Identifying mechanism-of-action targets for drugs and probes

Elisabet Gregori-Puigjané[a,1], Vincent Setola[b,1], Jérôme Hert[a], Brenda A. Crews[c], John J. Irwin[a], Eugen Lounkine[d], Lawrence Marnett[c], Bryan L. Roth[b,e,2], and Brian K. Shoichet[a,2]

[a]Department of Pharmaceutical Chemistry, University of California, 1700 Fourth Street, San Francisco, CA 94143-2550; [b]Department of Pharmacology, University of North Carolina at Chapel Hill School of Medicine, 4009 Genetics Medicine CB7365, Chapel Hill, NC 27599-7365; [e]National Institute of Mental Health Psychoactive Drug Screening Program, Division of Chemical Biology and Medicinal Chemistry, Program in Neuroscience and Lineberger Cancer Center, 4009 Genetics Medicine CB7365, Chapel Hill, NC 27599-7365; [c]Department of Biochemistry, Vanderbilt University School of Medicine, 23rd Avenue South and Pierce Room 850 RRB, Nashville, TN 37232-0146; and [d]Novartis Institutes for Biomedical Research, 250 Massachusetts Avenue, Cambridge, MA 02139

Edited by Stephen J. Benkovic, The Pennsylvania State University, University Park, PA, and approved May 8, 2012 (received for review March 16, 2012)

Notwithstanding their key roles in therapy and as biological probes, 7% of approved drugs are purported to have no known primary target, and up to 18% lack a well-defined mechanism of action. Using a chemoinformatics approach, we sought to "de-orphanize" drugs that lack primary targets. Surprisingly, targets could be easily predicted for many: Whereas these targets were not known to us nor to the common databases, most could be confirmed by literature search, leaving only 13 Food and Drug Administration—approved drugs with unknown targets; the number of drugs without molecular targets likely is far fewer than reported. The number of worldwide drugs without reasonable molecular targets similarly dropped, from 352 (25%) to 44 (4%). Nevertheless, there remained at least seven drugs for which reasonable mechanism-of-action targets were unknown but could be predicted, including the antitussives clemastine, cloperastine, and nepinalone; the antiemetic benzquinamide; the muscle relaxant cyclobenzaprine; the analgesic nefopam; and the immunomodulator lobenzarit. For each, predicted targets were confirmed experimentally, with affinities within their physiological concentration ranges. Turning this question on its head, we next asked which drugs were specific enough to act as chemical probes. Over 100 drugs met the standard criteria for probes, and 40 did so by more stringent criteria. A chemical information approach to drug-target association can guide therapeutic development and reveal applications to probe biology, a focus of much current interest.

chemical tools | drug target identification | polypharmacology

Recent studies suggest that for many approved drugs, a primary target is unknown. In an influential review, Drews reported that 7% of approved drugs lack a defined molecular target (1), and in their seminal paper on modern drug development, Overington et al. were only able to assign mechanism-of-action protein targets to about 82% of Food and Drug Administration (FDA) approved drugs (2). Given the critical role small molecule drugs play in medicine, and their potential to serve as tools in biology, the lack of a commonly accepted primary target for so many molecules seemed almost provocative.

Almost all drug discovery now begins with activity of molecules on a molecular target, and it is hard to imagine how the primary target of such molecules would be unknown. However, at least half of drugs date from the premolecular era, when their action was explored against whole tissues, rarely on isolated proteins, and target identities were only inferred from tissue-based responses. Even today, 37% of first-in-class drugs derive from phenotypic screens (3), and the targets for some new drugs remain unknown. Recent examples include the emergent polypharmacology of imatinib and olanzipine, where the multiple targets not only explain side effects but also therapeutic efficacy (4). Thus, much effort has been spent on target discovery, including large-scale experimental screening of approved therapeutics against G-protein coupled receptors (GPCRs) (5) and kinases (6). Meanwhile, an ingenious chemical biology tool set has emerged for target identification (7–9). This not only improved our understanding and use of these drugs but has enabled their deployment as tools to probe biology.

Most of these chemical biology approaches focus on a particular target class or experimental strategy and do not lend themselves to the target identification for the diverse set of drugs for which targets remain unknown. We therefore investigated the chemoinformatic similarity ensemble approach (SEA) to linking drugs to targets, followed by experimental testing in direct binding assays. Whereas SEA and other chemoinformatics approaches (10–12) have liabilities (to which we shall return), they can comprehensively and systematically interrogate all targets for which ligands are known. Previously, we used SEA to discover off-targets and mechanism-of-action targets for over 30 drugs against over 40 targets (10, 13). In this strategy, a likely target for a drug is identified when the known ligands for a target resemble the "bait" drug much more closely than would be expected at random; over 2,500 molecular targets have enough ligands to be interrogated by this method. Extending this work, we screened almost 1,000 target "orphan" drugs, finding for many high-likelihood predictions of sensible targets. Surprisingly, it often happened that these predicted targets were already reported in the literature, even though they were unknown to the drug-target databases; only rarely did the target orphan status of a drug survive close scrutiny. Still, several did, and for these we tried to predict and experimentally test sensible primary targets. Also, with so many drugs well associated with targets, we thought to turn this approach on its head and identify those drugs for which targets and specificities were well-enough known to act as biological probes, a subject of much current interest in the community. We considered the application of this approach to investigating mechanism-of-action targets for new molecules emerging from phenotypic programs in drug discovery, and for the relatively rapid identification of drugs and drug candidates that may be used as biological probes.

