Research Article

Many Approved Drugs Have Bioactive Analogs With Different Target Annotations

Ye Hu,¹ Eugen Lounkine,^{1,2} and Jürgen Bajorath^{1,3}

Received 10 April 2014; accepted 12 May 2014; published online 29 May 2014

ABSTRACT. Close structural relationships between approved drugs and bioactive compounds were systematically assessed using matched molecular pairs. For structural analogs of drugs, target information was assembled from ChEMBL and compared to drug targets reported in DrugBank. For many drugs, multiple analogs were identified that were active against different targets. Some of these additional targets were closely related to known drug targets while others were not. Surprising discrepancies between reported drug targets and targets of close structural analogs were often observed. On one hand, the results suggest that hypotheses concerning alternative drug targets can often be formulated on the basis of close structural relationships to bioactive compounds that are easily detectable. It is conceivable that such obvious structure–target relationships are frequently not considered (or might be overlooked) when compounds are developed with a focus on a primary target and a few related (or undesired) ones. On the other hand, our findings also raise questions concerning database content and drug repositioning efforts.

KEY WORDS: approved drugs; bioactive compounds; drug targets; matched molecular pairs; polypharmacology; structural relationships.

INTRODUCTION

Over the past decade, increasing evidence has accumulated that drugs often elicit their therapeutic effects through interactions with multiple targets, which is rationalized as polypharmacology (1–4). The notion of polypharmacology has strong implications for drug development in some (but not all) therapeutic areas (4, 5) and has also catalyzed significant efforts to better understand and predict side effects of drugs (6, 7) or reposition drugs for alternative therapeutic applications (8). Given the increasing popularity of polypharmacology, a variety of computational methods have been developed or adapted for target analysis or prediction (9).

Compound promiscuity, defined as specific interactions of a small molecule with multiple targets, represents the molecular basis of polypharmacological effects (10–12).

Electronic supplementary material The online version of this article (doi:10.1208/s12248-014-9621-8) contains supplementary material, which is available to authorized users.

¹Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität Bonn, Dahlmannstr. 2, D-53113, Bonn, Germany.

²Novartis Institutes for Biomedical Research, In Silico Lead Discovery, 250 Massachusetts Ave., Cambridge, Massachusetts 02139, USA.

³To whom correspondence should be addressed. (e-mail: bajorath@bit.uni-bonn.de)

Target annotations have been assembled from various databases to estimate the degree of promiscuity among different types of compounds including screening hits, bioactive compounds, and drugs (10–14). On the basis of initial estimates, drugs interacted on average with ~2.7 targets and with ~6.3 targets when predicted interactions were also taken into account (14). It has also been estimated that more than 50% of current drugs interact with >5 targets and that drugs active against G protein coupled receptors (GPCRs) have on average six to seven targets (10). However, analysis of target annotations on the basis of high-confidence activity data indicates that bioactive compounds are generally less promiscuous than often thought (11–13). For example, bioactive compounds from ChEMBL (15) were found to be on average only active against one to two targets and confirmed screening hits from PubChem (16) only against two to three targets (12, 13). Similar observations were made for experimental drugs taken from DrugBank (17), which were on average active against ~1.8 targets, whereas approved drugs interacted with on average ~6 targets (13). By contrast, ChEMBL compounds active against target families generally assumed to be promiscuous, *i.e.*, protein kinases and class A GPCRs, were annotated with on average only one to two targets on the basis of high-confidence activity data (13).

It is evident that data selection criteria and the types of experimental measurements that are considered significantly affect the results of promiscuity analysis. As a representative example, for the drugs milrinone and buspirone, ChEMBL version 14 reported a total of 42 and 72 potential targets, respectively, whereas these drugs were only annotated with 1

and 2 targets in DrugBank, respectively (18), which represents a significant discrepancy. However, when only ligand–target interactions at the highest confidence level in ChEMBL and (assay-independent) equilibrium constants as activity measurements were considered, the number of target annotations for milrinone and buspirone was dramatically reduced to one and two, respectively, consistent with the drug target annotations in DrugBank (18).

