UCSF

UC San Francisco Previously Published Works

Title

Ultra-large library docking for discovering new chemotypes

Authors

Lyu, Jiankun
Wang, Sheng
Balius, Trent E
et al.

Peer reviewed

Ultra-large library docking for discovering new chemotypes

Jiankun Lyu[1,2,†], **Sheng Wang**[3,4,†], **Trent E. Balius**[1,†], **Isha Singh**[1,†], **Anat Levit**[1], **Yurii S. Moroz**[5a,5b], **Matthew J. O'Meara**[1], **Tao Che**[4], **Enkhjargal Algaa**[1], **Kateryna Tolmachova**[5c], **Andrey A. Tolmachev**[5c], **Brian K. Shoichet**[1,*], **Bryan L. Roth**[4,6,7,*], and **John J. Irwin**[1,*]

[1]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94158, USA

[2]State Key Laboratory of Bioreactor Engineering, Shanghai Key Laboratory of New Drug Design, School of Pharmacy, East China University of Science & Technology, Shanghai 200237, China

[3]State Key Laboratory of Molecular Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences; University of Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China

[4]Department of Pharmacology, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599-7365, USA

[5a]National Taras Shevchenko University of Kyiv, 60 Volodymyrska Street, Kyiv 01601, Ukraine

[5b]Chemspace, ilukstes iela 38-5, Riga, LV-1082, Latvia

*Corresponding authors: Correspondence and requests for materials should be addressed to Brian Shoichet (bshoichet@gmail.com), Bryan Roth (zendudest@gmail.com), John Irwin (jir322@gmail.com).

Code Availability:
DOCK3.7 is freely available for non-commercial research http://dock.compbio.ucsf.edu/DOCK3.7/. A web-based version is freely available to all at http://blaster.docking.org/.

Data Availability Statement:
Structures of AmpC β-lactamase determined with the new docking hits are available from the Protein Databank with accession numbers 6DPZ, 6DPY, 6DPX and 6DPT. The compounds docked in this study are freely available from our ZINC database, http://zinc15.docking.org. All active compounds are available either from the authors or may be purchased from Enamine. Figures with associated raw data include: Fig. 2, electron density and reflection files are deposited with the PDB; Fig. 3, underlying activities are uploaded in a excel file; Fig. 4, underlying numbers are uploaded in a excel file; Extended Data Fig. 1, underlying numbers are supplied in Supplementary Table 1; Extended Data Fig. 5, underlying activities are uploaded in a excel file; Extended Data Fig. 6, raw clustering/no-clustering rank numbers are uploaded as Supplementary Table 8 and 9. Further underlying data are provided in Supplementary Tables 3 and 5 (aggregation assays for AmpC inhibitors and D4 ligands); Extended Data Table 1 (crystallographic data collection & refinement); Supplementary Tables 9–10 and Supplementary Data 12–15 (chemical purity of active ligands, and their spectra); Supplementary Data 11 and 14 (synthetic routes to compounds).

Data availability:
All data used in the preparation of this manuscript are available as follows: Four crystal structures with PDB codes: 6DPZ, 6DPY, 6DPX and 6DPT; Prism files used in the preparation of curves are in the supporting information; All other data are available from the authors on request.

[5c]Enamine Ltd., 78 Chervonotkatska Street, Kyiv 02094, Ukraine

[6]Division of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7360, USA

[7]National Institute of Mental Health Psychoactive Drug Screening Program (NIMH PDSP), School of Medicine, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599-7365, USA

Abstract

Despite intense interest in expanding chemical space, libraries of hundreds-of-millions to billions of diverse molecules have remained inaccessible. Here, we investigate structure-based docking of 170 million make-on-demand compounds from 130 well-characterized reactions. The resulting library is diverse, representing over 10.7 million scaffolds otherwise unavailable. The library was docked against AmpC β-lactamase and the D_4 dopamine receptor. From the top-ranking molecules, 44 and 549 were synthesized and tested, respectively. This revealed an unprecedented phenolate inhibitor of AmpC, which was optimized to 77 nM, the most potent non-covalent AmpC inhibitor known. Crystal structures of this and other new AmpC inhibitors confirmed the docking predictions. Against D_4, hit rates fell monotonically with docking score, and a hit-rate vs. score curve predicted 453,000 D_4 ligands in the library. Of 81 new chemotypes discovered, 30 were sub-micromolar, including a 180 pM sub-type selective agonist.

In a famous footnote, Bohacek and colleagues suggested that there are over 10^{63} drug-like molecules[1]. This is too many to even enumerate, and other estimates of drug-like chemical space have been proposed[2–4]. What is clear is that the number of possible drug-like molecules is many orders-of-magnitude higher than exists in early discovery libraries, and that this number grows exponentially with molecular size[3]. As most optimized chemical probes and drug candidates resemble the initial discovery hit[5], there is much interest in expanding the number of molecules and chemotypes that can be explored in early screening.

Expanding chemical space

An early effort to enlarge chemical libraries focused on the enumeration of side chains from central scaffolds. Though such combinatorial libraries can be very large, efforts to produce and test them often foundered on problems of synthesis, assay artifacts[6], and lack of diversity. More recently, a related strategy using DNA encoded libraries (DELs)[7] has overcome many of these deficits[8]. Still, most DEL libraries are limited to several reaction types or core scaffolds[9], reducing diversity.

In principle, structure-based docking can screen virtual libraries of great size and diversity, selecting only the best fitting molecules for synthesis and testing. These advantages are balanced by grave deficits: docking cannot calculate affinity accurately[10], and the technique has many false-positives. Accordingly, docking of readily-available molecules is crucial. For virtual molecules, such accessibility has been problematic. Worse still, a large library screen could exacerbate latent docking problems, giving rise to new false positives. Thus, while

docking screens of several million molecules have found potent ligands for multiple targets[11–22], docking much larger virtual libraries has remained largely speculative.

To overcome the problem of compound availability in a make-on-demand library, we focused on molecules from 130 well-characterized reactions using 70,000 building blocks from Enamine (Fig. 1). The resulting reaction products are often functionally congested—displaying multiple groups from a compact scaffold—with substantial 3-dimensionality; less than 3% are commercially available from another source. Addition of new reactions and building blocks has steadily grown the library (Fig. 1a). As of this writing there are over 350 million make-on-demand molecules in ZINC (http://zinc15.docking.org) in the lead-like range[23] (i.e., MWT 350, cLogP 3.5). Over 1.6 billion readily synthesizable molecules have been enumerated, and the dockable library should soon grow beyond 1 billion molecules (Fig. 1b orange bars). Meanwhile, diversity is retained: a new scaffold is added for every ~20 new compounds (Fig. 1c). Naturally, a library of this size is almost entirely virtual.

Even if the make-on-demand molecules are readily accessible, inaccurate scores and a vast chemical space could conspire to overwhelm the true actives with docking decoys. Accordingly, we simulated how hit rates would vary as the library grew from tens-of-thousands to hundreds-of-millions of molecules. In a first approach, we docked tens-to-hundreds of known ligands mixed with thousands of property-matched decoys[24] (Extended Data Fig. 1a & 1b). From the resulting rank distributions, we simulated the effect of varying the ligand-to-decoy ratio in a growing library. Performance was judged by the number of ligands in the top 1000 ranked molecules for any library size, a stringent criterion. When ligands were enriched in the smaller libraries, performance typically improved with library size (Extended Data Fig. 1c). Conversely, when docking performed poorly in small benchmarks, performance often deteriorated with library size.

In a second approach, we investigated ligand enrichment against the full make-on-demand library. We counted known binders as well as their close analogs in the library as ligands; the rest of the library were considered decoys (Methods). For targets with well-formed binding sites, known ligands and ligand-analogs were found among the top docking hits, even from libraries of over 170 million molecules (Extended Data Fig. 1d **and** Supplementary Table 1).

Docking 99 million molecules vs. β-lactamase

Fortified by these simulations, we turned to prospective prediction of new compounds. We targeted two unrelated proteins: the enzyme AmpC β-lactamase and the D_4 dopamine G protein-coupled receptor (GPCR). Against AmpC, we docked the make-on-demand lead-like library, then composed of 99 million molecules. Each was fit in the enzyme active site in an average of 4054 orientations, and for each orientation 280 conformations were sampled. Each configuration was scored for energetic fit, using the physics-based DOCK3.7 scoring function. The top-ranked million molecules were clustered by scaffold[25] and by topological similarity, reducing redundancy. Molecules were excluded that resembled known AmpC inhibitors from ChEMBL[26] (ECFP4 Tanimoto coefficient (Tc) > 0.45) or that resembled any molecule in the 3.5 million in-stock library (ECFP4 Tc > 0.5). Thus, we sought molecules new to the enzyme and new to the planet.

Fifty-one top-ranking molecules—each a different scaffold—were selected for testing, of which 44 (86%) were successfully synthesized (Supplementary Table 9, Supplementary Data 11 and 12). Five measurably inhibited AmpC, with K_i values ranging from 1.3 to 400 μM (Extended Data Fig. 2 **and** Extended Data Fig. 3), an 11% hit rate. All were selective, competitive inhibitors, did not aggregate, nor did they inhibit counter-screening enzymes like trypsin and chymotrypsin (Supplementary Table 2 and 3). Notably, the 1.3 μM ZINC339204163 engages the crucial oxyanion hole of AmpC with a phenolate. Not only is **'4163** the most potent reversible AmpC inhibitor found from any type of screen, its phenolate warhead is unprecedented for β-lactamases and has few precedents even for other amidases and proteases[27]. To optimize the five initial hits, we chose 90 well-scoring analogs from within the make-on-demand library (Methods). Over half were active on testing, improving the affinity of each of the five hits by 3- to 17-fold (Extended Data Fig. 2 **and** Supplementary Table 2). This included the 77 nM ZINC549719643, an analog of the phenolate **'4163**, the most potent non-covalent inhibitor characterized for AmpC. The ability to optimize affinity by finding analogs within the library attests to its depth of coverage for many chemotypes.

Crystal structures of three of the new ligand families, and of the 77 nM **'9643**, were determined to a resolution ranging from 1.50 to 1.91 Å. Unambiguous electron density confirmed their fidelity to the docking predictions, with RMSDs varying from 0.98 to 1.52 Å (Fig. 2, Extended Data Table 1 **and** Extended Data Fig. 4). The RMSD rises to 1.98 Å for ZINC275579920, but this largely reflects a rotation of the terminal ring, which makes no polar interactions with the enzyme in either conformation. For the central core of '**9920**, the RMSD is 1.20 Å and all five hydrogen bonds predicted by docking are found crystallographically (Fig. 2b). Such polar interactions corresponded well between docked and crystallographic poses in all four structures, including that of the phenolate of **'9643**, which forms the three docking-predicted hydrogen bonds with AmpC's oxyanion-hole (Fig. 2e).

Docking 138 million molecules vs. the D$_4$

The prospective screen against the D$_4$ dopamine receptor (D$_4$) had two goals. The first was to see if we could discover new receptor chemotypes, as with most docking campaigns. A second goal was to investigate something largely unexplored in molecular docking: how success varies with docking rank. Accordingly, we tested 549 make-on-demand molecules drawn from not only high-ranking molecules, but also mid- and low-ranked ones (Fig. 3a).

Seeking new chemotypes[28,29], the now over 138 million library molecules were docked against the structure of D$_4$ dopamine receptor[14]. About 70 trillion complexes were sampled in the orthosteric site, requiring 43,563 core hours or about 1.2 calendar days on 1500 cores. Here again, the ranked library was clustered by topology and by scaffold,[25] reducing redundancy. To increase novelty, molecules found in the 3.5 million in-stock library, or that resembled the ~28,000 dopaminergic, serotonergic, or adrenergic ligands in ChEMBL (Tc 0.35 by ECFP4 fingerprints), were excluded. Of the 589 molecules selected, 549 (93%) were successfully synthesized (Supplementary Table 10, Supplementary Data 11 and 13). From the top 1000-ranking clusters, 124 molecules were selected by visual inspection for

favorable and diverse interactions with the D_4 site, and for lack of internal strain[30]; another 444 were selected automatically, by docking score alone, across the rank-ordered list (19 were in both lists). At 10 μM, 122 of the 549 molecules displaced more than 50% [^{3}H]-N-methylspiperone specific D_4 binding (Fig. 3a). Dose response curves for 81 compounds revealed K_i values ranging from 18.4 nM to 8.3 μM (Fig. 3b, Extended Data Table 2 **and** Supplementary Table 4).

Many of the highly ranked molecules were functionally congested, and often docked to interact with residues that are rarely simultaneously engaged (Fig. 3c, Extended Data Fig. 5). Most filled the pocket defined by helix 5 and 6 residues such as S196$^{5.42}$, F410$^{6.51}$ and F411$^{6.52}$, and ion-paired with D115$^{3.32}$, both common interactions among dopaminergic ligands (Fig. 3c, superscripts use Ballesteros-Weinstein and GPCRdb nomenclature[31,32]). Less common among previously known ligands, but frequently observed here, was engagement of the D_4 selectivity pocket, defined by F91$^{2.61}$ and L111$^{3.28}$, which distinguishes this subtype from the D_2 and D_3 dopamine receptors (Fig. 3c). This may explain the 30- to 500-fold subtype selectivity of many of the hits (Extended Data Table 2). Finally, some compounds docked to further hydrogen-bond with backbone atoms in extra-cellular loop 2 (Fig. 3c), which is thought to influence signaling bias[33].