Results

Finding Primary Targets for Drugs. Our initial goal was to find molecular targets for drugs for which the primary target was

Author contributions: B.K.S. designed research; E.G.-P., V.S., J.H., and B.A.C. performed research; E.G.-P., J.H., J.J.I., E.L., L.M., B.L.R., and B.K.S. analyzed data; and E.G.-P. and B.K.S. wrote the paper.

The authors declare no conflict of interest..

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]E.G.-P. and V.S. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: bryan_roth@med.unc.edu or shoichet@cgl.ucsf.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1204524109/-/DCSupplemental.

unknown. We began with the 1,382 approved drugs from the DrugBank database. From the 127 drugs with an unknown primary target, 44 were discarded as topically applied, acting via a nonprotein target (e.g., chelators) or as diagnostics, while 5 were discarded for having a molecular weight greater than 1,000 daltons. This yielded 78 approved drugs, 5.6% of those in DrugBank, a figure consistent with the literature (1).

To discover molecular targets for these 78 drugs, we looked for chemical similarities to ligand sets for over 2,500 targets, using SEA (10, 14). SEA describes each target by its known ligands, as represented by topological fingerprints (here extended connectivity fingerprints [ECFP] ECFP_4 (15)). The query molecule is compared to the set of known ligands for a target by summing all pair-wise similarities. This raw score is normalized for the size bias by comparing it to a score expected at random for similar set sizes, using the basic local alignment search tool (BLAST) algorithms (16). The final scores, as in BLAST, are expectation values (E-values); the smaller this E-value, the more significant the drug-target association.

For 10 of the 78 drugs, the SEA-predicted target exactly matched one of the chemical libraries we interrogated. For 60 of the remaining 68 drugs, E-values ranged from 9.84×10^{-6} to 1.85×10^{-251}. We selected those predictions that were related to the drug's indication. For 19 drugs, the predicted targets were confirmed by literature (Table S1). A mechanistically sensible target could be found manually in the literature for another 31 drugs for which it could not be predicted using SEA, leaving only 19 drugs (1.51%) for which a reasonable therapeutic target could not be identified.

Before submission of this paper, DrugBank was updated to version 3 (17), resulting in a more comprehensive list of targets. Many of the drugs that previously had no target had one annotated in the new version; only nine drugs had a target found in the literature but not in DrugBank (Table S2). Reassuringly, in 53 cases these targets corresponded to those predicted by SEA and confirmed in the literature. Intriguingly, two of the new DrugBank targets differed from the SEA predictions. Here, the SEA predictions seem more consistent with the drug's indication. Thus, olsalazine is annotated to interferon-γ and thiopurine S-methyltransferase. However, olsalazine is a prodrug of mesalamine, active on cyclooxygenase 1 and 2 and arachiodonate 5-lipoxygenase, targets predicted by SEA and widely thought to be related to the anti-inflammatory indication for olsalazine and mesalamine. Conversely, thiopurine *S*-methyltransferase is most likely an off-target of the drug (18). Similarly, DrugBank associates silver sulfadiazine with DNA as a target, but the target of the core molecule is almost certainly dihydropteroate synthetase (19), as predicted chemoinformatically. The results of the updated annotations from DrugBank and our own SEA-guided literature searching (Fig. 1) left only 13 drugs for which a sensible primary target could not be determined (Table S3). These are considerably fewer than might have been anticipated from the previous literature.

Because we were most interested in finding targets for drugs that truly lacked them, we also considered the 1,431 worldwide-approved drugs from the MDL Drug Data Report (MDDR). After combining the information already present in MDDR and ChEMBL database, the initial number of compounds with no known protein target was 352. Here too, we used SEA to suggest targets that could be verified in the literature. A sensible molecular target was found for 308 compounds, leaving a set of 41 approved drugs for which either no target was predicted or the prediction could not be confirmed (3%) (Table S4).

Finally, we also analyzed the drugs in the NCGC Pharmaceutical Collection (NPC) v1.1.0 (20) (Fig. 2). There, we extracted the set of human-approved drugs, excluding those in DrugBank v3. This yielded a total of 6,554 entries. From these, 5,339 were discarded for being, for instance, pharmaceutical aids (233) or for

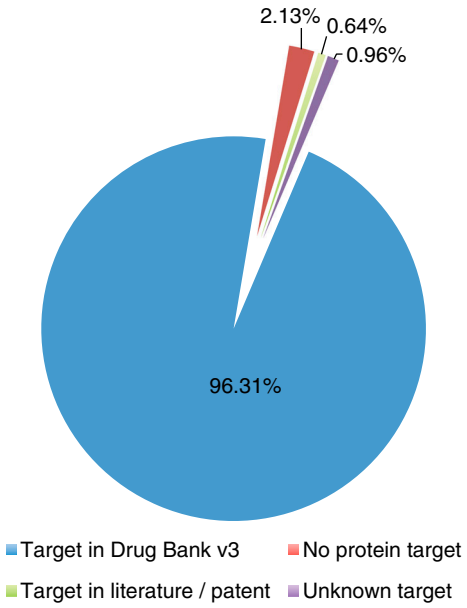

Fig. 1. Distributions of the DrugBank v3 approved drugs by to whether they have a database-assigned target (blue), a literature one (green), act by a nonprotein target (red), or if their molecular target is unknown (purple).