In addition to analyzing target annotations and compound or drug promiscuity, structural relationships between drugs and other active compounds have also been intensely studied, for example, to assess drug-likeness (19, 20), search for privileged substructures (21), or facilitate activity predictions (22, 23). Such analyses have often been carried out at the level of molecular frameworks or scaffolds (24) extracted from bioactive compounds or drugs (25). For example, in a comprehensive analysis of drug scaffolds extracted from approved drugs in DrugBank, it was found that each drug scaffold represented on average ~1.8 approved drugs and that ~79% of all drug scaffolds represented only a single drug (26). Moreover, when drug scaffolds were systematically compared to scaffolds extracted from bioactive compounds, 221 of 700 drug scaffolds were not detected in bioactive compounds (26). Scaffolds have also been associated with polypharmacology. In bioactive compounds, >400 scaffolds were identified to represent compounds active against targets from at least two different target families. A subset of 83 of these scaffolds was found to be active against targets from 3 to 13 families, and 17 of these promiscuous scaffolds were detected in >200 approved drugs (27).

To set the stage for our current analysis, we summarize some of the trends discussed above: On the basis of currently available data, bioactive compounds have a much lower degree of promiscuity than approved drugs. In addition, the most promiscuous frameworks from bioactive compounds are well represented in approved drugs, whereas many scaffolds exclusively occur in drugs. Taken together, these findings would indicate that structural relationships between drugs and bioactive compounds are well explored and that it might be difficult to deduce additional target annotations for drugs from bioactive compounds.

In light of the above, we have set out to address the following question: How well do target annotations of approved drugs and bioactive compounds agree if we examine narrowly confined chemical space around drugs; in other words, if we only consider structural analogs of approved drugs? To these ends, we have systematically identified bioactive compounds structurally most similar to drugs and compared target annotations of drugs and their bioactive analogs. How might these target annotations compare? A reasonable expectation can be formulated by considering that structural analogs often act against the same target(s) and that approved drugs generally display a higher degree of promiscuity than bioactive compounds. Accordingly, one might anticipate that structural analogs of approved drugs might often be active against the same set or a subset of drug targets. By contrast, in our current analysis, a different picture has emerged. We have detected many examples of approved drugs for which multiple analogs were annotated with additional and often unrelated targets. These findings were surprising. They suggested additional target hypotheses

for many approved drugs but also raised questions about database content and strategies underlying drug repositioning efforts.

MATERIALS AND METHODS

Data Collection

From ChEMBL (release 17) (15), compounds with direct interactions (*i.e.*, target relationship type “D”) against human targets at the highest confidence level (*i.e.*, target confidence score 9) were assembled. Two types of activity measurements were considered, *i.e.*, equilibrium constants (K_i) and assay-dependent IC_{50} values. In order to ensure a high level of data confidence, only compounds with explicitly defined K_i or IC_{50} values were considered. Approximate measurements such as “>,” “<,” and “~” were discarded. For each qualifying compound, a target profile was generated by collecting all of its available targets.

From DrugBank 3.0 (17), approved small molecule drugs with available structures and activity information were collected. For selected drugs, all available drug action targets, metabolizing enzymes, transporters, and carriers were assembled, thus representing a spectrum of targets similar to ChEMBL.

Matched Molecular Pairs

The matched molecular pair (MMP) concept was applied to systematically investigate structural relationships between approved drugs and bioactive compounds. An MMP is defined as a pair of compounds that only differ at a single site by a substructure change (28), which is termed a chemical transformation (29). Depending on the site at which the chemical transformation takes place, the substructure change could result in a modification in the core structure or substituents. MMP generation according to Hussain and Rea (29) involved the systematic deletion of individual exocyclic single bonds in a compound as well as the deletion of combinations of two or three bonds. The resulting fragments represented a molecular core and substituent. Three transformation size restrictions were introduced to limit the size of substructure changes and preferentially generate MMPs from structural analogs (30). Specifically, a core structure had to be at least twice the size of the exchanged substructure. In addition, the difference in size between two exchanged substructures forming a chemical transformation was limited to at most eight non-hydrogen atoms. Furthermore, the maximal size of an exchanged fragment was 13 non-hydrogen atoms (30). All possible transformation size-restricted MMPs formed between approved drugs and bioactive compounds were calculated using an in-house implementation of the algorithm by Hussain and Rea (29) that utilizes the OpenEye toolkit (31).