In functional assays, several of the high-ranking molecules were potent. For instance, ZINC621433143 appeared to be a 2.3 nM full agonist (see below), ZINC465129598 and ZINC270269326 were 24 and 17 nM full agonists, respectively, while ZINC464771011 was a 10 nM partial agonist (Fig. 3d and Extended Data Table 2). Two antagonists were also found: ZINC413570733 (IC$_{50}$ 5.9 μM) and ZINC130532671 (IC$_{50}$ 10.8 μM) (Extended Data Table 2 **and** Extended Data Fig. 5). All six lacked detectable activity at the D_2 or D_3 subtypes (Extended Data Table 2). Meanwhile, ZINC615622500 had no detectable G_i activity but was a 3 μM β-arrestin-biased agonist (Extended Data Table 2 **and** Extended Data Fig. 5).

Intriguingly, the potent agonist **'3143** (above) was tested as a diastereomeric mixture. Several of its diastereomers, each independently docked, also scored well—an example is ZINC621433144, which differs from **'3143** by adopting the (3R,4S) rather than the (3S,4S) stereoisomer around the tetrahydropyrrole; the two stereoisomers superpose well in their D_4 docked poses (Fig. 3c). Accordingly, the four diastereomers were independently synthesized and tested. Compound **'3144** is a 180 pM full agonist, with 2,500-fold sub-type selectivity, making it among the most potent, selective full agonists characterized for the D_4 receptor. **'3144** was also functionally selective, with a 17-fold bias towards G_i signaling versus β-arrestin recruitment, versus the characteristic agonist quinpirole (Fig. 3e). Two of the other diastereomers, **'1264** and **'1265** had G_i biases of 26 and 11, respectively, and the third (**'3143**) had a β-arrestin bias of 7 (Extended Data Table 2); here stereoisomerization at a single center flips the bias of a potent agonist.

The make-on-demand library will soon exceed one billion lead-like molecules (Fig. 1b), and it is tempting to dock only cluster representatives, rather than every single molecule. Indeed, doing so reduced docking time 22-fold. Unfortunately, the best cluster representative for a protein is unknowable without docking all cluster members. We found that only docking a

single cluster representative, chosen by multiple criteria (Methods), substantially worsened the docking scores, especially for the best ranked molecules (Fig. 3f and Extended Data Fig. 6a). This had a devastating effect on experimentally-active scaffold families. For instance, the 47 confirmed actives among the top 3000 ranked molecules were replaced with different cluster representatives, and these fell in rank by an average of 1,121,443; only two of the original active scaffolds remained (Supplementary Table 7). Similar effects were observed for β-lactamase (Extended Data Fig. 6b and Supplementary Table 8). Screening the entire library was essential for the discovery of the compounds reported here.

Docking hit rates vary regularly with score

A longstanding question in docking is how well rank predicts binding likelihood. In most docking screens, only tens of molecules are tested, and then only from among the top-ranks. With the great expansion of the library, it seemed interesting to sample also from lower ranks, with enough molecules to be statistically meaningful. Accordingly, we modeled potential "hit-rate" curves as a function of docking score. Using distributions of prior probabilities from Bayesian statistics, we developed ranges of docking scores over which we should test molecules to experimentally define the curve (Fig. 4). From these simulations, the 549 make-on-demand molecules were spread among 12 scoring bins covering the highest-ranking (−75 to −63 kcal/mol), mid-ranking (−61 to −46 kcal/mol), and low-ranking scores where most molecules had unconvincing receptor interactions (−43 to −35 kcal/mol). Typically, 35 to 40 molecules were tested per bin, with more in the highest scoring bins to maximize the number of actives found. Overall, 444 molecules were picked automatically while 124 were picked by visual inspection (above). All molecules were tested *in vitro* using the same protocol.

Intriguingly, hit-rates fell almost monotonically with score, after a plateau defined by the highest-ranking molecules. Among these, hit rates ranged from 22 to 26%, but past scores of −65 kcal/mol they fell steadily, to 12% by a docking score of −54 kcal/mol, and by scores of −43 kcal/mol the hit rate reached zero, where it remained at the next two (worse) scoring bins. We fit a response curve to these observations, with a top hit rate at 24%, a bottom hit rate at 0%, a mid-point at −54 kcal/mol, and a mid-point hit-rate slope of −1.7%/(kcal/mol). The regularity of this curve suggests that, at least for the D_4 dopamine receptor, ligand activity is well-predicted by docking score, notwithstanding a high false-positive rate and an inevitable false negative rate.

From this curve we can model the total number of D_4 actives in the library. Assuming that all molecules in a scoring range have the same hit-rates, we can multiply the total number of library molecules in any such range by the observed hit rate in that range, and sum (Fig. 4). Among the 138 million molecule library there are calculated to be over 453,000 D_4 active molecules, in over 72,600 scaffolds, with estimated K_i values of 10 μM or better (Fig. 4a and 4c). The number of actives drops to 158,000 at a more stringent 1 μM cutoff (Fig. 4b and 4d). Admittedly, these predictions have uncertainties, with 95% confidence ranging from 188,000 to 1,035,000 actives molecules and from 38,000 to 129,000 active scaffolds. Still, in some ways the estimates are conservative—for instance, we assume a 0% rate of compound discovery below a docking score of −40 kcal/mol (Fig. 4a and 4b). Had we assumed a higher

random hit-rate, the number of discoverable compounds would have risen, as most of the library scored worse than −35 kcal/mol (Fig. 4). Finally, we note that this unusually large set of 549 confirmed actives and inactives, all with docking poses[34], may be a useful benchmark for the field (Supplementary Table 4).

Human vs. machine.

We wondered whether molecules prioritized by docking and human visual evaluation would perform better than those prioritized by docking alone. From among the top 1000 ranked molecules, we selected 124 that, on inspection, had favorable interactions, and deprioritized those with strained internal energies (above)[30]. Another 114 high-ranking molecules were selected by docking score alone, from the same ranks. Unexpectedly, the hit rates were about the same at around 24% (Extended Data Fig. 7a). However, the human prioritized molecules typically had better affinities: 44% of these were sub-μM, which was true of only 27% of those prioritized by docking score alone. Correspondingly, a disproportionate number of the most potent agonists, such as the 180 pM **'3144** and the 14 nM **'1011**, were human-prioritized (Extended Data Fig. 7b–c).

The docking results here may be compared to those from earlier high-throughput screening (HTS) and docking campaigns. For AmpC, the direct docking active **'4163** is over 20-times more potent than any earlier non-covalent inhibitor[35–37], and its optimized analog **'9643** is the most potent non-covalent AmpC inhibitor yet characterized. Partly this reflects the simple absence of phenolates from the much smaller libraries previously screened. Similarly, the low and mid-nanomolar agonists **'9598**, **'9326** and **'1011** are 10-fold more potent than any D_4 screening hits of which we are aware, even from campaigns biased toward dopaminergic chemotypes[38], and also more selective. Meanwhile, the 180 pM **'3144** is among the most potent selective agonists reported for this target[39–41]. Comparing this study to a recent docking screen of 600,000 "in-stock" compounds against the D_4 dopamine receptor[14], the initial lead from the smaller library was a 260 nM agonist, and even after three rounds of optimization this series only reached an EC_{50} of 4 nM. Here again, there is no compound topologically similar to **'3144** in the smaller, "in stock" library. It is the great expansion of the make-on-demand library, both in compounds and chemotypes, that has enabled discovery of the new ligands.

Certain caveats bear airing. The variation of hit-rate with docking score, while sigmoidal, was not fully monotonic, with variability among the top-ranking tranches. Naturally, the estimation of actives is valid only for the D_4 dopamine receptor; it has wide error margins (Fig. 4). Whereas molecules were docked as pure stereoisomers and diastereomers, they were often tested as stereochemical mixtures. Finally, the longstanding challenges of sampling and scoring in molecular docking screens remain[42]. Whereas the hit rate vs. docking score curve (Fig. 4) supports an ability to prioritize actives, our raw docking scores remain off-set from true binding energies, and we cannot confidently even rank-order molecules for activity. Finally, docking undoubtedly continues to suffer from false negatives.

These caveats should not obscure our principal observations. First, docking rank predicts the likelihood that a molecule will bind to the D_4 dopamine receptor (Fig. 4). This suggests that

docking methods[43–50], at least for well-formed binding sites, can efficiently prioritize new molecules from a large chemical space. Second, the discovery of novel and potent chemotypes for both targets suggests that the ultra-large libraries contain molecules better suited to a given receptor structure than found within the smaller "in-stock" libraries, and that docking can recognize them. Third, the well-behaved hit-rate vs. score curve (Fig. 4) allows one to predict the total number of expected actives for a target within a library, including those unrelated to known ligands. Integrating under this curve predicts that there are an astonishing 453,000 D_4 ligands in over 72,000 scaffold families in the make-on-demand library. As daunting as these numbers are, we expect them to grow, with the library itself anticipated to exceed one billion lead-like molecules by 2020. This represents a great challenge but also a great opportunity: a 1000-fold expansion of the molecules and chemotypes readily available to chemical biology and to drug discovery, openly accessible to the community (http://zinc15.docking.org).

Online Methods

Database generation.

Dockable ligand databases are downloadable from ZINC (http://zinc15.docking.org), and protonation states and tautomers (Jchem v15.11.23.0, www.chemaxon.com), 3D structures (Corina v3.6.0026, www.molecularnetworks.com), conformational ensembles (omega v2.5.1.4, www.eyesopen.com)[51], atomic charges[52] and desolvation energies[53,54] are calculated as described[55]. For both the AmpC and D_4 campaigns, library molecules were protonated according to experimental testing near neutral pH, using pKa values calculated according to Jchem. Whereas AmpC is known to prefer anionic molecules, and the dopamine receptors are known to prefer cations, there is precedent for uncharged molecules binding to both[56,57]. Accordingly, the full library, unfiltered for charge state except by lead-like characteristics, was docked against both targets. The full list of library molecules docked, by ZINC number, SMILES, and docking scores, are deposited in FigShare[58,59]; from this, full charge and structural representation may be found in http://zinc15.docking.org.

Toy model for database growth.

We constructed a model of ligand enrichment with library size, using the distribution of ligand and decoy docking scores. Except for the D_4 receptor, the ligands and decoys are drawn from the DUD-E benchmark; for the D_4 receptor, 48 ligands were downloaded from IUPHAR (http://www.guidetopharmacology.org) and the corresponding decoys were generated by the DUD-E web server (http://dude.docking.org/generate). Inputs to the model are ligand-to-decoy ratio and number of molecules in databases. From these two parameters, the distributions are sampled. We generated distributions by fitting the skewed-normal distribution to that observed for the DUD-E ligands and decoys from docking, using the statistical library in SciPy (Extended Data Fig. 1a–c and Supplementary Table 1).

Simulating hit rates from full-library docking.

We docked the full make-on-demand library to investigate the ranking of ligands vs. decoys. All known ligands for each target were drawn from ChEMBL[26]. Their analogs in the make-

on-demand library were defined by ECFP4 Tc similarity   0.5, 0.6 or 0.7 for each target (Extended Data Fig. 1d). Together, the known actives and their analogs were defined as ligands while the rest of the docked molecules were defined as decoys. The full library was then docked. To investigate the effect of library size on the ability to enrich "ligands" among the top 1000 ranked compounds, 10^5, $3*10^5$, 10^6, $3*10^6$, 10^7, $3*10^7$ and 10^8 sets of molecules were randomly selected from the full docking-ranked list and the number of ligands among the 1000 was counted. Each set was pulled twenty times with random selection from the larger library.

Bemis-Murcko scaffold analysis.

The SMILES of all the make-on-demand lead-like molecules in ZINC were downloaded from http://zinc15.docking.org/tranches/home/ on February 28, 2018. The program mitools (www.molinspiration.com) calculated scaffolds for all 233 million lead-like molecules using the Bemis and Murcko method.[25]

Large-scale-docking.

The AmpC campaign used the structure in PDB 1L2S, while the D_4 campaign used PDB 5WIU. In each, 45 matching spheres were calculated around and including the ligand atoms —a 26 μM thiophene carboxylate for AmpC and nemonapride for D_4 structures were prepared and AMBER united atom charges assigned[14]. The magnitude of the partial atomic charges for five residues in AmpC were increased without changing the net residue charge[56]. For both targets, the low protein dielectric was extended into the binding site using pseudo-atom positions representing possible ligand docking sites, the radius was 1.0 Å and 2.0 Å for D_4 and AmpC respectively[14,54,60]. For the D_4 dopamine receptor, the desolvation volume of the site was also increased by similar atom positions, using a radius of 0.3 Å. This improved ligand charge-balance in benchmarking calculations, reducing the number of high-ranking dications. Energy grids representing the AMBER van der Waals potential[61], Poisson-Boltzmann electrostatic potentials using QNIFFT[62,63], and ligand desolvation from the occluded volume of the target for different ligand orientations[54] were calculated. Using DOCK3.7.2[64], over 99 million and over 138 million library molecules were docked against AmpC and the D_4 dopamine receptor, respectively. Each library molecule was sampled in about 4054 and 3300 orientations and, on average, 280 and 479 conformations for AmpC and D_4, respectively, and were rigid-body minimized with a simplex minimizer. The throughput averaged 1 second per library compound.

Clustering.