simply lacking an indication (4,323). For the remaining 1,259 compounds, 420 were found in DrugBank—336 as approved drugs. Thus, our final set consisted of 839 worldwide-approved drugs, from which 208 had a protein target annotation in the NPC, 70 in ChEMBL, and 5 in MDDR. We then mined the literature for known therapeutic targets for the remaining 556 drugs, again guided by SEA. This revealed targets for 340 of the drugs, all of which were checked for consistency with the drug's indication. For 121 drugs (36%), the target was predicted using SEA and then confirmed by literature search. For 219 drugs (64%) the target was not predicted by SEA but was found in the literature.

Relevant Targets for Seven Drugs. For seven worldwide-approved drugs that continued to lack a therapeutically relevant target, we were able to find and test in vitro a therapeutically relevant target (Table 1, Fig. 3, and Fig. S1). Three of these drugs—cloperastine, clemastine, and nepinalone—are antitussives. Although cloperastine was discovered based on phenotypic experiments on animals looking for antihistamine response (21), histaminergic activity explains a side effect, somnolence, but not its activity in cough. SEA found three interesting targets for cloperastine. As expected, histamine H1 and H3 were predicted with E-values of 3.28×10^{-8} and 6.21×10^{-24}, respectively. It also predicted activity on σ receptors (E-value $= 2.03 \times 10^{-12}$), of which subtype 1 has previously been associated with cough (22). Since the affinity

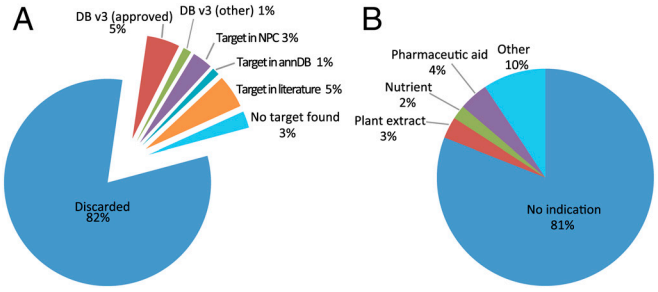

Fig. 2. Distribution of the human-approved drugs set from NPC v1.1.0. (*A*) All human-approved drugs set excluding DrugBank. (*B*) Characteristics of NPC drugs were discarded.

Table 1. Experimentally confirmed primary target predictions for approved drugs.

Drug name	Indication	*E*-value	Predicted target	K_i (nM)
Cloperastine	Cough suppressant	2.03×10^{-12}	σ receptor	σ1 20 σ2 900
Nepinalone		3.28×10^{-8} 6.21×10^{-24}	Histamine H1 receptor Histamine H3 receptor	4 2,148
	Cough suppressant	1.93×10^{-17}	σ receptor	σ1 30 σ2 405
Clemastine	Cough suppressant	N/A	σ receptor	σ1 67 σ2 15
Benzquinamide	Antiemetic Antipsychotic	2.04×10^{-19}	Adrenoceptor α_2 receptor	α_{2A} 1,365 α_{2B} 691 α_{2C} 545
Lobenzarit		N/A	Dopamine receptor	D2 4,369 D3 3,592 D4 574
	Anti-inflammatory	1.8×10^{-5}	COX-2	128,000*
Cyclobenzaprine	Muscle relaxant	N/A	Muscarinic M1 receptor Muscarinic M2 receptor Muscarinic M3 receptor	25 60 6
Nefopam	Analgesic	2.95×10^{-5} 2.95×10^{-5} 2.95×10^{-5} 7.04×10^{-4} N/A 1.10×10^{-1}	Serotonin receptor 2A Serotonin receptor 2B Serotonin receptor 2C Dopamine transporter Serotonin transporter Norepinephrine transporter	1,685 330 56 531 29 33

*Indicates IC_{50} value instead of K_i.

of cloperastine for histamine receptors could not be found in the literature, all four predictions were tested in vitro by ligand displacement assay. The observed K_i values were 3.8 nM for H1, 2,148 nM for H3, 20 nM for σ1, and 900 nM for σ2 (Fig. 3).

Similarly, clemastine is reported to be a selective antihistamine and anticholinergic agent (23). These targets explain its sedating and antipruritic effects, but its antitussive effect remains unexplained. Given their high chemical similarity and the cross-pharmacology between the histaminergic and σ receptors, we tested clemastine against the σ1 and σ2 receptors in vitro, observing K_i values of 67 and 15 nM, respectively. The third antitussive, nepinalone, was also predicted to act via σ, with an E-value of 1.93×10^{-17}. This prediction was confirmed in vitro with K_i values of 30 nM for σ1 and 404.6 nM for the σ2 receptor.