Drug Hubs and Potential Drug Targets

Drugs that formed MMPs with at least 10 bioactive compounds were identified and classified as “drug hubs” (based upon a network representation discussed below). The

activity profiles of drug hubs and their bioactive analogs were compared as illustrated in Fig. 1a. The target annotations of drug hubs were extracted from DrugBank. The activity profiles of bioactive analogs were generated by collecting all available target annotations from ChEMBL (on the basis of high-confidence activity data, as specified above). The

ChEMBL targets of bioactive compounds were identified that were not available in DrugBank for the corresponding drug hub. Such targets provided hypotheses for alternative/potential drug targets. For example, in the schematic representation in Fig. 1a, the drug hub on the left was assigned three additional potential drug targets. By contrast, the drug

Fig. 1. Potential drug targets. **a** The identification of potential targets for drug hubs (red nodes) is schematically illustrated. For each drug hub, the activity profile was assembled from DrugBank (shown in red boxes). In addition, target annotations of its bioactive MMP partners (structural analogs; blue nodes) were collected from ChEMBL. The ChEMBL targets that were not available in DrugBank (shown in blue boxes) were considered potential/alternative targets of the drug hub. For example, the drug hub shown on the left was annotated with three targets in DrugBank, i.e., T₁, T₂, and T₃. Its bioactive analogs were active against a total of four targets, i.e., T₂, T₄, T₅, and T₆. Therefore, T₄, T₅, and T₆ were designated as potential/alternative targets of the drug. Accordingly, for the drug shown on the right, no additional ChEMBL target was available. **b** The assessment of potential targets for drug hubs is illustrated. Each drug hub was mapped to ChEMBL. If the drug was available in ChEMBL with high-confidence activity data, the target profile was assembled and compared to the set of potential/alternative targets. In this schematic example, the drug was present in ChEMBL and active against four targets, two of which were potential/alternative targets. Therefore, these two potential targets (i.e., T₄ and T₆) were considered confirmed

hub on the right had no other potential targets because its bioactive analogs were active against a subset of its DrugBank targets.

Furthermore, potential drug targets were assessed as illustrated in Fig. 1b. Drug hubs were mapped to ChEMBL. If high-confidence activity data were available for the drug, its target profile was assembled and compared to the set of potential targets according to Fig. 1a. Potential targets the drug hub was found to be annotated with in ChEMBL were designated as “confirmed targets”. Hence, this protocol identified inconsistencies between drug target annotations in DrugBank and ChEMBL and also “validated” additional drug targets deduced from bioactive analogs at a conceptual level.

RESULTS AND DISCUSSION

Analysis Concept

In this study, we have analyzed narrowly confined chemical space around approved drugs and associated target information with this space. Chemical space was screened by systematically exploring close structural relationships between drugs and bioactive compounds. Accordingly, the study stringently depends on the systematic generation of close structural analogs of approved drugs. The MMP formalism provides the method of choice to limit the detection of structural relationships to close analogs. In addition, it is descriptor independent, similarity metric free, and chemically intuitive. Different computational approaches are available to evaluate structural similarity including, for example, molecular fingerprints, pharmacophore models or shape representations. However, none of these alternative approaches primarily focuses on structural analogs. Rather, fingerprint, pharmacophore, or shape queries are typically employed to identify increasingly diverse structures having similar activity, different from the major goal of our analysis. Therefore, in our study, the assessment of structural similarity was deliberately carried out by calculating all transformation size-restricted “drug-to-bioactive compound” MMPs (in the following simply referred to as MMPs). Hence, for each drug, all MMPs formed with bioactive compounds were determined, yielding a set of structural analogs for each MMP-forming drug. Then, drugs involved in most structural relationships (*i.e.*, forming at least 10 different MMPs) were selected. These drugs, for which narrow chemical space was best explored, were considered drug hubs (in analogy to network terminology). Known targets of drug hubs were compared to targets of their structural analogs. As a control, drugs were also mapped into bioactive compound space and qualifying target annotations were assembled. Finally, targets from bioactive analogs were assigned to drug hubs for which no prior evidence existed in source databases.