To increase novelty, the high-ranking molecules from both screens were filtered for similarity to previously known ligands, and for similarity to the molecules in the 3.5 million in-stock library (we have deposited tools to do this at https://github.com/docking-org/ChemInfTools). To increase diversity, the docking-ranked molecules were clustered into related families of compounds. For the AmpC screen, the top 1 million ranked molecules were best-first clustered using an ChemAxon ECFP4 Tc of 0.5 for cluster inclusion (using the Tc_c_tool that we have deposited in https://github.com/docking-org/ChemInfTools). For the D_4 screen, we wanted to sample through the docking scoring range, and thus used a hybrid clustering approach to treat many more molecules. To cluster the 53,588,665

molecules with DOCK scores better than −30 kcal/mol against the D_4 dopamine receptor, we used best-first clustering on the first 2 million molecules (DOCK score to −49.38 kcal/mol). This resulted in 126,287 clusters. Bemis-Murcko scaffolds were calculated for the full 53,588,665 molecules, resulting in 4,893,388 scaffold-based clusters. The ECFP4-based clusters and the scaffold-based clusters were combined, and ECFP4 best-score first clustering was run on the best scoring members of each cluster, again using a 0.5 Tc cutoff. This left 423,656 hybrid clusters, each represented by its top-scoring member.

Analysis of full library docking vs. pre-clustering library docking.

The scaffold analysis of all docked molecules against AmpC and against D_4 dopamine receptor used Bemis-Murcko scaffolds, as above. For the full library docking, the best-scoring member was selected to represent the scaffold. To investigate the impact of only docking cluster representatives, rather than docking the full library, scaffold representatives were picked by four different methods: 1) the closest member to the centroid by molecular weight and cLogP; 2) the closest member to the centroid of molecular weight alone; 3) the member with the largest molecular weight and 4) the member with the smallest molecular weight. The molecular weight values are calculated and the cLogP values are predicted by Rdkit (http://www.rdkit.org).

Analoging within the library.

The 90 AmpC analogs from within the make-on-demand library were selected based on topological similarity to the primary docking hits: each had an ECFP4-based Tc values 0.5, or shared the same substructure as the initial hit. All prioritized analogs also had favorable docking scores to the enzyme.

Make-on-demand synthesis.

Compounds were synthesized using 70,000 qualified in stock building blocks and 130 well-characterized, two component reactions at Enamine. Historically, molecules have been synthesized in three to four weeks with an 85% fulfilment rate; in this project delivery time was six weeks, but with a 93% fulfilment rate. Each reaction is well tested for conditions including temperatures, completion time, and mixing, as described[65]. Typically, compounds are made in parallel by combining reagents and solvents in a single vial in the appropriate conditions to allow the reaction to proceed to completion. The product-containing vial is filtered by centrifugation into a second vial to remove precipitate and the solvent is evaporated under reduced pressure; the product is then purified by HPLC. Identity and purity is assessed by LC/MS and, when necessary, ^{1}H NMR. All compounds were shipped 90% pure or better (Supplementary Table 9 and 10, Supplementary Data 11–14).

AmpC β-lactamase crystallography.

All four inhibitors, **'3290**, **'9920**, **'4163**, and **'9643**, were co-crystalized from 1.7 M potassium phosphate with microseeding at pH values that varied for 8.7 to 8.9, as described[66]. Crystals were cryo-cooled in a solution containing a reservoir solution and 25% sucrose. Reflections were measured at beamline 8.3.1 of the Advanced Light Source (Berkeley, CA) with wavelength of 1.11583 nm at a temperature of 100K. Complexes with

'3290, '9920, '4163 and '9643 were measured to a resolution of 1.50 Å, 1.91 Å, 1.90 Å and 1.79 Å, respectively (Extended Data Table 1). All four complexes crystallized in the C2 space group with two molecules in the asymmetric unit[66]. The datasets were processed, scaled, and merged using *XDS* and *AIMLESS*[67]. *MOLREP* was used for molecular replacement using the protein model from PDB 1KE4, giving unbiased electron density for the inhibitor in initial electron density maps. Initial model fitting and water addition was done in *COOT*[68] followed by refinement in *REFMAC*[69]. Geometry restraints of inhibitor molecules were created in *eLBOW*-PHENIX. Following inhibitor modeling in *COOT*, refinement was carried out using *PHENIX*[70]. For each structure, geometry was assessed using MolProbity. The final models of '3290, '9920, '4163 and '9643 in complex with AmpC were refined to Rwork and Rfree values of 19.1 and 22.3%, 19.4 and 23.2%, 17.1 and 20.3%; and 18.6 and 22%, respectively. Coordinates have been deposited with PDB identifiers 6DPZ, 6DPY, 6DPX and 6DPT, respectively. Model quality was confirmed using PROCHECK. The total number of residues located in the most favorable and allowed region of the Ramachandran plot for the complexes with '3290, '9920, '4163, and '9643 were 97.89% and 2.11%, 98.03% and 1.97%, 98.31% and 1.69%, and 98.03% and 1.97%, respectively. The data measurement and refinement statistics are summarized in Extended Data Table 1.

AmpC β-lactamase enzymology.

All candidate inhibitors were dissolved in DMSO at 30 mM, and more dilute DMSO stocks were prepared as necessary so that DMSO concentration was held constant at 1% v/v in 50mM sodium cacodylate buffer, pH 6.5. AmpC activity and inhibition was monitored spectrophotometrically using either CENTA or nitrocephin as substrates[71]. All assays included 0.01% Triton-X-100 to reduce compound aggregation artifacts[72]. Active compounds were further investigated for aggregation by dynamic light scattering and by inhibition of three counter-screening enzymes: trypsin, chymotrypsin, and malate dehydrogenase[37]; unless otherwise stated, no active compound was found to form aggregates nor did they inhibit any of the three counter-screening enzymes (Supplementary Table 2 and 3). IC_{50} values reflect percent inhibition fit to a dose response equation in GraphPad Prism (GraphPad Software Inc.), while K_i values were calculated directly from Lineweaver-Burk plots for all compounds except for '1339, '1516, '0178, and '3290, where the Cheng-Prusoff equation was used.

D_4 Dopamine Receptor radioligand binding assay.

Binding was measured using HEK293T membrane preparations transiently expressing human D_2 (D_2 long receptor), D_3, and D_4 ($D_{4.4}$ variant). HEK293T cells (ATCC CRL-11268; 59587035; mycoplasma free) were transfected and membrane preparation and radioligand binding assays were set up in 96-well plates in the standard binding buffer (50 mM Tris, 10 mM $MgCl_2$, 0.1 mM EDTA, 0.1% BSA, pH 7.4)[14]. For primary screening, 10 μM compounds were incubated with membrane and radioligands (0.8–1.0 nM [³H]-N-methylspiperone) (PerkinElmer). For displacement experiments, test compounds with increasing concentrations were incubated with membrane and radioligands (0.8–1.0 nM [³H]-N-methylspiperone). Reactions, either primary screening or displacement experiments, were incubated for 2 h at room temperature in the dark and terminated by rapid vacuum

filtration onto chilled 0.3% PEI-soaked GF/A filters followed by three quick washes with cold washing buffer (50 mM Tris HCl, pH 7.4) and quantified as described previously[73]. Results (with or without normalization) were analyzed using GraphPad Prism 5.0 using one-site shift models where indicated.

cAMP inhibition assay.

To measure D_4 $G\alpha_{i/o}$-mediated cAMP inhibition, HEK 293T (ATCC CRL-11268; 59587035; mycoplasma free) cells were co-transfected with human D_4 ($D_{4.4}$ variant) along with a luciferase-based cAMP biosensor (GloSensor; Promega) and assays were performed as described[14]. After 16 h, transfected cells were seeded in poly-L-lysine coated 384-well white clear bottom cell culture plates (Greiner; 10,000 cells/well, 40 μL/well) in DMEM containing 1% dialyzed FBS. The next day, ligand solutions were prepared in fresh buffer [20 mM HEPES, 1X HBSS, 0.3% bovine serum album (BSA), pH 7.4] at 3X drug concentration. Plates were decanted and received 20 μl per well of ligand buffer followed by addition of 10 μl of ligand solution (3 wells per condition) for 15 min in the dark at room temperature. To measure agonist activity for $G\alpha_{i/o}$-coupled receptors, 10 μL luciferin (4 mM final concentration) supplemented with isoproterenol (400 nM final concentration was added to activate G_s via endogenous β_2-adrenergic receptors) and luminescence intensity was quantified 10 min later. Data were analyzed using "log(agonist) vs. response" in GraphPad Prism 5.0.

Bioluminescence Resonance Energy Transfer (BRET) assay—To measure D_4-mediated G protein activation, HEK293T cells were co-transfected with human D_4, $G\alpha i1$ containing C-terminal *Renilla* luciferase (*R*Luc8), $G\beta$ and $G\gamma$ containing a C-terminal GFP (at mass ratio 1 : 0.3 : 2 : 2, respectively). To measure D_4-mediated arrestin recruitment, HEK293T cells were co-transfected with human D_4 containing C-terminal *Renilla* luciferase (*R*Luc8), and β-arrestin2 containing a N-terminal YFP at ratio 1:3. After at least 16 hours, transfected cells were plated in poly-lysine coated 96-well white clear bottom cell culture plates in plating media (DMEM + 1% dialyzed FBS) at a density of 40,000–50,000 cells in 200 μl per well and incubated overnight. The next day, media was decanted and cells were washed twice with 60 μL of drug buffer (20 mM HEPES, 1X HBSS, pH 7.4), then 60 μL of the *R*Luc substrate, coelenterazine 400a for G protein assay, and coelenterazine h for β-arrestin2 assay (Promega, 5 μM final concentration in drug buffer), was added per well, incubated an additional 5 minutes to allow for substrate diffusion. Afterwards, 30 μL of drug (3X) in drug buffer (20 mM HEPES, 1X HBSS, 0.1% BSA, pH 7.4) was added per well and incubated for another 5 minutes. Plates were immediately read for luminescence at 400 nm and GFP fluorescent emission at 515 nm (G protein assay); 485 nm and fluorescent eYFP emission at 530 nm (β-arrestin2 assay) for 1 second per well using a Mithras LB940 multimode microplate reader. The ratio of GFP/*R*Luc or eYFP/*R*Luc was calculated per well and the net BRET ratio was calculated by subtracting the GFP/*R*Luc or eYFP/*R*Luc from the same ratio in wells without GFP or eYFP present. The net BRET ratio was plotted as a function of drug concentration using Graphpad Prism 5 (Graphpad Software Inc., San Diego, CA).

Hit-rate curve prediction and estimation of maximum number of hits.

To define the docking scoring ranges from which molecules would be picked for experimental testing, we used distributions of prior probabilities from Bayesian statistics for highest, mid-point, and random hit-rates, and for the slope of the curve. To advance the argument, we conjectured that docking hit-rates would behave like a dose response curve as a function of docking energy, e_i:

$$hitpercent(e_i) = \frac{Top - Bottom}{1 - e^{-S(e_i - Dock_{50})}} + Bottom$$

This function is defined by four parameters: (1) Top is the maximum hit-percent; (2) $Dock_{50}$ is the dock energy in kcal/mol at $Top/2$; (3) $S = Slope * 4 / Top$ where slope is the change in hit-percent at $Dock_{50}$ in hit-percent/(kcal/mol); and (4) $Bottom$ is the minimum hit-rate that we fixed at zero. To define the prior probability distribution, four authors graded 440 compounds across 11 energy slices (Extended Data Fig. 8e), from which we chose independent Bayesian prior probabilities for each parameter, $p(Top) = Beta(\alpha = 20, \beta = 80)$, $p(Dock_{50}) = Normal(\mu = -60, \sigma = 15)$, $p(S) = Normal(\mu = -0.2, \sigma = 0.1)$ (Extended Data Fig. 8b–d). To sample curves from the posterior distribution given the prior distribution and given the results of testing the 549 compounds, we used Hamiltonian Monte Carlo with No-U-Turn Sampling with Stan[74] (4 chains of 50,000 warm-up and 50,000 sampling steps each and $adapt\_delta = 0.99$, and $max\_treedepth = 12$ control parameters) (Fig. 4a and Extended Data Fig. 8b–d, Red). To select the most informative compounds to test, we evaluated the Shannon Information Gain of six candidate designs, defined as the expected difference in posterior minus prior entropy over the prior-predictive distribution, by nested Monte Carlo[75,76]. We selected Design 5 favoring higher information gain over number of active compounds (Extended Data Fig. 8f). To estimate the number of active compounds (Fig. 4b) and scaffolds (Extended Data Fig. 8g), the energies of the compounds and scaffold cluster heads were integrated over the uncertainty in the posterior hit-rate model (Extended Data Fig. 8h and 8i).
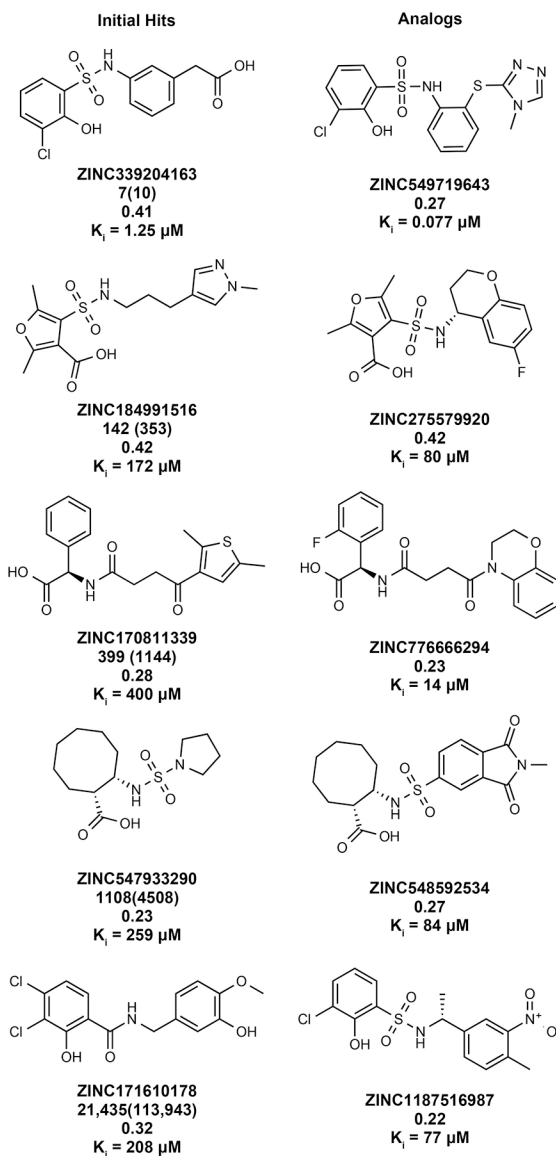
Extended Data
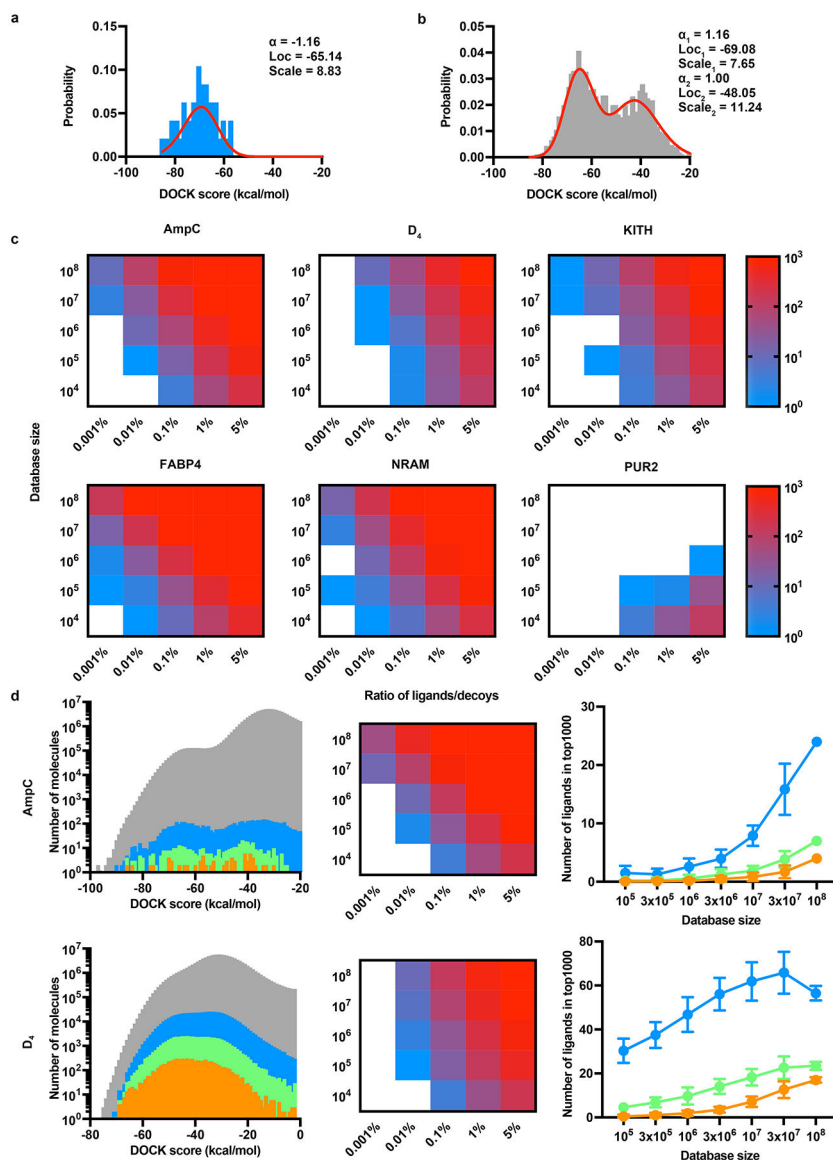
Extended Data Fig. 1 |. Simulating the effect of library size on ligand enrichment among the top 1,000 docked molecules.

The energy distribution of **a**. ligands and **b**. decoys from docking enrichment calculations against AmpC β-lactamase. The skewed normal fitting curves are plotted in red lines. The fitting parameters (α, loc and scale values) are shown. **c**. Heatmaps of number of active molecules in the top 1,000 docked molecules for six targets. The number of ligands in the top 1,000 docked molecules for a given library size and the ratio of ligands/decoys is colored in a $\log_{10}$ scale from 1 (blue) to 1,000 (red). Cells with zero ligands are colored white. **d**. Large-library docking screens of AmpC (top, N=99 million molecules) and D_4 (bottom, N=138 million molecules). Known binders and close analogs are treated as ligands and the rest of the molecules are treated as decoys. Panel on the left: the energy distributions of decoys (grey), ligands defined by ECFP4 Tc similarity   0.5 (blue), 0.6 (green) and 0.7

(orange) to ligands from ChEMBL. Middle Panel: heatmaps of number of ligands in the top 1000 docked molecules based on fit on full-library docking with the ligands (AmpC, Tc 0.5, green; D_4, Tc 0.6, orange) and decoys (grey) distributions. Right panel: number of ligands in the top 1,000 docked molecules as the library grows based on actual distributions plotted in left most panel. The data are the mean ± SD from 20 samples (See Supplementary Table 1 for retrospective performance on three more targets).

Initial Hits

ZINC339204163
7(10)
0.41
K_i = 1.25 μM

ZINC184991516
142 (353)
0.42
K_i = 172 μM

ZINC170811339
399 (1144)
0.28
K_i = 400 μM

ZINC547933290
1108(4508)
0.23
K_i = 259 μM

ZINC171610178
21,435(113,943)
0.32
K_i = 208 μM

Analogs

ZINC549719643
0.27
K_i = 0.077 μM

ZINC275579920
0.42
K_i = 80 μM

ZINC776666294
0.23
K_i = 14 μM

ZINC548592534
0.27
K_i = 84 μM

ZINC1187516987
0.22
K_i = 77 μM

**Extended Data Fig. 2 |. Initial hits and selected analogs against AmpC β-lactamase.**
5 initial hits are shown in the first column. For each compound, the first row is the ZINC ID; the second row is the cluster rank (position in cluster head list sorted by DOCK score) with global rank (position in unclustered hit list sorted by DOCK score) in the brackets; the third row is the Tc value (Tanimoto coefficient to known AmpC inhibitors in ChEMBL); the fourth row is the K_i value. Five selected analogs for the corresponding hits are shown in the

second column. For each compound, the first row is the ZINC ID; the second row is the Tc value; the third row is the K_i value.
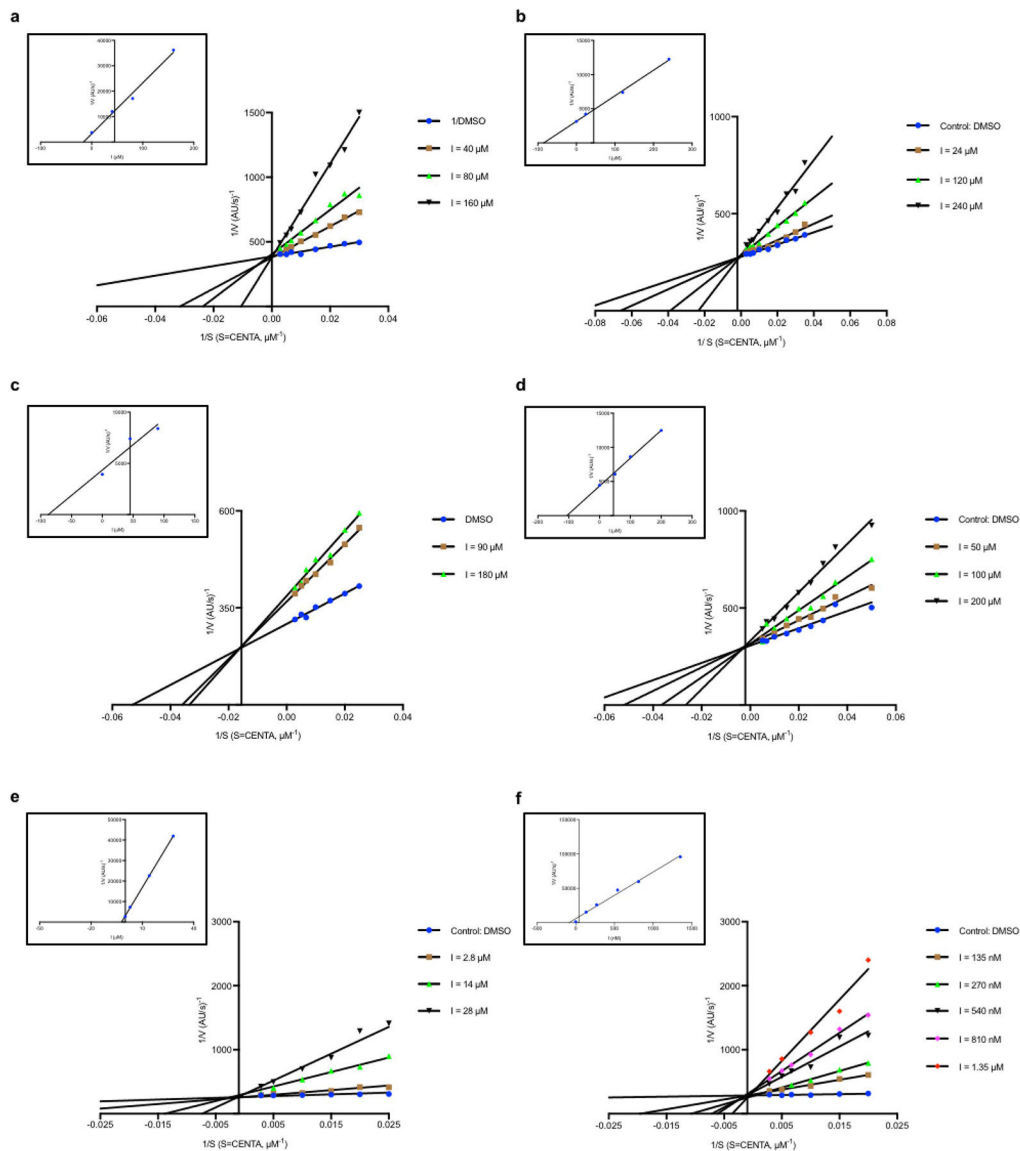