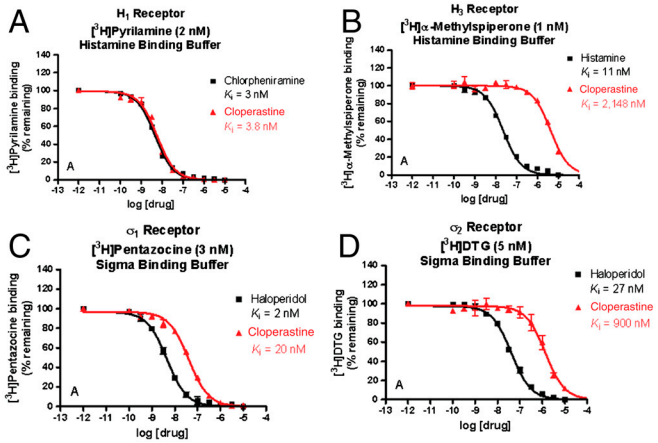

Fig. 3. Dose-response curves for prospective primary target predictions of cloperastine (*A*) Histamine H1. (*B*) Histamine H3. (*C*) σ1. (*D*) σ2.

Lobenzarit is an immunomodulator for which no target is known. SEA predicts cyclooxygenase-2 (COX-2) with an E-value of 1.8×10^{-5}. Although the two anions in lobenzarit are unprecedented among COX-2 ligands, in vitro testing confirmed this target with an IC_{50} value of 128 μM. Whereas this appears high for a drug, we note that lobenzarit's maximum plasma concentration (C_{max}) is 70 μM (24). This IC_{50}/C_{max} ratio is not unprecedented among COX inhibitors: indomethacin has a C_{max} of 3 μM and an IC_{50} of 180 μM, whereas even aspirin has a C_{max} of 1.11 mM and an IC_{50} of 18 mM.

Nefopam is a widely used antinociceptive with safety advantages over the opiates and the NSAIDs, although its molecular targets remain uncertain. Whereas phenotypic patterns in animal models have suggested activities via the serotonin, glutamate, and dopamine circuits (25–28), these suggestions and experiments were made and undertaken before direct assays were available, except as in brain homogenates. For instance, though evidence from classical pharmacology supports a role for the monoamine transporters in nefopam's activity (25, 29), molecular binding affinities and selectivities remain unknown. Whereas chemoinformatic inference finds little similarity to glutaminergic ligands, SEA does predict serotonergic activity (E-value = 1.40×10^{-26}).

Though its prediction against the dopamine transporter (DAT) was less substantial (E-value = 7.04×10^{-4}), the known cross activity between the two targets and the speculation from the classical literature led us to test both predictions. Nefopam had K_i values for 5-HT_{2A}, 5-HT_{2B}, and 5-HT_{2C} of 1,685, 329.5, and 56 nM, respectively. Similarly, it had a K_i of 531 nM against DAT. This, in turn, led us to test the molecule against the norepinephrine (NET) and serotonin (SERT) transporters, where its K_i values, at 33 and 29 nM, were more substantial still. Conversely, whereas a role for dopamine receptors has also been mooted (28), we observed no activity against either the D1 and D2 subtypes. As nefopam achieves plasma concentrations of 48.5 to 183.1 nM (30), these results support a role for the three aminergic transporters (SERT, NET, and DAT) and the $5HT_2$ receptors for the drug's analgesic activity. As with other antinociceptives targeting transporters, like duloxetine, nefopam's efficacy may relate to the ratio of its transporter K_i values, with its serotonin-norepinephrine reuptake inhibition (SNRI) slightly over 10-fold more potent than against the dopamine transporter, which it nevertheless seems to engage.

For other drugs, such as benzquinamide, targets are not only unknown but are known wrong. It is accepted in the field and reported in DrugBank that the primary targets for this drug are the histamine H1 and the muscarinic M1-5 receptors. However, there is no direct experimental evidence of this activity in the literature, despite much assertion. By SEA, the drug was much more similar to the ligand set of the α_{2A} adrenergic receptor (E-value = 2.04×10^{-19}), with no significant similarity to the histamine H1 or any muscarinic receptor (M1–5) ligand sets. Upon experimental testing, no substantial modulation of any of the H1 or M1–5 receptors was observed (maximum inhibition of 16% at 10 μM). Conversely, and consistent with prediction, benzquinamide did bind to the α_{2A}, α_{2B}, and α_{2C} adrenergic receptors (α2-AR) with K_i values of 1,365, 691, and 545 nM, respectively. This activity may partially explain the anxiolytic activity effect of the drug (31). Although there are studies linking the α2-AR and emesis (32), this is not an established target for this indication. We therefore adopted a target-hopping strategy, looking for nausea-related targets with strong chemoinformatic associations with α2-AR. This led us to the dopamine D2 receptor, which by ligand-set similarity resembles α2-AR (13) and is an accepted target for emesis. On testing benzquinamide on the D2, D3, and D4 receptors, we observed K_i values of 3,964, 3,592, and 574 nM, respectively. Notwithstanding the fact that the α2-AR values are lower than the D2 values, it is the D2 activity that may be the most relevant for emesis.