Approved Drugs and Bioactive Compounds

From DrugBank, 1,241 approved drugs were collected that were annotated with a total of 1,012 protein targets representing 7,514 distinct drug-target interactions (Table I). From ChEMBL, 133,378 qualifying bioactive compounds with high-confidence activity annotations were obtained,

Table I. Data Sets

Number of	Compound type	
	Bioactive compounds (ChEMBL)	Approved drugs (DrugBank)
Compounds	133,378	1,241
Targets	1,308	1,012
Interactions	204,247	7,514

The number of bioactive compounds and approved drugs, their targets, and interactions are reported

which were active against 1,308 targets, yielding a total of 204,247 interactions (Table I). Thus, 100+ times more bioactive compounds than approved drugs were available. However, bioactive compounds and drugs covered target space of comparable size (*i.e.*, 1,308 vs. 1,012 targets).

Drug-to-Bioactive Compound MMPs

A total of 3,579 MMPs were formed between 617 approved drugs and 2,392 bioactive compounds. Thus, nearly 50% of all approved drugs but only less than 2% of bioactive compounds participated in the formation of MMPs. Approximately half of the approved drugs had no bioactive analogs, despite the large number of bioactive compounds that were evaluated. This finding further emphasized that many drugs are characterized by unique structural features (as discussed above). Nonetheless, on average, ~6 structural relationships were detected per MMP-forming drug, thus delineating narrow chemical space around a subset of approved drugs.

MMP-Based Network

Figure 2 shows an approved drug-to-bioactive compound bipartite network representing all 3,579 MMPs formed by 617 approved drugs and 2,392 bioactive compounds. The topology of the network reveals that a subset of drugs formed MMPs with multiple bioactive analogs, giving rise to characteristic cluster formation, whereas the majority of bioactive compounds formed only a single MMP. On average, a drug was involved in 5.8 MMPs but a bioactive compound only in 1.5 MMPs. Many clusters were formed around individual drugs (centrally located red nodes). The distribution of node degrees for approved drugs is reported in Fig. 3. Approximately 30% of the drugs formed MMPs with a single bioactive compound but the majority of drugs were involved in multiple MMPs. Nearly 30% of all drugs formed MMPs with >5 bioactive analogs and 96 of 617 drugs formed MMPs with 10 or more analogs, representing the largest clusters around individual drugs in Fig. 2.

Drug Hubs

Structures of the 96 approved drugs forming MMPs with at least 10 bioactive analogs were inspected and 17 were found to represent simple pharmaceutically active substances such as, for example, acetic acid, calcium acetate, or aminolevulinic acid. These 17 substances were not further considered. The remaining 79 approved drugs were selected as drug hubs for further analysis. Two drug hubs from the

Fig. 2. MMP-based drug-to-bioactive compound network. MMPs formed between approved drugs a-drugs and bioactive compounds are visualized in a network representation. Red and blue nodes represent approved drugs and bioactive analogs, respectively. Edges indicate the formation of drug-to-bioactive compound MMPs

network and representative bioactive analogs are shown in Fig. 4. Rimobant, an anti-obesity agent, formed MMPs with 96 bioactive analogs (Fig. 4a). In addition, indomethacin, a non-steroidal anti-inflammatory agent, formed 31 MMPs (Fig. 4b). Table II reports the top 20 drug hubs forming MMPs with more than 25 bioactive analogs. Two stereoisomers, vidarabine and adenosine, formed 110 MMPs each, the largest numbers we detected. The drug hubs in Table II were annotated in DrugBank with a variety of targets, ranging from 1 to 29 targets per drug.

