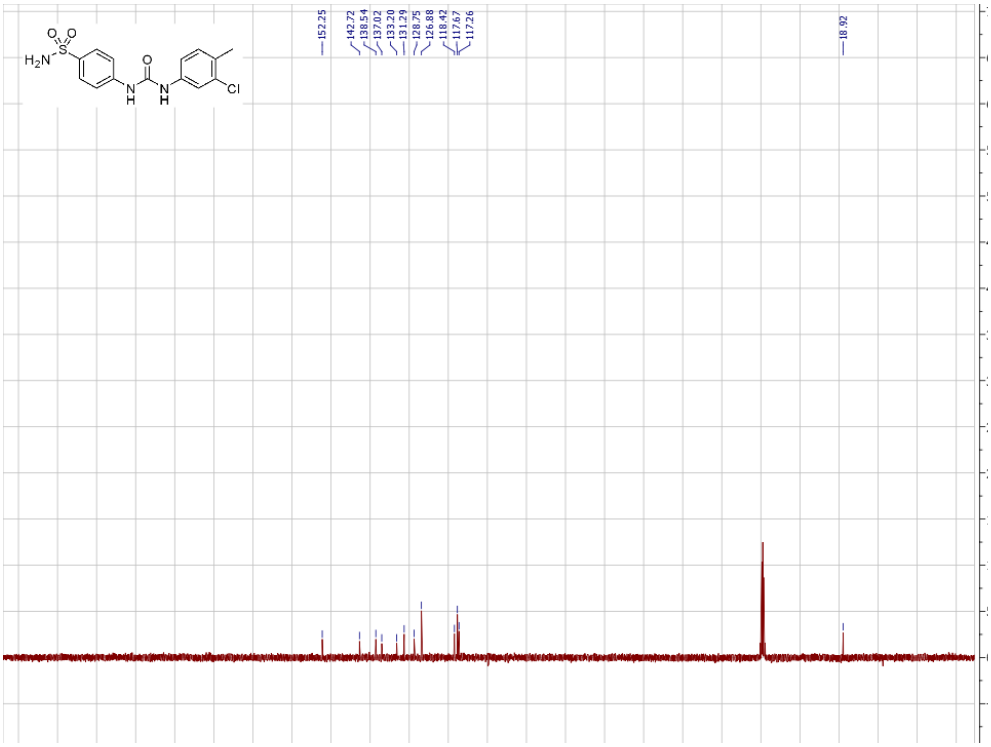


*Table 1 main text, Compound **11***





Computational Details

Methods

Compound database for QSAR Generation and Validation. Data points for historical compounds which serve as the training and test sets for the QSAR in the FRESH program were collected from the ChEMBL database for each specific protein target.[1] Chiral molecules were excluded from the present analyses because of uncertainties in stereochemical assignments, the inability of ECFP descriptors to differentiate chirality and the synthesis challenges. To make the retro-studies more realistic, two filters were used to exclude all compounds with either the same structure or those that appeared after the publication year. For the Bayesian evaluations, 10 nM was chosen as a cutoff value to differentiate "Active" and "Inactive" analogs. The selected ChEMBL database compounds were divided by a 2:1 ratio for training and test sets. The training set structures were used to build the Bayesian QSAR and the test set analogs were used to generate the ROC plots. For the two receptor-based cases, all compounds were scored by Glide (and for the PI3K case MM-GBSA) to obtain the corresponding ROC plots.

Fragment Preparation. The database used to acquire building blocks for the library enumeration step was "Zinc bb now", a collection from several suppliers and claimed to be building blocks immediately available.[2] The building blocks were queried and the corresponding fragments were filtered by "Fragment Rules of 3" and for functional groups with potential toxic liabilities and reactivity concerns. (See SI Filter Details)

Molecule Selection. A series of drug-like filters based on widely-accepted Rules (Rules of Three and Five; The Rule of Four can be readily implemented if protein-protein blockers are the molecular targets.[3,4,5]) were applied to retain structures with desirable drug-like properties. The scores from Bayesian, Glide (and MM-GBSA for the PI3K case) calculations were used to construct QSAR models for the cases treated herein, but any other activity measure can be readily applied as well. Structures with superior Glide and MM-GBSA scores relative to the reference compound were pooled, and the final list was sorted by the Bayesian scores (the best QSAR score with the highest AUC value). For the prospective study, existing structures in ChEMBL or PubMed databases were excluded to guarantee novelty.

Drug-like Property Calculations. Simple properties like molecular weight and number of H-bond acceptors were calculated directly using Pipeline Pilot. Other properties were obtained by Qikprop in the Schrodinger 2012 Maestro package. The 2D molecule structures were prepared by Ligprep in Maestro.