This same target-hopping strategy was used for the muscle relaxant cyclobenzaprine. Though first developed as an antidepressant and an antipsychotic, the drug is mostly used today as a muscle relaxant, and none of the drug's known targets, such as transporters, are consistent with this indication. As with benzquinamide, SEA did not directly suggest a target that was closely linked with muscle relaxation. The method did, however, predict H1 histaminergic activity for cyclobenzaprine, with an E-value of 2.21×10^{-56}, which has recently been tested and confirmed with a K_i value of 21 nM (33). By ligand similarity, the H1 receptor is associated with muscarinic receptors, which are well-accepted targets for muscle relaxation. Consistent with this prediction, cyclobenzaprine had K_i values of 25.0, 60.0, and 6.3 nM for the muscarinic M1, M2, and M3 receptors. Given that the drug has maximum plasma concentrations of 16 to 31 nM and an AUC of 325 to 647 nM·h, this affinity is consistent with the drug's role.

Drugs as Probes. An unexpected lesson of this work is that most drugs do in fact have reasonable primary molecular targets, and so we wondered if this point could be taken a step further: For how many drugs is potency and specificity well-enough known to identify them as chemical probes of biology? The distinction between drugs, which must be therapeutically efficacious but can often be promiscuous, and probes, which must have a high-fidelity to their targets but can be therapeutically ineffective, has been well established (34, 35). We wondered, nevertheless, whether some drugs might have the specificity requirements to meet the common criteria of probes (36). If so, they might find wide utility, not least because the burden of developing new, genuinely effective probes has been high (37). We began with the NIH Molecular Libraries Screening Centers Network (MLSCN) definition of a biological probe (38), a small molecule with activity on a target of 100 nM or better and at least 10-fold selectivity over any related target. We found over 100 drugs that fit these criteria, acting on 41 targets. These covered the major protein families including enzymes, ion channels, nuclear hormone receptors, and transporters. Notwithstanding the well-known promiscuity of GPCRs, we nevertheless found 30 drugs that could be used as probes for 17 members of this family with at least 10-fold selectivity to any other member of the family for which they have been tested.

Despite being the NIH standard, a 10-fold selectivity cutoff may be too permissive, so we increased the selectivity threshold to 100-fold. We further insisted that any drug be cell- and typically tissue-penetrant and used SEA to calculate potential off-targets. This found a smaller but higher-confidence set in which all the drugs for which a related target could be chemoinformatically predicted with an E-value of 10^{-5} or lower (better) were removed. Pregabalin, for example, was discarded because it is predicted by SEA to hit the glycine receptor, a target closely related to its presumed target, the voltage-gated calcium channel.

This high-confidence set consists of 40 drugs that can serve as probes for 25 unique protein targets (Table S5). This is almost certainly only a subset of a full list of high-specificity probe-like drugs, as we have made no effort to be comprehensive here. Still, these 40 probe-like drugs illustrate how such active molecules might be found. Many, like selective serotonin reuptake inhibitors (SSRIs), are designed and checked for selectivity versus other monoamine transporters and receptors. For SSRIs, inhibiting α1 adrenergic, muscarinic, and histamine H1 receptors leads to undesired side effects, so selectivity is closely associated with compound progression. Thus, fluvoxamine is 500-fold selective versus the next-best target, the muscarinic M1 receptor (39), and lacks substantial SEA E-values to other related targets (Fig. 4). Similarly, the proteasome inhibitor bortezomib was advanced due not only to its potency but also for its selectivity profile over antitargets such as chymotrypsin and elastase (40) (Fig. 4). Similarly, anastrozole (Fig. 4) is a selective aromatase inhibitor that has little activity against other cytochromes. A very

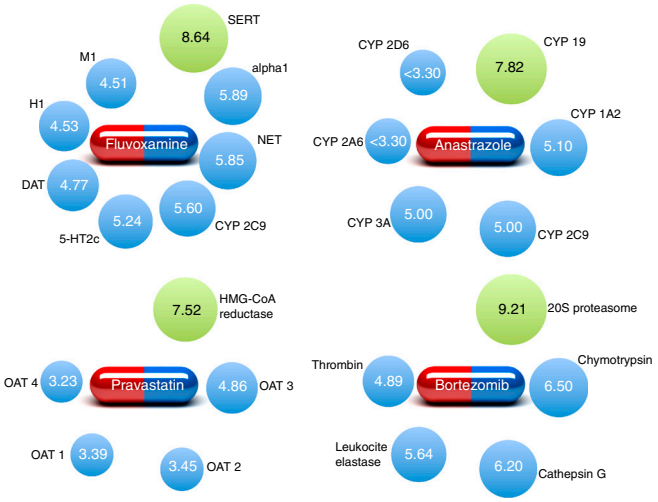

Fig. 4. Example drugs that can be used as chemical probes. Activity is shown as a circle proportional to the pKi or pIC_{50}. The main target is marked in green.

different example is that of pravastatin, a statin that has not been tested for off-targets responsible of undesired side effects, but for activity against organic anion transporting polypeptides (OAT 1–4), the transporters responsible for the drug's active transport into hepatocytes (Fig. 4).