Extended Data Fig. 3 |. Lineweaver-Burk plot and K_i analysis for analogs of each of the five series of AmpC inhibitors.
(a-f) Lineweaver-Burk plots for **'6291** (**a**), **'9920** (**b**), **'2532** (**c**), **'6987** (**d**), **'4163** (**e**), and **'9643** (**f**) indicating competitive inhibition. IC_{50} values were determined by non-linear regression fit in GraphPad Prism, and K_i values calculated by a replot of the slope of each Lineweaver-Burk plot versus the corresponding inhibitor concentration.
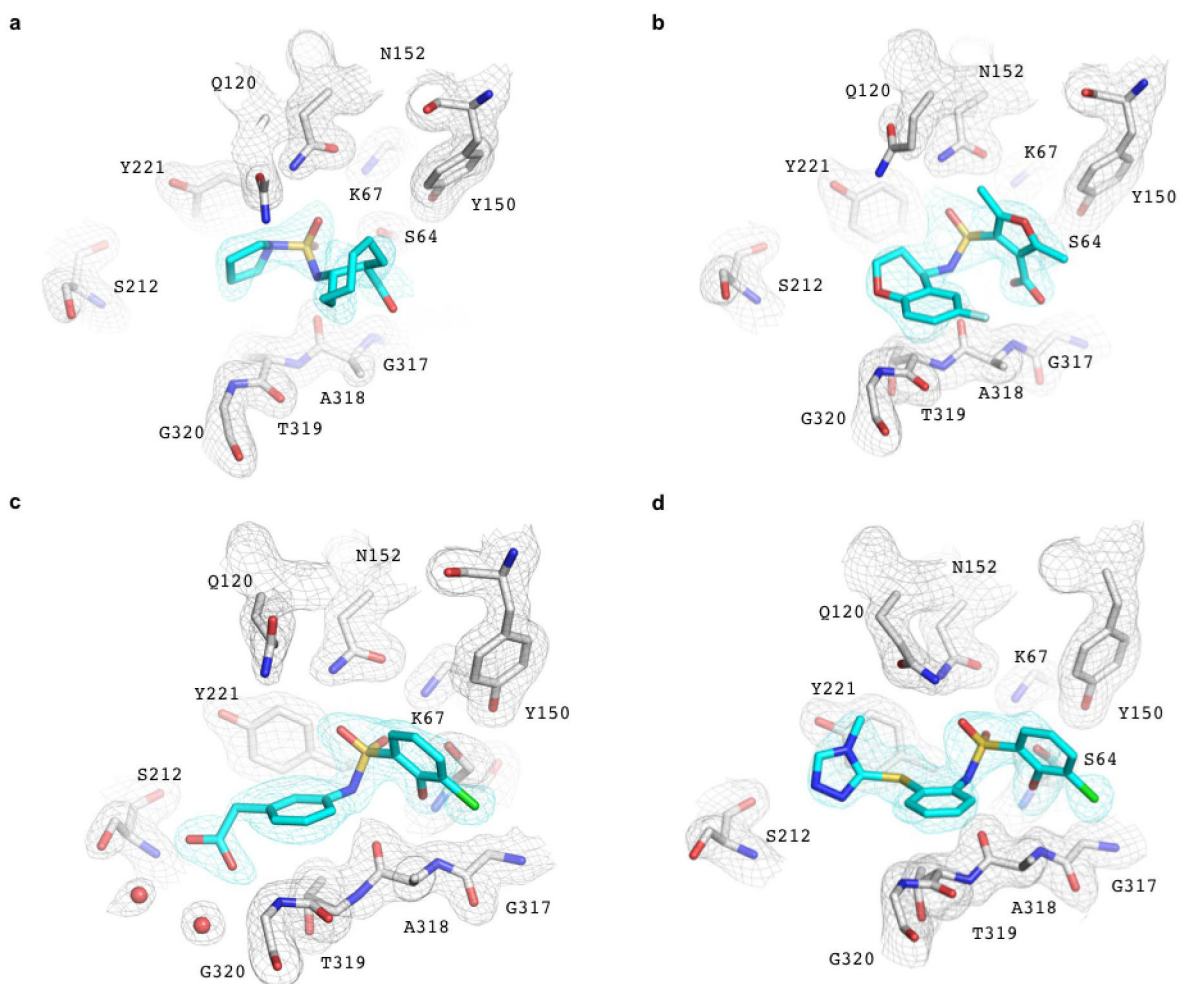
Extended Data Fig. 4 |. Electron density maps for AmpC/inhibitor complexes.
The initial Fo-Fc electron density map contoured at 2.5σ around the inhibitor (density in cyan) with refined 2Fo-Fc electron density contoured at 1σ for enzyme residues for the complexes with compounds **a.** '**3290**, **b. '9920**, **c. '4163** and **d. '9643**. Inhibitor carbons in cyan and enzyme carbons in grey, oxygens red, nitrogens blue, sulfurs yellow and chlorides green.
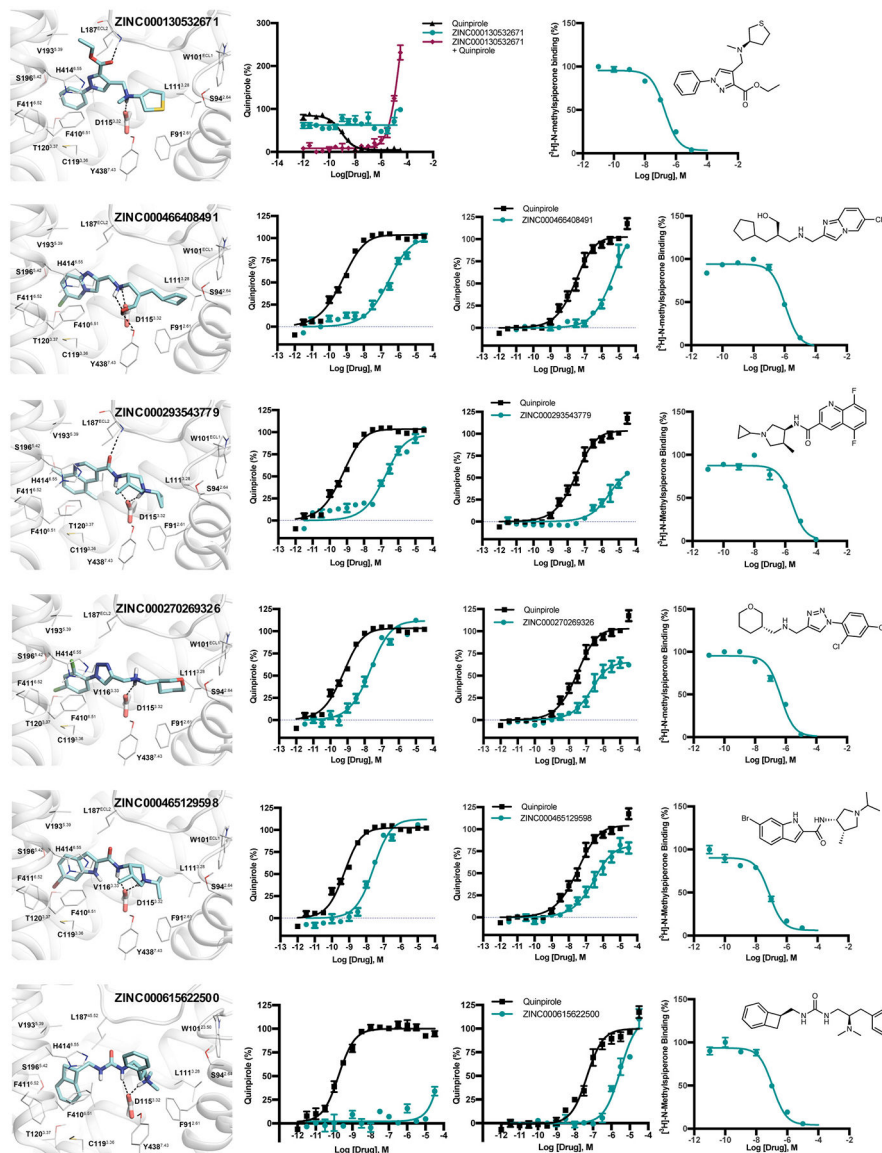
Extended Data Fig. 5 |. Selected D$_4$ hits from docking 138 million make-on-demand molecules.
Six ligands with docked poses (first column), cAMP Gα$_{i/o}$ activities (second column),
Tango β-arrestin activities (third column) and [^{3}H]-N-methylspiperone displacement and
chemical drawing (fourth column) are shown. The receptor structure is in grey and ligand
carbons are in teal. Ballesteros-Weinstein residue numbering in superscript. Functional
assays represent normalized concentration-response curves of the ligands in cloned human
D$_4$-mediated activation of Gα$_{i/o}$ and β-arrestin translocation. The data are the mean ± SEM
from three assays. The first row shows an example of an antagonist identified among the D$_4$
hits. Both agonist (teal curve) and antagonist (purple curve) modes are shown for
ZINC000130532671 in the third panel; the concentration of Quinpirole in the antagonist
mode was 100 nM.
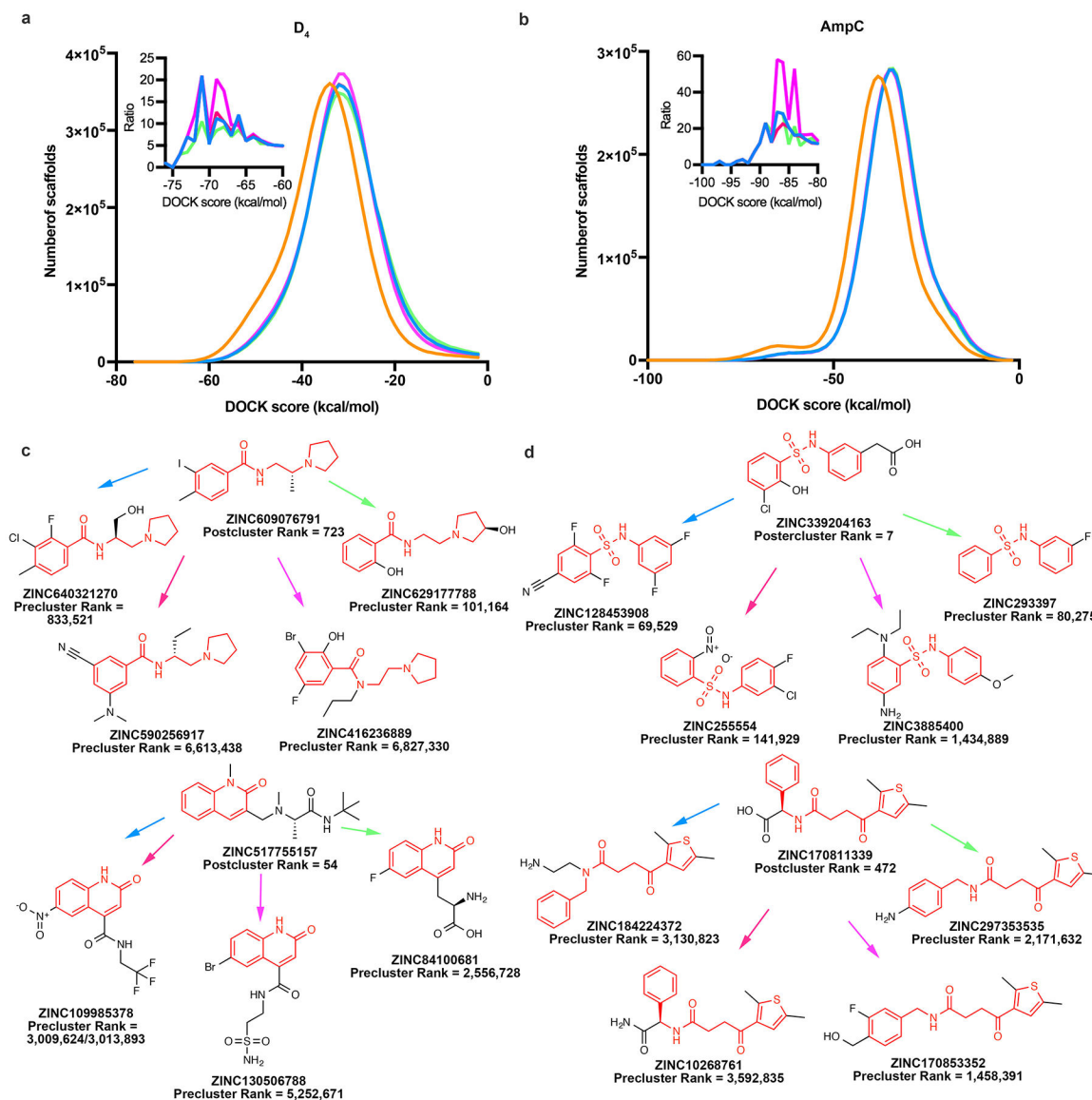
Extended Data Fig. 6 |. Pre-clustering the docking library yields much worse scores of scaffold representatives compared to full library docking.

Comparison of energy distributions of scaffold representatives between full library docking (orange) and pre-clustered library docking for **a**) D_4 and **b**) AmpC using four strategies: the closest member to the centroid of molecular weights and clogP (blue), the closest member to the centroid of molecular weights (pink), the member with the largest molecular weights (magenta) and the member with the smallest molecular weights (green). The inset shows the ratio of the number of molecules at a given docking score for full library docking divided by the number at that score when only cluster representatives are docked (colored by clustering method). For each target, two examples illustrate the effect on our experimentally active scaffold families. **c**) D_4, **d**) AmpC. The scaffold for each molecule is highlighted in red. The ZINC ID, post-cluster rank and pre-cluster rank are labelled for each pair. The arrow color is as for the pre-clustering methods in panels **a** and **b**
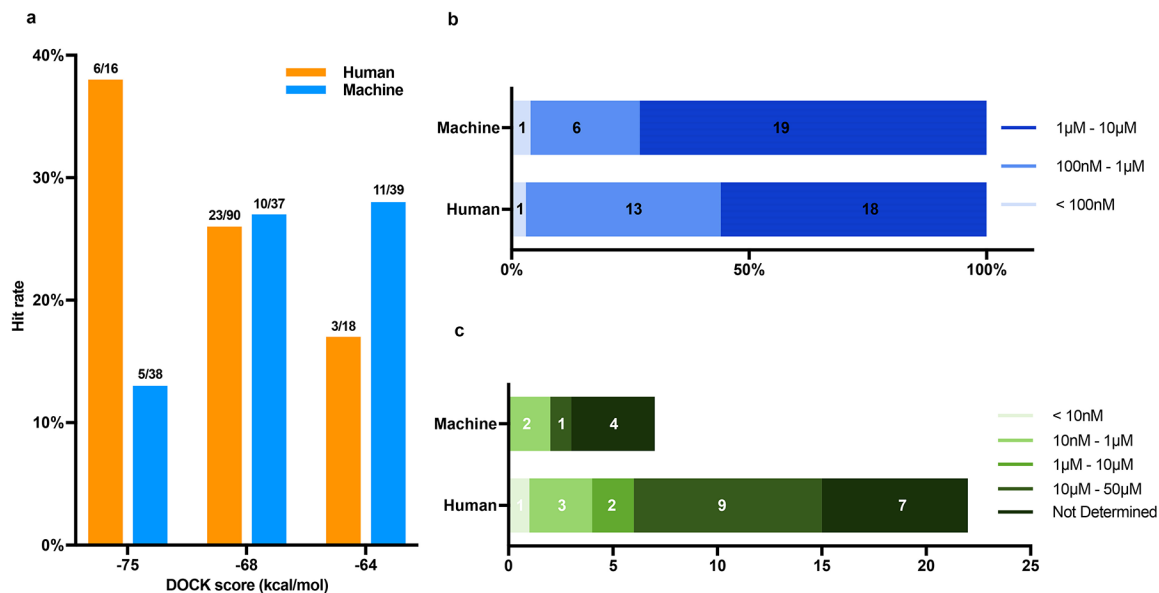
Extended Data Fig. 7 |. Comparison of hit rates achieved by combined docking score and human prioritization vs. by docking score alone.