Potential Targets

Target profiles of drug hubs and their bioactive analogs were compared, as illustrated in Fig. 1. Targets of bioactive analogs that were not reported for drug hubs in DrugBank were considered alternative/potential drug targets. For the top 20 drug hubs with largest numbers of MMPs, potential targets are reported in Table II. The number of these potential targets per drug ranged from 0 to 61, which did not correlate with the number of designated DrugBank targets or the number of MMPs formed by drug hubs. For example, aliskiren (rank 4 in Table II) was annotated with two DrugBank targets and formed 61 MMPs. All 61 bioactive

analogs were active against renin, *i.e.*, one of two known drug targets of aliskiren. Thus, in this case, no additional target was suggested. By contrast, promazine (rank 7) had 18 DrugBank targets and formed 35 MMPs. There were 35 additional targets of promazine's bioactive analogs that were not reported for the drug in DrugBank.

Fig. 3. Node degree distribution. Reported is the node degree distribution of approved drugs in the network in Fig. 2

Fig. 4. Exemplary drug hubs. Shown are two exemplary drug hubs. **a** Rimonaabant and **b** Indomethacin. The network cluster of each drug hub is shown. In addition, six representative bioactive analogs forming MMPs with the drug hub are depicted. Structural differences between the drug and its analog are highlighted in red

For 8 of the 79 drug hubs, no alternative/potential targets were suggested on the basis of bioactive analogs. However, for the remaining 71 drug hubs, between 1 and 61 potential targets were found (which we considered a surprisingly large number of drug hubs and potential targets). For these 71 drug hubs, complete profiles of potential targets are provided in Table S1 of the Supporting Information, representing a total

of 887 potential drug-target interactions (again, a surprisingly large number). For 14 of the 71 drug hubs, no high-confidence activity data was available in ChEMBL. Thus, in these cases, no suggested potential target could be confirmed. For 37 of the remaining 57 drug hubs with qualifying activity data, a total of 173 potential drug-target interactions were confirmed in ChEMBL. Table III lists 18 drug hubs with more than two

Table II. Top-Ranked Drug Hubs

DrugID	Name	No. of targets (DrugBank)	No. of MMPs	No. of potential targets
DB00194	Vidarabine	5	110	18
DB00640	Adenosine	6	110	13
DB06155	Rimonabant	2	96	3
DB01258	Aliskiren	2	61	0
DB00259	Sulfanilamide	6	46	17
DB00834	Mifepristone	8	37	0
DB00420	Promazine	18	35	35
DB01246	Trimeprazine	2	35	33
DB01069	Promethazine	14	34	25
DB04896	Milnacipran	2	32	1
DB00328	Indomethacin	27	31	5
DB04794	Bifonazole	6	30	61
DB00392	Ethopropazine	4	29	24
DB00902	Methdilazine	1	29	33
DB01261	Sitagliptin	4	29	2
DB02546	Vorinostat	6	29	4
DB00568	Cinnarizine	15	28	25
DB00780	Phenelzine	14	28	26
DB00985	Dimenhydrinate	1	27	36
DB01238	Aripiprazole	29	26	1

Drug hubs are ranked by the number of MMPs with bioactive analogs. Reported are the top-20 drug hubs forming MMPs with more than 25 bioactive compounds. For each drug hub, its ID, name, and number of protein targets from DrugBank are provided. In addition, the number of potential/alternative targets for drug hubs identified in our analysis is given.

confirmed interactions. In several instances, more than half of the potential targets were confirmed. Table IV reports the complete target profile of promethazine (rank 1 in Table III) that had 14 DrugBank targets and 25 potential targets suggested on the basis of 34 MMPs. In ChEMBL, high-confidence activity data for 22 targets were available for

promethazine including 14 potential targets (highlighted in Table IV). Most of these 14 targets were members of the monoamine GPCR subfamily and hence related to promethazine's DrugBank targets. The remaining suggested targets were unrelated and included enzymes such as acetylcholinesterase or cyclooxygenases.