Occasionally, the protein target for which the drug can be used as a probe is not its therapeutic target. Thus, mofezolac (Disopain®), like most NSAIDs before the introduction of the coxibs, has activity on both COX-2, its primary mechanism-of-action target, and COX-1, the off-target to which its gastrointestinal side effects have been attributed. Unlike most NSAIDs, and far different from the coxibs, mofezolac is actually *specific* for COX-1, with an IC_{50} of 1.4 nM for this target and an IC_{50} of only 440 nM for COX-2. Whereas specificity for what is commonly considered a toxic off-target may seem perverse, for a tool it may help deconvolute the pathways modulated by the COX-2 specific and the nonspecific NSAIDs, and may better elucidate the roles and even possible therapeutic modalities of COX-1.

Discussion

Four key observations emerge from this study. First, the number of drugs with unknown primary targets can be reduced from the relatively high figures widely accepted in the field to 1%, in the case of approved drugs included in DrugBank. Second, we find that chemoinformatic methods are useful guides to at least first-pass efforts to assign drug or candidate targets when these are unknown. Third, we were able to chemoinformatically predict, and confirm in vitro, therapeutically relevant targets for seven drugs where the primary targets were genuinely unknown. Fourth, many approved drugs can qualify as chemical probes—highly sought tools in biological studies.

It was surprising how often likely primary targets for drugs that supposedly lacked them were found in the literature. In contrast to the much higher figures often implied (1, 2), and what seem to be common conceptions in the field, only 4% of worldwide drugs in the MDDR, only 3% of drugs in the NPC database, and less than 1% of the approved drugs in DrugBank lack a well-established target in the literature. This discrepancy reflects a simple lack of annotation in the databases, something largely corrected, at least for DrugBank, in its most recent update. What does seem true is that chemoinformatic prediction, using SEA or related methods (11, 12), can rapidly fill these gaps for established drugs and for new investigational molecules that may emerge from nontarget-based drug discovery, such as phenotypic screening (41).

Of course, a more compelling illustration of this approach is the prospective prediction of therapeutically sensible targets for drugs. Both by direct drug-target association and by target-hopping, the chemoinformatic method identified plausible mechanism-of-action targets for seven approved but until now target-orphan drugs. Especially interesting are the identification of the σ1 receptor as the target for three cough suppressant drugs (cloperastine, nepinalone, and clemastine). Although σ1 has a checkered history in pharmacology (42), and is notorious for its promiscuity, perhaps the only physiology to which it is reliably linked is cough, and so the 20 to 67 nM affinities of these drugs may illuminate their mechanism of action and that of other drugs in this class. Similarly, the association of benzquinamide clarifies what had been a widespread (17) but false association with muscarinic and histaminergic signaling, for which we found no evidence in vitro. These results, and those with lobenzarit and cyclobenzaprine, support a role for chemoinformatic association in target identification for leads emerging from phenotypic or other nontarget-based methods.

In undertaking this study, we initially hoped to discover primary targets for drugs with unknown targets. As we learned, there are in fact few drugs with truly unknown targets. We wondered if this lack of drug-target association also affects the identification of those drugs that are truly target specific, a quality important for their use as biological probes. We therefore considered an inversion of our original hypothesis: Are there many drugs that are likely so specific, potent, and penetrant that they may act as biological probes, notwithstanding the well-recognized arguments that drugs often are poor probes (34, 43)? Using the National Institutes of Health Molecular Libraries Probe Production Centers criteria for a probe (38) (at least 100 nM activity and 10-fold specificity against related targets), we found over 100 qualifying drugs. Because a 10-fold selectivity may be too permissive, we then considered drugs with at least a 100-fold selectivity and further insisted that no related target have a SEA E-value better (less) than 10^{-5} for the putative tool. These more stringent criteria culled the list to 40 tool drugs for 25 targets. This is likely a lower limit as we have made no attempt to be comprehensive, and many drugs that may meet these criteria were not considered. Our point is that, in fact, there are many drugs that meet the criteria for probes, and many of these may be overlooked, notwithstanding their outstanding physical and biological properties that are, after all, substantially better than all but the most highly developed probes. Given the high cost and difficulty to develop a probe for any given protein, having access to a set of high-quality chemical compounds such as approved drugs may be broadly useful to the field.

Despite these efforts, there remains a small but important group of drugs to which we could not assign reasonable targets, such as metformin, among the world's most highly prescribed molecules. Even for the drugs for which we could find a relevant target in the literature, only 36% of them were initially guided by SEA, while the rest did not have relevant predictions and the targets were found through manual literature search. This reflects the limitations of the chemoinformatics method, which is essentially inferential. When targets are poorly annotated for ligands, or when the drug does not share chemotypes with the ensemble of annotated ligands, SEA will miss the association. More subtly, demonstrating a target does not demonstrate a mechanism of action. Establishing target-engagement mechanistically in vivo remains difficult, ultimately demanding full animal physiology, and even then unanticipated modalities may play a role (witness the evolving understanding of even a recent drug such as imatinib (44, 45)). Thus whereas the targets we predict and test are reasonable, based on in vitro binding and a known biological role for the protein, their status as mechanism-of-action targets is, for now, no more than reasonable. Also, many drugs exert their effect through the modulation of several targets, and so identifying one protein target for a drug may leave others important for its mechanism unknown.