a) The hit rates from selecting compounds at different scoring ranges by each strategy: human prioritization and docking score (orange), docking score alone (blue). Hit rate is actives/tested; the raw numbers appear at the top of each bar. **b**) Binding affinity level distribution among the hits from panel a. There are 32 hits from human prioritization and docking score, and 26 from docking score alone. These are divided into three affinity ranges: < 100 nM (pale blue); 100 nM - 1 μM (blue); 3) 1 – 10 μM (**dark blue**). **c**) Functional activity distribution among the hits from panel **b**. There are 22 molecules from human prioritization and docking score, and 7 molecules from docking score alone. These are divided in five activity ranges: < 10 nM (pale green); 10 nM - 1 μM (light green); 1 μM - 10 μM (olive); 10 μM - 50 μM (forest green); 5) not determined (**dark green**).
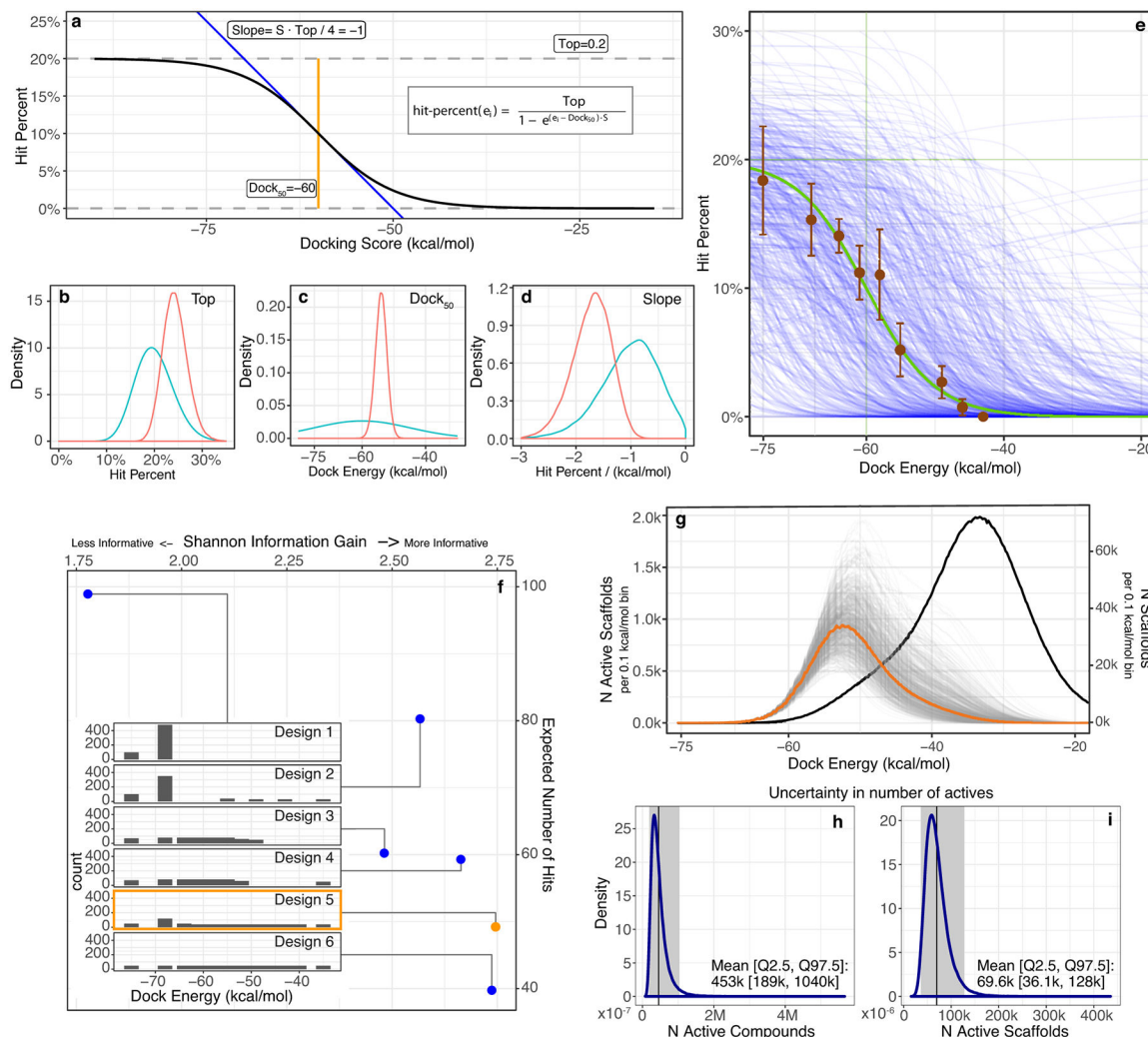
Extended Data Fig. 8 |. Bayesian Prior modeling for balancing information gain and ligand discovery in molecule-selection design and error estimation.
a) Sigmoid functional form for the hit-rate model. **b-d**) Marginal Bayesian prior (teal) and posterior (red) distributions (n=200,000) for each model parameter **b**) Top, **c**) $Dock_{50}$ and **d**) Slope. **e**) Estimated hit-rate based on evaluation by the authors of the docked poses ***before any molecules were tested*** (brown: mean (n compound = 200, 220, 230, 230, 285, 235, 210, 230, 200) ± stddev. (n experts = 5,4,4,4,4,4,4,4,4)), the prior mean (green), and samples (n=200) from the prior (blue). **f**) Candidate (blue) and chosen (orange) experimental designs (Inset Designs 1–6), with expected number of hits and information gain for each. **g**) Expected number of active scaffolds (orange: mean, gray: posterior draws n=200,000) superimposed on the total number of scaffold cluster heads (black). **h-i**) Marginal distribution of the number of active compounds (**h**) and scaffolds (**i**) over the posterior distributions (n=200,000).

Extended Data Table 1 |

Data collection and refinement statistics of β-lactamase AmpC inhibitors.

	ZINC547933290	ZINC275579920	ZINC339204163	ZINC549719643
Data collection				
Space group	C2	C2	C2	C2
Cell dimensions				
a, b, c (Å)	97.70, 77.73, 115.68	97.44, 77.56, 118.41	98.02, 77.7, 115.94	98.17, 77.56, 115.51
a, b, g (°)	90.00, 113.26, 90.00	90.00, 116.05, 90.00	90.00, 113.35, 90.00	90.00, 113.04, 90.00
Resolution (Å)	62.75–1.50 (1.53–1.50)[*]	62.67–1.91 (1.95–1.91)	62.76–1.90 (1.94–1.90)	87.87–1.79 (1.83–1.79)
R_{sym} or R_{merge}	5.6(17.1)	14(20.7)	5.3 (35)	6.9(10.9)
I/s/	13.6(1.0)	10(1.1)	19.6 (4.3)	13.5(1.7)
Completeness (%)	99.5 (98.6)	97.4 (96)	96.4 (71.1)	99.8 (99.6)
Redundancy	6.6 (6.4)	6.9 (7.1)	6.4 (5.0)	6.6 (6.0)
Refinement				
Resolution (Å)	62.7–1.5	58.64–1.91	45.8–1.90	58.46–1.79
No. reflections	126186	59879	60717	74953
R_{work} / R_{free}	19.2/22.3	19.5/23.2	17.2/20.2	18.8/21.9
No. atoms				
Protein	5499	5483	5513	5499
Ligand/ion	40	50	44	50
Water	593	237	326	269
B-factors				
Protein	29.34	33.09	29.76	34.57
Ligand/ion	35.14	38.60	36.59	42.63
Water	37.86	35.87	34.59	38.55
R.m.s. deviations				
Bond lengths (Å)	0.007	0.007	0.008	0.007
Bond angles (°)	0.86	0.84	0.86	0.85

(One crystal for each structure)

[*] Values in parentheses are for highest-resolution shell.

Extended Data Table 2 |

36 of the highest-affinity direct docking hits for D_4 dopamine receptor.

See Supplementary Table 4 for all 549 compounds tested. See Supplementary Table 6 for novelty.

ZINC ID	Cluster Rank*	TC†	cAMP EC$_{50}$ (nM)	G$_i$ BRET EC$_{50}$ (nM)	Tango EC$_{50}$ (nM)	Arrestin BRET EC$_{50}$ (nM)	Bias Factor	K$_i$ (nM)		
								D$_4$	D$_2$	D$_3$
ZINC621433144	611	0.33	0.18	0.56	57.3	2.3	17 to G protein	4.30	>10,000	>10,000
ZINC621433143 (diastereomeric mixture)	611	0.33	2.3	22.8	74.1	13.03	1	18	>10,000	>10,000
ZINC361131264	611	0.33	7.3	21.93	151	890.1	26 to G protein	72	>10,000	>10,000
ZINC361131265	611	0.33	59	1,778	556	ND‡	11 to G protein	185	>10,000	>10,000
ZINC621433143	611	0.33	350	3,148	898	1,246	7 to Arrestin	669	>10,000	>10,000
ZINC464771011	937	0.32	10.4	NT§	1,141	NT	3 to Arrestin	140	>10,000	>10,000
ZINC270269326	1,510	0.32	17	11.7	172.3	NT	2 to Arrestin	500	>10,000	>10,000
ZINC465129598	1,565	0.32	24	22.9	210.1	NT	2 to Arrestin	80	>10,000	>10,000
ZINC278933042	5,796	0.27	41	NT	22,340	NT	NT	270	>10,000	>10,000
ZINC293543779	420	0.31	130	NT	1,684	NT	1	1200	>10,000	>10,000
ZINC466408491	274	0.27	270	NT	8,529	NT	4 to Arrestin	3,900	4,780	5,430
ZINC567992445	5,854	0.33	4,820	NT	NT	NT	NT	850	>10,000	>10,000
ZINC413570733	81,969	0.28	IC$_{50}$ = 5,900	NT	NT	NT	NT	130	>10,000	>10,000
ZINC155719879	132	0.31	7,200	NT	NT	NT	NT	460	>10,000	>10,000
ZINC569686370	17,659	0.34	7,800	NT	NT	NT	NT	550	>10,000	>10,000
ZINC268141382	734	0.3	9,900	NT	NT	NT	NT	560	>10,000	>10,000
ZINC127859549	42,379	0.33	10,400	NT	NT	NT	NT	260	>10,000	3,570
ZINC437075156	1,542	0.34	10,400	NT	NT	NT	NT	190	>10,000	>10,000
ZINC130532671	89	0.3	IC$_{50}$ = 10,800	NT	NT	NT	NT	320	>10,000	>10,000
ZINC51842964	1,631	0.33	22,100	NT	NT	NT	NT	120	>10,000	>10,000
ZINC651262870	1,328	0.32	31,000	NT	NT	NT	NT	440	>10,000	>10,000
ZINC609076791	572	0.33	31,500	NT	NT	NT	NT	330	>10,000	>10,000
ZINC362128724	1,615	0.34	32,200	NT	NT	NT	NT	160	8,610	6,040
ZINC467716766	1,171	0.31	34,800	NT	NT	NT	NT	700	>10,000	>10,000
ZINC375299581	198	0.32	ND	NT	NT	NT	NT	180	>10,000	>10,000
ZINC247739858	225	0.35	ND	NT	NT	NT	NT	210	>10,000	>10,000

ZINC ID	Cluster Rank[*]	TC[†]	cAMP EC$_{50}$ (nM)	G$_i$ BRET EC$_{50}$ (nM)	Tango EC$_{50}$ (nM)	Arrestin BRET EC$_{50}$ (nM)	Bias Factor	K$_i$ (nM)		
								D$_4$	D$_2$	D$_3$
ZINC176752603	92	0.3	ND	NT	NT	NT	NT	300	>10,000	>10,000
ZINC480408888	75	0.34	ND	NT	NT	NT	NT	400	>10,000	>10,000
ZINC449156693	5,802	0.28	ND	NT	NT	NT	NT	760	>10,000	>10,000
ZINC64028109	255	0.29	ND	NT	NT	NT	NT	810	>10,000	>10,000
ZINC92642352	42,403	0.23	ND	NT	NT	NT	NT	820	>10,000	>10,000
ZINC572302473	577	0.29	ND	NT	NT	NT	NT	910	>10,000	>10,000
ZINC601953994	235	0.32	ND	NT	NT	NT	NT	980	>10,000	>10,000
ZINC615622500	204	0.34	ND	NT	3,074	NT	ND	150	>10,000	5,400
ZINC358506216	5,857	0.33	NT	NT	NT	NT	NT	250	5,880	6,540
ZINC41555869 7	17,674	0.27	NT	NT	NT	NT	NT	990	2,000	5,870
ZINC237938360	17,654	0.32	NT	NT	NT	NT	NT	180	5,720	1,070
Quinpirole	ND	ND	0.32	1.82	63.3	NT	1	14.2	ND	ND
Dopamine	ND	ND	0.65	0.91	79.4	NT	2 to G protein	6.30	ND	ND

*
Cluster Rank: position in cluster head (see Methods) list sorted by DOCK score;

†
TC: Tanimoto Coefficient to dopaminergic, serotonergic, or adrenergic knowns from ChEMBL;

‡
Not determined, ND. The compound was tested but no measurable value was observed.

§
Not Tested, NT

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

References

1. Bohacek RS, McMartin C & Guida WC The art and practice of structure-based drug design: A molecular modeling perspective. Medicinal Research Reviews 16, 3–50, (1996). [PubMed: 8788213]

2. Ertl P Cheminformatics Analysis of Organic Substituents: Identification of the Most Common Substituents, Calculation of Substituent Properties, and Automatic Identification of Drug-like Bioisosteric Groups. Journal of Chemical Information and Computer Sciences 43, 374–380, (2003). [PubMed: 12653499]

3. Fink T, Bruggesser H & Reymond JL Virtual Exploration of the Small-Molecule Chemical Universe below 160 Daltons. Angewandte Chemie International Edition 44, 1504–1508, (2005). [PubMed: 15674983]

4. Chevillard F & Kolb P SCUBIDOO: a large yet screenable and easily searchable database of computationally created chemical compounds optimized toward high likelihood of synthetic tractability. Journal of chemical information and modeling 55, 1824–1835 (2015). [PubMed: 26282054]

5. Keserü GM & Makara GM The influence of lead discovery strategies on the properties of drug candidates. Nature Reviews Drug Discovery 8, 203, (2009). [PubMed: 19247303]

6. McGovern SL, Caselli E, Grigorieff N & Shoichet BK A Common Mechanism Underlying Promiscuous Inhibitors from Virtual and High-Throughput Screening. Journal of Medicinal Chemistry 45, 1712–1722, (2002). [PubMed: 11931626]

7. Brenner S & Lerner RA Encoded combinatorial chemistry. Proceedings of the National Academy of Sciences 89, 5381–5383 (1992).