Receptor-based Glide and MM-GBSA Scores (Calculated independent of FRESH, but large databases of structures are readily imported into FRESH for analysis). For the PI3Ka case study, a homology model using the PI3Kg template (PDB code: 1E8Z) was constructed with the sequence of PI3Ka (uniprot id: P42336[6]; 41% sequence identity and 51% similarity within the kinase domain[7]) and used subsequently for receptor-based docking analysis. The homology model was generated using the Prime – Homology Modeling package and the associated "Structure Prediction Wizard" in Maestro.[8] The

PI3Ka sequence was loaded from the above mentioned uniprot id. The default alignment option and the "model built" method were used for the homology modeling. The heavy atoms of the backbone and residues within a 10 Å radius of the catalytic center of the homology model align well (RMSD ~ 1.2 Å) with a PI3Ka crystal structure (PDB code: 3ZIM[9]) reported following our homology model construction. The resulting protein model was utilized for subsequent docking, QSAR analysis and implementation of the latter into FRESH to give structures with predicted potencies superior to the reference compound. For CA II, the catalytic binding site of the latter crystal structure (PDB code 1LUG[10]) was chosen for docking, score estimation by Glide and MM-GBSA and exploitation to conserve the well-established binding mode between sulfonamide inhibitors and the protein. Thus, constraints were imposed during Glide docking on the catalytic center N-Zn bond (2.5 Å) and an H-bond between the ligand SO and the backbone NH of Thr198 as supported by several crystal structures.

Compound Enumeration and Ranking. For the PI3K case study, initially, ~44,000 compounds were extracted from the ChEMBL database. As stated in the manuscript, these were pre-filtered by various physical/ADMET properties as well as two 3D QSAR scores (Glide & MMGBSA). The latter were used to rank the ligands as active or inactive relative to a reference compound with a known K_i. The final rankings were subsequently obtained as Bayesian scores. If 3D scores are eliminated from this procedure (i.e. removal of 2 filters), then the 3rd best PI3K structure (p3 text, left column) drops to the 4th best structure.

For the CA II the study, initially, ~1,500 compounds available from ChEMBL were pre-filtered by Glide scores and physical/ADMET properties, and the final rankings were also obtained as Bayesian scores. If the 3D contributions are removed, the 4th ranked structure (p3 text, right column) falls in rank by 6 slots. In addition, one compound with known activity (3,5-dimethyl substitution, K_i = 1.8 uM) appears on the top 10 list as a false-positive, well above the nM hot list. In sum, in the Bayesian framework, the 3D scores are essential for proper ranking of highly potent compounds.

For the HDAC case, initially, ~8,400 available compounds were filtered by physical/ADMET properties, and the final rankings were also obtained by Bayesian scores.

References

(1) Gaulton A.; Bellis L. J.; Bento A. P.; Chambers J.; Davies M.; Hersey A.; Light Y.; McGlinchey S.; Michalovich D.; Al-Lazikani B.; Overington J. P. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, 1100–1107.

(2) Irwin J. J.; Sterling T.; Mysinger M. M.; Bolstad E. S.; Coleman R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.*, **2012**, *52*, 1757–1768.

(3) Lipinski C. A.; Lombardo F.;Dominy B. W.; Feeney P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3–25.

(4) Jorgensen W. L. Efficient drug lead discovery and optimization *Acc. Chem. Res.* **2009**, *42*, 724-733.

(5) Morelli, X.; Bourgeas, R.; Roche P. Chemical and structural lessons from recent successes in protein–protein interaction inhibition (2P2I). *Curr. Opin. Chem. Biol.* **2011**, *15,* 1-7

(6) http://www.uniprot.org/uniprot/P42336

(7) Walker E. H.; Pacold M. E.; Perisic O.; Stephens L.; Hawkins P. T.; Wymann M. P.; Williams R. L. Structural determinants of phosphoinositide 3-kinase inhibition by wortmannin, LY294002, quercetin, myricetin, and staurosporine. *Mol. Cell.* **2000**, *6*, 909-919.

(8) http://www.schrodinger.com/Maestro; last accessed 12/11/2014.

(9) Nacht M.; Qiao L.; Sheets M. P.; St Martin T.; Labenski M.; Mazdiyasni H.; Karp R.; Zhu Z.; Chaturvedi P.; Bhavsar D.; Niu D.; Westlin W.; Petter R. C.; Medikonda A. P.; Singh J. Discovery of a potent and isoform-selective targeted covalent inhibitor of the lipid kinase PI3Kα. *J. Med. Chem.* **2013**, *56*, 712-721.

(10) Behnke C.A.; Le Trong I.; Godden J.W.; Merritt E.A.; Teller D.C.; Bajorath J.; Stenkamp R.E. Atomic resolution studies of carbonic anhydrase II. *Acta Crystallogr.* **2010**, *66*, 616-627.