These caveats should not obscure the central themes of this study. In the teeth of current opinion, we found that few drugs lack reasonable primary targets, at least among domain experts. The ability of chemoinformatics methods to rapidly identify targets, and to predict those which can be experimentally confirmed, supports the idea that this and related methods (11, 12, 46–48) will be useful in target identification for drugs that emerge from nontarget-based methods. Even in this molecular era of drug discovery, there remain new investigational drugs whose molecular targets are unclear, restricting their optimization and broad use in disease. The same chemoinformatic strategy may be used to illuminate those drugs that can, in fact, be used as biological probes, meeting and typically surpassing the specificity, potency, and biological penetrance criteria now current in the field. Given the resources devoted to probe discovery and development on a national level, it may be estimated that between $1 million to $2 million are required to develop a new probe molecule. There may be a substantial number of overlooked probes that already exist in our drug armentarium, and these may be deployed even today, at little expense, to illuminate biology.

Methods

Test Sets. Four collections of drugs with annotated targets were used: the approved drugs set of DrugBank version 2 (1,382 FDA approved drugs); the approved drugs set of DrugBank version 3 (1,410 drugs); the launched drugs set of MDDR version 2006.1 (1,431 drugs); and the human approved drugs set excluding those in DrugBank v3 of NPC v.1.1.0 (6,554 drugs). For each of these databases, the drugs that do not have an associated target have been taken as the initial set of "drug with unknown protein target." These have been further explored by SEA (10, 14) and by a literature search to get the final set of drugs with no known targets.

Ligand-Based Virtual Screening. Each drug was computationally screened against 2,521 ligand-target sets with activity of 10 μM or better from the ChEMBL 09 (49) and MDDR databases. The virtual profiling was performed using SEA, as described (10). More detail on the SEA calculations can be found in *SI Methods*.

Known Target Assignment. See the *SI Methods*.

Experimental Testing. See *SI Methods* and Table S6.

ACKNOWLEDGMENTS. We thank Dr. Mark Burlingame for purifying nepinalone and M. Keiser, N. Hodge, C. Laggner, and A. Doak for reading this manuscript. Supported by National Institutes of Health Grants GM71896 (to J.I. and B.K.S.), AG002132 (to S.P. and B.K.S.), the National Institutes of Mental Health Psychoactive Drug Screening Program (B.L.R.), CA89450 (to L.M.), and the Fulbright and Generalitat de Catalunya Postdoctoral Fellowships (E.G.-P.).