8. Ahn S et al. Allosteric "beta-blocker" isolated from a DNA-encoded small molecule library. Proceedings of the National Academy of Sciences 114, 1708 (2017).

9. Goodnow RA, Jr, Dumelin CE & Keefe AD DNA-encoded chemistry: enabling the deeper sampling of chemical space. Nature Reviews Drug Discovery 16, 131, (2016). [PubMed: 27932801]

10. Jorgensen WL The many roles of computation in drug discovery. Science 303, 1813–1818 (2004). [PubMed: 15031495]

11. De Graaf C et al. Crystal structure-based virtual screening for fragment-like ligands of the human histamine H1 receptor. Journal of medicinal chemistry 54, 8195–8206 (2011). [PubMed: 22007643]

12. Katritch V et al. Structure-based discovery of novel chemotypes for adenosine A2A receptor antagonists. Journal of medicinal chemistry 53, 1799–1809 (2010). [PubMed: 20095623]

13. Manglik A et al. Structure-based discovery of opioid analgesics with reduced side effects. Nature 537, 185 (2016). [PubMed: 27533032]

14. Wang S et al. D4 dopamine receptor high-resolution structures enable the discovery of selective agonists. Science 358, 381–386 (2017). [PubMed: 29051383]

15. Negri A et al. Discovery of a novel selective kappa-opioid receptor agonist using crystal structure-based virtual screening. Journal of chemical information and modeling 53, 521–526 (2013). [PubMed: 23461591]

16. Jazayeri A, Andrews SP & Marshall FH Structurally enabled discovery of adenosine A2A receptor antagonists. Chemical reviews 117, 21–37 (2016). [PubMed: 27333206]

17. Lane JR et al. Structure-based ligand discovery targeting orthosteric and allosteric pockets of dopamine receptors. Molecular pharmacology 84, 794–807 (2013). [PubMed: 24021214]

18. Langmead CJ et al. Identification of novel adenosine A2A receptor antagonists by virtual screening. Journal of medicinal chemistry 55, 1904–1909 (2012). [PubMed: 22250781]

19. Becker OM et al. G protein-coupled receptors: in silico drug discovery in 3D. Proceedings of the National Academy of Sciences 101, 11304–11309 (2004).

20. Kooistra AJ et al. Function-specific virtual screening for GPCR ligands using a combined scoring method. Scientific reports 6, 28288 (2016). [PubMed: 27339552]

21. Congreve M et al. Discovery of 1, 2, 4-triazine derivatives as adenosine A2A antagonists using structure based drug design. Journal of medicinal chemistry 55, 1898–1903 (2012). [PubMed: 22220592]

22. Kiss R et al. Discovery of novel human histamine H4 receptor ligands by large-scale structure-based virtual screening. Journal of medicinal chemistry 51, 3145–3153 (2008). [PubMed: 18459760]

23. Oprea TI & Gottfries J Chemography: the art of navigating in chemical space. Journal of combinatorial chemistry 3, 157–166 (2001). [PubMed: 11300855]

24. Mysinger MM, Carchia M, Irwin JJ & Shoichet BK Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. Journal of medicinal chemistry 55, 6582–6594 (2012). [PubMed: 22716043]

25. Bemis GW & Murcko MA The Properties of Known Drugs. 1. Molecular Frameworks. Journal of Medicinal Chemistry 39, 2887–2893, (1996). [PubMed: 8709122]

26. Gaulton A et al. The ChEMBL database in 2017. Nucleic acids research 45, D945–D954 (2016). [PubMed: 27899562]

27. Katz BA et al. A novel serine protease inhibition motif involving a multi-centered short hydrogen bonding network at the active site1. Journal of molecular biology 307, 1451–1486 (2001). [PubMed: 11292354]

28. Congreve M, Langmead CJ, Mason JS & Marshall FH Progress in structure based drug design for G protein-coupled receptors. Journal of medicinal chemistry 54, 4283–4311 (2011). [PubMed: 21615150]

29. Vaidehi N Dynamics and flexibility of G-protein-coupled receptor conformations and their relevance to drug design. Drug discovery today 15, 951–957 (2010). [PubMed: 20831898]

30. Irwin JJ & Shoichet BK Docking Screens for Novel Ligands Conferring New Biology. Journal of Medicinal Chemistry 59, 4103–4120, (2016). [PubMed: 26913380]

31. Vass M et al. in Computational Methods for GPCR Drug Discovery 73–113 (Springer, 2018).

32. Isberg V et al. Generic GPCR residue numbers–aligning topology maps while minding the gaps. Trends in pharmacological sciences 36, 22–31 (2015). [PubMed: 25541108]

33. McCorvy JD et al. Structural determinants of 5-HT2B receptor activation and biased agonism. Nature Structural & Molecular Biology 25, 787–796, (2018).

34. Lyu J, Balius T, Levit A, Shoichet BK & Irwin JJ D4_benchmark_mols.mol2. figshare doi: 10.6084/m9.figshare.7367288.v2 (2018).

35. Powers RA, Morandi F & Shoichet BK Structure-based discovery of a novel, noncovalent inhibitor of AmpC β-lactamase. Structure 10, 1013–1023 (2002). [PubMed: 12121656]

36. Feng BY et al. A High-Throughput Screen for Aggregation-Based Inhibition in a Large Compound Library. Journal of Medicinal Chemistry 50, 2385–2390, (2007). [PubMed: 17447748]

37. Babaoglu K et al. Comprehensive Mechanistic Analysis of Hits from High-Throughput and Docking Screens against β-Lactamase. Journal of Medicinal Chemistry 51, 2502–2511, (2008). [PubMed: 18333608]

38. Rowley M et al. 5-(4-Chlorophenyl)-4-methyl-3-(1-(2-phenylethyl) piperidin-4-yl) isoxazole: a potent, selective antagonist at human cloned dopamine D4 receptors. Journal of medicinal chemistry 39, 1943–1945 (1996). [PubMed: 8642551]

39. Enguehard-Gueiffier C et al. 2-[(4-phenylpiperazin-1-yl) methyl] imidazo (di) azines as selective D4-ligands. Induction of penile erection by 2-[4-(2-methoxyphenyl) piperazin-1-ylmethyl] imidazo [1, 2-a] pyridine (PIP3EA), a potent and selective D4 partial agonist. Journal of medicinal chemistry 49, 3938–3947 (2006). [PubMed: 16789750]

40. Löber S, Hübner H & Gmeiner P Synthesis and biological investigations of dopaminergic partial agonists preferentially recognizing the D4 receptor subtype. Bioorganic & medicinal chemistry letters 16, 2955–2959 (2006). [PubMed: 16563764]

41. Lindsley CW & Hopkins CR Return of D4 Dopamine Receptor Antagonists in Drug Discovery: Miniperspective. Journal of medicinal chemistry 60, 7233–7243 (2017). [PubMed: 28489950]

42. Tirado-Rives J & Jorgensen WL Contribution of conformer focusing to the uncertainty in predicting free energies for protein– ligand binding. Journal of medicinal chemistry 49, 5880–5884 (2006). [PubMed: 17004703]

43. Abagyan R, Totrov M & Kuznetsov D ICM—a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. Journal of computational chemistry 15, 488–506 (1994).

44. Halgren TA et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. Journal of medicinal chemistry 47, 1750–1759 (2004). [PubMed: 15027866]

45. Goodsell DS & Olson AJ Automated docking of substrates to proteins by simulated annealing. Proteins: Structure, Function, and Bioinformatics 8, 195–202 (1990).

46. Kufareva I, Katritch V, Stevens RC & Abagyan R Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges. Structure 22, 1120–1139 (2014). [PubMed: 25066135]

47. Kramer B, Rarey M & Lengauer T Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. Proteins: Structure, Function, and Bioinformatics 37, 228–241 (1999).

48. McGann M FRED pose prediction and virtual screening accuracy. Journal of chemical information and modeling 51, 578–596 (2011). [PubMed: 21323318]

49. Jones G, Willett P, Glen RC, Leach AR & Taylor R Development and validation of a genetic algorithm for flexible docking. Journal of molecular biology 267, 727–748 (1997). [PubMed: 9126849]

50. Corbeil CR, Williams CI & Labute P Variability in docking success rates due to dataset preparation. Journal of computer-aided molecular design 26, 775–786 (2012). [PubMed: 22566074]

References to Online Methods

51. Hawkins PC, Skillman AG, Warren GL, Ellingson BA & Stahl MT Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. Journal of chemical information and modeling 50, 572–584 (2010). [PubMed: 20235588]

52. AMSOL 7.1, University of Minnesota, Minneapolis (2004).

53. Wei BQ, Baase WA, Weaver LH, Matthews BW & Shoichet BK A model binding site for testing scoring functions in molecular docking. Journal of molecular biology 322, 339–355 (2002). [PubMed: 12217695]

54. Mysinger MM & Shoichet BK Rapid context-dependent ligand desolvation in molecular docking. Journal of chemical information and modeling 50, 1561–1573 (2010). [PubMed: 20735049]

55. Sterling T & Irwin JJ ZINC 15 – Ligand Discovery for Everyone. Journal of Chemical Information and Modeling 55, 2324–2337, (2015). [PubMed: 26479676]

56. Barelier S et al. Increasing chemical space coverage by combining empirical and computational fragment screens. ACS chemical biology 9, 1528–1535 (2014). [PubMed: 24807704]

57. Gray DL et al. Impaired β-arrestin recruitment and reduced desensitization by non-catechol agonists of the D1 dopamine receptor. Nature communications 9, 674 (2018).

58. Balius T, Lyu J, Shoichet BK & Irwin JJ AmpC_screen_table.csv.gz. figshare doi:10.6084/m9.figshare.7359626.v2 (2018).

59. Balius T, Lyu J, Levit A, Shoichet BK & Irwin JJ D4_screen_table.csv.gz. figshare doi:10.6084/m9.figshare.7359401.v3 (2018).

60. Carlsson J et al. Ligand discovery from a dopamine D3 receptor homology model and crystal structure. Nature chemical biology 7, 769–778 (2011). [PubMed: 21926995]

61. Meng EC, Shoichet BK & Kuntz ID Automated docking with grid-based energy evaluation. Journal of computational chemistry 13, 505–524 (1992).

62. Sharp KA, Friedman RA, Misra V, Hecht J & Honig B Salt effects on polyelectrolyte–ligand binding: Comparison of Poisson–Boltzmann, and limiting law/counterion binding models. Biopolymers 36, 245–262 (1995). [PubMed: 7492748]

63. Gallagher K & Sharp K Electrostatic contributions to heat capacity changes of DNA-ligand binding. Biophysical journal 75, 769–776 (1998). [PubMed: 9675178]

64. Coleman RG, Carchia M, Sterling T, Irwin JJ & Shoichet BK Ligand pose and orientational sampling in molecular docking. PloS one 8, e75992 (2013). [PubMed: 24098414]

65. Tolmachev A et al. Expanding Synthesizable Space of Disubstituted 1, 2, 4-Oxadiazoles. ACS combinatorial science 18, 616–624 (2016). [PubMed: 27548754]

66. Eidam O et al. Design, synthesis, crystal structures, and antimicrobial activity of sulfonamide boronic acids as β-lactamase inhibitors. Journal of medicinal chemistry 53, 7852–7863 (2010). [PubMed: 20945905]

67. Kabsch WXDS. Acta Crystallographica Section D: Biological Crystallography 66, 125–132, (2010). [PubMed: 20124692]

68. Emsley P, Lohkamp B, Scott WG & Cowtan K Features and development of Coot. Acta Crystallographica Section D: Biological Crystallography 66, 486–501 (2010). [PubMed: 20383002]

69. Murshudov GN, Vagin AA & Dodson EJ Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallographica Section D: Biological Crystallography 53, 240–255 (1997). [PubMed: 15299926]

70. Adams PD et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D: Biological Crystallography 66, 213–221 (2010). [PubMed: 20124702]

71. Eidam O et al. Fragment-guided design of subnanomolar β-lactamase inhibitors active in vivo. Proceedings of the National Academy of Sciences 109, 17448–17453 (2012).

72. Feng BY & Shoichet BK A detergent-based assay for the detection of promiscuous inhibitors. Nature protocols 1, 550 (2006). [PubMed: 17191086]

73. Allen JA et al. Discovery of β-arrestin–biased dopamine D2 ligands for probing signal transduction pathways essential for antipsychotic efficacy. Proceedings of the National Academy of Sciences 108, 18488–18493 (2011).

74. Carpenter B et al. Stan: A Probabilistic Programming Language. Journal of Statistical Software; *Vol* 1, *Issue* 1 (2017) (2017).

75. R. EG, D. CC, M. JM & P. AN A Review of Modern Computational Algorithms for Bayesian Optimal Design. International Statistical Review 84, 128–154, (2016).