1. Drews J (2000) Drug discovery: A historical perspective. *Science* 287:1960–1964.
2. Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? *Nat Rev Drug Discov* 5:993–996.
3. Swinney DC, Anthony J (2011) How were new medicines discovered? *Nat Rev Drug Discov* 10:507–519.
4. Dewar AL, et al. (2005) Macrophage colony-stimulating factor receptor c-fms is a novel target of imatinib. *Blood* 105:3127–3132.
5. Roth BL, Kroeze WK (2006) Screening the receptorome yields validated molecular targets for drug discovery. *Curr Pharm Des* 12:1785–1795.
6. David AD (2002) Mining the human kinome. *Drug Discov Today* 7:1121–1123.
7. Mayer TU, et al. (1999) Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* 286:971–974.
8. Harding M, Galat A, Uehling D, Schreiber S (1989) A receptor for the immunosuppressant FK506 is a cis-trans peptidyl-prolyl isomerase. *Nature* 341:758–760.
9. Leung D, Hardouin C, Boger DL, Cravatt BF (2003) Discovering potent and selective reversible inhibitors of enzymes in complex proteomes. *Nat Biotechnol* 21:687–691.
10. Keiser MJ, et al. (2007) Relating protein pharmacology by ligand chemistry. *Nat Biotechnol* 25:197–206.
11. Gregori-Puigjané E, Mestres J (2008) A ligand-based approach to mining the chemogenomic space of drugs. *Comb Chem High Throughput Screen* 11:669–676.
12. Nidhi , Glick M, Davies JW, Jenkins JL (2006) Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J Chem Inf Model* 46:1124–1133.
13. Keiser MJ, et al. (2009) Predicting new molecular targets for known drugs. *Nature* 462:175–181.
14. Hert J, Keiser MJ, Irwin JJ, Oprea TI, Shoichet BK (2008) Quantifying the relationships among drug classes. *J Chem Inf Model* 48:755–765.
15. Rogers D, Hahn M (2010) Extended-connectivity fingerprints. *J Chem Inf Model* 50:742–754.
16. Pearson WR (1998) Empirical statistical estimates for sequence similarity searches. *J Mol Biol* 276:71–84.
17. Knox C, et al. (2011) DrugBank 3.0: A comprehensive resource for "Omics" research on drugs. *Nucl Acid Res* 39(suppl 1):D1035–D1041.
18. Szumlanski C, Weinshilboum R (1995) Sulphasalazine inhibition of thiopurine methyltransferase: Possible mechanism for interaction with 6-mercaptopurine and azathioprine. *Br J Clin Pharmacol* 39:456–459.
19. McCullough JL, Maren TH (1974) Dihydropteroate synthetase from plasmodium berghei: Isolation, properties, and inhibition by dapsone and sulfadiazine. *Mol Pharmacol* 10:140–145.
20. Huang R, et al. (2011) The NCGC pharmaceutical collection: A comprehensive resource of clinically approved drugs enabling repurposing and chemical genomics. *Sci Transl Med* 3 80ps16.
21. Takagi K, Flucuda H, Fuije K, Matsui K, Sato M (1961) Studies on antitussives. IV. Various drugs and ω-(diphenilmetoxy) alkylamine compounds. *J Pharm Soc Jpn* 81:261.
22. Kamei J, Iwamoto Y, Misawa M, Kasuya Y (1993) Effects of rimcazole, a specific antagonist of [sigma] sites, on the antitussive effects of non-narcotic antitussive drugs. *Eur J Pharmacol* 242:209–211.
23. Kubo N, Shirakawa O, Kuno T, Tanaka C (1987) Antimuscarinic effects of antihistamines: Quantitative evaluation by receptor-binding assay. *Jpn J Pharmacol* 43:277–282.
24. Cynshi O, et al. (1990) Anti-oxidative profile of lobenzarit disodium (CCA). *Biochem Pharmacol* 40:2117–2122.
25. Fuller RW, Snoddy HD (1993) Evaluation of nefopam as a monoamine uptake inhibitor in vivo in mice. *Neuropharmacology* 32:995–999.
26. Marazziti D, Rotondo A, Ambrogi F, Cassano G (1991) Analgesia by nefopam: Does it act through serotonin? *Drugs Exp Clin Res* 17:259–261.
27. Verleye M, Andre N, Heulard I, Gillardin J-M (2004) Nefopam blocks voltage-sensitive sodium channels and modulates glutamatergic transmission in rodents. *Brain Res* 1013:249–255.
28. Esposito E, Romandini S, Merlo-Pich E, Mennini T, Samanin R (1986) Evidence of the involvement of dopamine in the analgesic effect of nefopam. *Eur J Pharmacol* 128:157–164.
29. Tresnak-Rustad NJ, Wood ME (1981) In vitro biochemical effects of nefopam hydrochloride, a new analgesic agent. *Biochem Pharmacol* 30:2847–2850.
30. Aymard G, et al. (2003) Comparative pharmacokinetics and pharmacodynamics of intravenous and oral nefopam in healthy volunteers. *Pharmacol Toxicol* 92:279–286.
31. Schramm NL, McDonald MP, Limbird LE (2001) The alpha2A-adrenergic receptor plays a protective role in mouse behavioral models of depression and anxiety. *J Neurosci* 21:4875–4882.
32. Hikasa Y, Ogasawara S, Takase K (1992) Alpha adrenoceptor subtypes involved in the emetic action in dogs. *J Pharmacol Exp Ther* 261:746–754.
33. Loukine E, et al. (2012) Large scale prediction and testing of drug activity on side-effect targets. *Nature*, in press.
34. Frye SV (2010) The art of the chemical probe. *Nat Chem Biol* 6:159–161.
35. Garcia-Serna R, Mestres J (2011) Chemical probes for biological systems. *Drug Discov Today* 16:99–106.
36. Castoreno AB, Eggert US (2010) Small molecule probes of cellular pathways and networks. *ACS Chem Biol* 6:86–94.
37. Edwards AM, Bountra C, Kerr DJ, Willson TM (2009) Open access chemical and clinical probes to support drug discovery. *Nat Chem Biol* 5:436–440.
38. Oprea TI, et al. (2009) A crowdsourcing evaluation of the NIH chemical probes. *Nat Chem Biol* 5:441–447.
39. Owens MJ, Knight DL, Nemeroff CB (2001) Second-generation SSRIs: Human monoamine transporter binding profile of escitalopram and R-fluoxetine. *Biol Psychiatry* 50:345–350.
40. Adams J, et al. (1998) Potent and selective inhibitors of the proteasome: Dipeptidyl boronic acids. *Bioorg Med Chem Lett* 8:333–338.
41. Laggner C, et al. (2012) Chemical informatics and target identification in a zebrafish phenotypic screen. *Nat Chem Biol* 8:144–146.
42. Maurice T, Su T-P (2009) The pharmacology of sigma-1 receptors. *Pharmacol Ther* 124:195–206.
43. Kodadek T (2010) Rethinking screening. *Nat Chem Biol* 6:162–165.
44. Schindler T, et al. (2000) Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* 289:1938–1942.
45. Breccia M, Muscaritoli M, Aversa Z, Mandelli F, Alimena G (2004) Imatinib mesylate may improve fasting blood glucose in diabetic Ph+ chronic myelogenous leukemia patients responsive to treatment. *J Clin Oncol* 22:4653–4655.
46. Weill N, Rognan D (2009) Development and validation of a novel protein-ligand fingerprint to mine chemogenomic space: Application to G protein-coupled receptors and their ligands. *J Chem Inf Model* 49:1049–1062.
47. Bottegoni G, Kufareva I, Totrov M, Abagyan R (2008) Four-dimensional docking: A fast and accurate account of discrete receptor flexibility in ligand docking. *J Med Chem* 52:397–406.
48. Grigoryan A, Kufareva I, Totrov M, Abagyan R (2010) Spatial chemical distance based on atomic property fields. *J ComputAided Mol Des* 24:173–182.
49. Gaulton A, et al. (2012) ChEMBL: A large-scale bioactivity database for drug discovery. *Nucl Acids Res* 40:D1100–1107.