76. Rainforth T, Cornish R, Yang H, Warrington A & Wood F On Nesting Monte Carlo Estimators. To appear at International Conference on Machine Learning 2018, (2018).
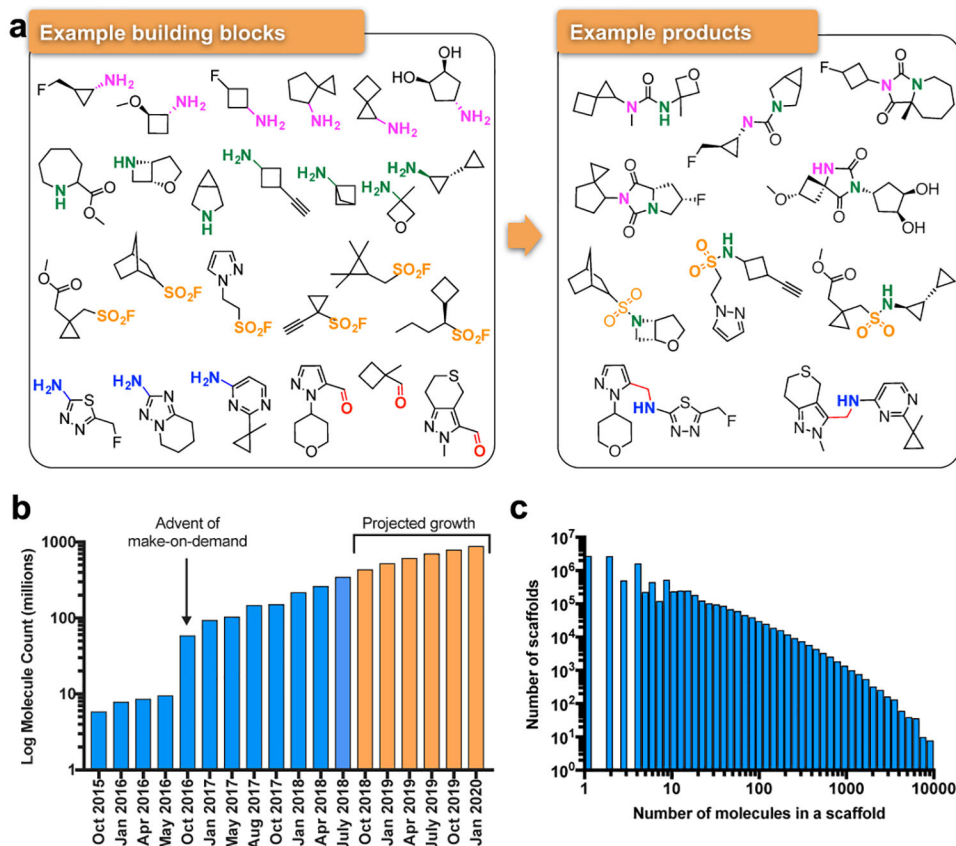
Fig. 1 |. Make-on-demand compounds are diverse and have increased exponentially.
a. Characteristic reagents, reactions, and products in the make-on-demand library. **b.** The expansion of the make-on-demand library; orange bars represent projected growth. **c.** The distribution of compounds among the 10.7 million scaffolds in the library.
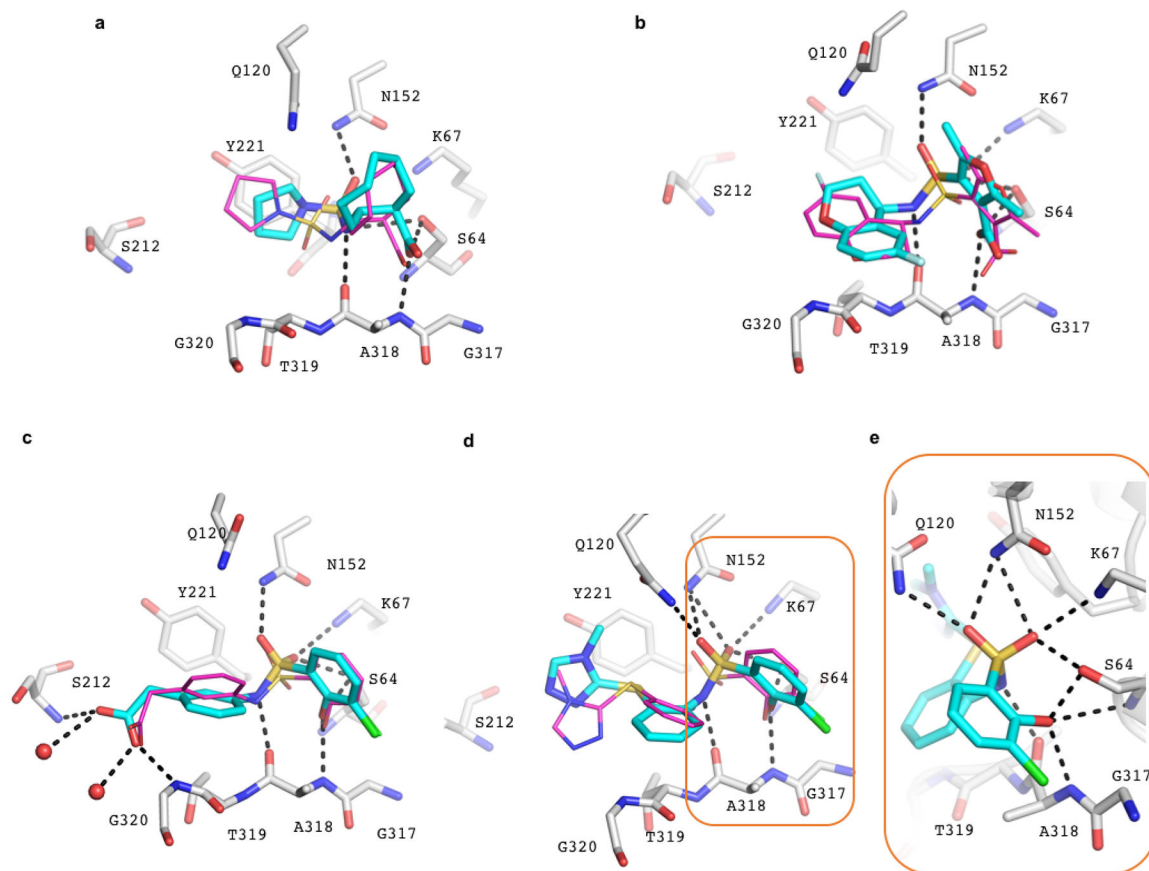
Fig. 2 |. Structural fidelity between docked-predicted and crystallographically-determined poses of the new β-lactamase inhibitors.

Crystal structures of the inhibitors (carbons in cyan) overlaid with their docking predictions (magenta). AmpC carbon atoms in grey, oxygens in red, nitrogens in blue, sulfurs in yellow, chlorides in green, fluorides in light blue. Hydrogen bonds are shown as black dashed lines. The AmpC complexes with **a**. **'3290** (PDB 6DPZ, RMSD 1.3 Å); **b**. **'9920** (PDB 6DPY, RMSD 1.2 Å for the warhead); **c**. The 1.3 μM inhibitor **'4163** (PDB 6DPX; RMSD 0.98 Å) and **d**. its 77 nM analog **'9643** (PDB 6DPT, RMSD 1.52 Å). **e**. Close up of the **'9643** phenolate in the oxyanion hole. Extended Fig. 4 shows electron density.
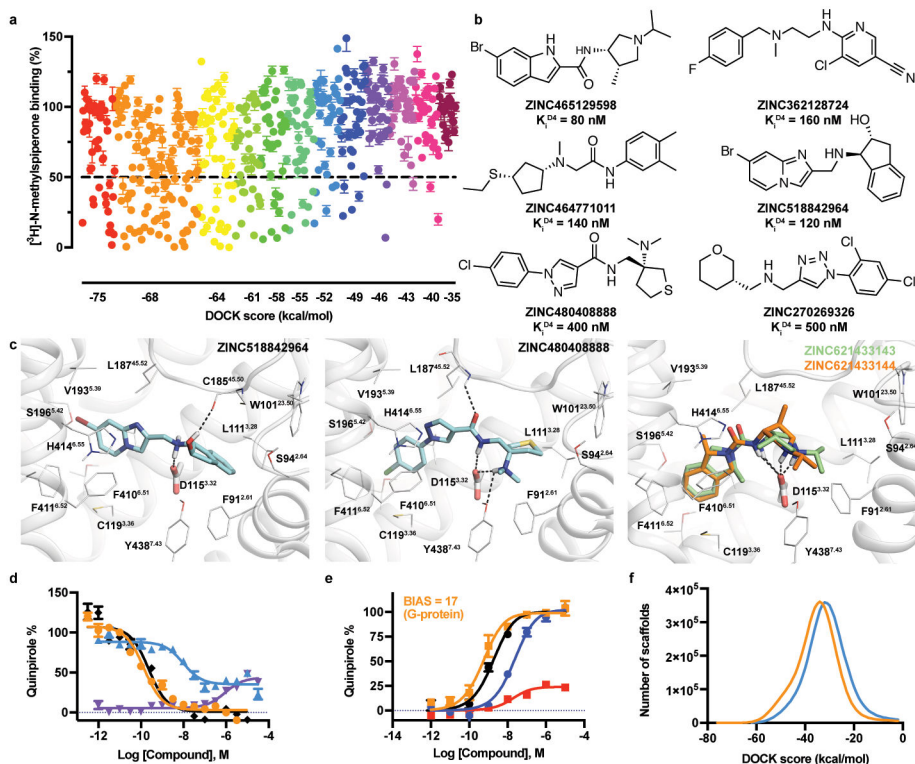
Fig. 3 |. Testing 549 molecules at different docking ranks against the D$_4$ dopamine receptor.
a. Displacement of the antagonist ^{3}H-N-methylspiperone by each of the 549 molecules tested at 10 μM (mean ± SEM of three assays). The molecules are colored by their docking score. The number of binders (<50% remaining radio-ligand—below the dashed line) diminish with docking score. **b**. Six actives, each a different scaffold. **c**. Docked poses of **'2964** (left panel), **'8888** (middle panel), and superposed **'3143** and **'3144** (right panel). The receptor helices are shown in ribbon, the conserved D115$^{3.32}$ is shown in stick, interacting residues within 4Å of the docked molecules are shown as lines. Ballesteros-Weinstein residue numbering in superscript. Modeled hydrogen bonds are in dashed lines. **d**. cAMP functional assays of the 180 pM full agonist **'3144** (orange) and the 10 nM partial agonist **'1011** (blue, agonist mode, purple, antagonist mode (**'1011**+100 nM Quinpirole)). The data are the mean ± SEM from three assays. **e**. Gα$_{i/o}$ BRET and arrestin BRET functional assays of the 180 pM full agonist **'3144** (Gα$_{i/o}$, orange; arrestin, red) and the unbiased ligand Quinpirole (Gα$_{i/o}$, black; arrestin, blue). The data are the mean ± SEM from three assays. **f**. The effect of pre-clustering on docking scores: the orange curve is the distribution of the best-scoring scaffold representative, the blue curve is the score distribution from pre-clustering and choosing only single cluster representatives to dock.
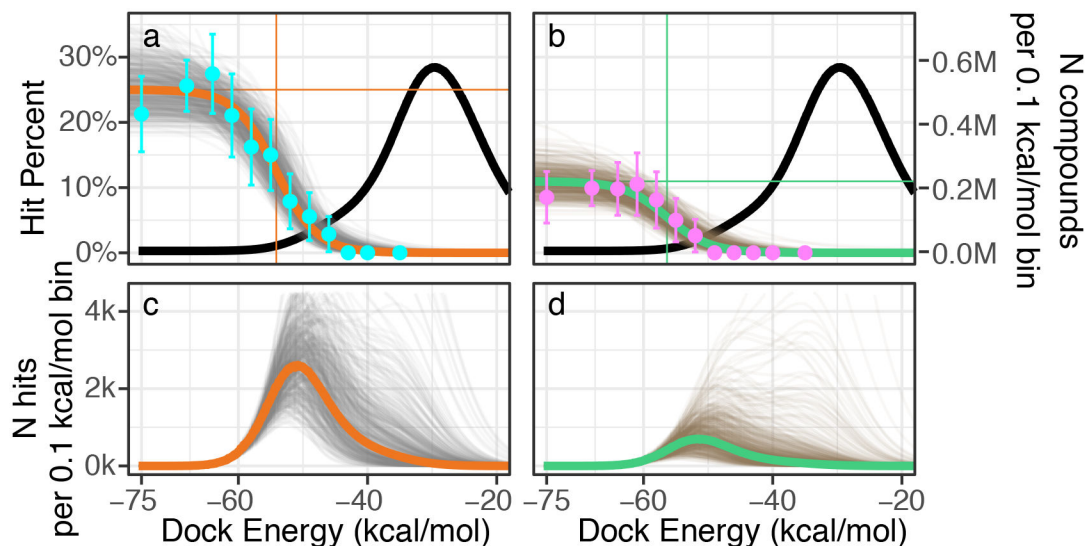
Fig. 4 |. Estimating the number of active D_4 dopamine receptor ligands in the 138 million compound library.

Top row). Left y-axis, the hit-rate of 549 tested compounds, right y-axis, distribution of library compounds by docking energy (black curve). **a.** Modeling the number of library compounds with K_i values 10 μM. Top = 24%; Bottom = 0%; $Dock_{50}$ = −54 kcal/mol; and $Slope_{50}$ = −1.7 % / (kcal/mol). Cyan points represent the hit-rate means and standard errors at each docking energy bin, with 47,121, 51, 38, 37, 40, 38, 36, 35, 36, 35, 35 compounds tested in each bin, from best to worst scoring. The gold curve gives the mode and the gray curves give the draws (n=500) from the Bayesian posterior distribution (i.e., the envelope of possible distributions). **b.** Modeling the number of library compounds with K_i values 1 μM. Top = 11%; Bottom = 0%; $Dock_{50}$ = −56 kcal/mol; and Slope = −2.8 % / (kcal/mol). Magenta points represent the hit-rate means and standard errors at each docking energy bin. The green curve gives the mode and the gray curves give the draws from the Bayesian posterior distribution. **Bottom row**). **c.** Predicted number of actives by docking energy under the hit-rate model for the 10 μM model and **d.** the 1 μM model, with the mode (gold; green) and draws (gray; brown) from the respective posterior distributions. Expected total actives for the 10 μM model = 453,000 (188,000–1,035,000, 95% inter-quantile range) and for the 1 μM model = 158,000 (38,000–489,000, 95% inter-quantile range).