Table III. Drug Hubs With Confirmed Potential Targets

DrugID	Name	No. of targets (DrugBank)	No. of MMPs	No. of potential targets	No. of confirmed potential targets
DB01069	Promethazine	14	34	25	14
DB00985	Dimenhydrinate	1	27	36	14
DB01176	Cyclizine	3	24	27	13
DB00568	Cinnarizine	15	28	25	12
DB00502	Haloperidol	16	17	16	12
DB00259	Sulfanilamide	6	46	17	11
DB00420	Promazine	18	35	35	11
DB00316	Acetaminophen	12	12	15	11
DB01075	Diphenhydramine	11	11	19	10
DB00245	Benzatropine	5	15	8	6
DB00328	Indomethacin	27	31	5	5
DB00257	Clotrimazole	18	15	13	5
DB04841	Flunarizine	11	14	11	5
DB02546	Vorinostat	6	29	4	4
DB00783	Estradiol	26	13	21	4
DB00972	Azelastine	13	10	4	4
DB01160	Dinoprost tromethamine	2	10	9	4
DB00472	Fluoxetine	11	12	10	3

Ranked are 18 drug hubs for which more than 2 potential targets deduced from bioactive analogs were confirmed by mapping drugs to ChEMBL. For each drug hub, the corresponding ID, name, and the number of protein targets from DrugBank, MMPs, potential targets, and confirmed potential targets are provided.

Table IV. Target Profile of Promethazine

ID	No. of targets (DrugBank)	No. of targets (ChEMBL)	Target list (DrugBank)	Potential/alternative targets
DB01069	14	22	5-hydroxytryptamine 2A receptor Alpha-1a adrenergic receptor Calmodulin Cytochrome P450 2D6 Cytochrome P450 2C9 Cytochrome P450 2B6 Dopamine D2 receptor Histamine H1 receptor Multidrug resistance protein 1 Muscarinic acetylcholine receptor M1 Muscarinic acetylcholine receptor M2 Muscarinic acetylcholine receptor M3 Muscarinic acetylcholine receptor M4 Muscarinic acetylcholine receptor M5	Acetylcholinesterase Adenosine A3 receptor <i>Alpha-1d adrenergic receptor</i> <i>Alpha-2a adrenergic receptor</i> <i>Alpha-2b adrenergic receptor</i> <i>Alpha-2c adrenergic receptor</i> Alpha-synuclein Butyrylcholinesterase Cannabinoid CB1 receptor Cyclooxygenase-1 Cyclooxygenase-2 Cytochrome P450 1A2 <i>Dopamine D1 receptor</i> <i>Dopamine D3 receptor</i> Dopamine transporter <i>Histamine H2 receptor</i> Monoamine oxidase A NADPH oxidase 1 <i>Norepinephrine transporter</i> <i>Serotonin 2b (5-HT2b) receptor</i> <i>Serotonin 2c (5-HT2c) receptor</i> <i>Serotonin 6 (5-HT6) receptor</i> <i>Serotonin transporter</i> <i>Sigma opioid receptor</i> <i>Solute carrier family 22 member 1</i>

For promethazine, the DrugBank ID and the number of designated targets in DrugBank and ChEMBL are reported. DrugBank targets are listed. In addition, alternative/potential targets deduced from promethazine's bioactive analogs are listed. Targets in ChEMBL also reported for promethazine are highlighted in bold italics. Without taking P450 isoforms into consideration, 10 potential targets remain.

Exemplary Drug Hubs With Potential Targets

A total of 714 interactions between drug hubs and alternative/potential targets were identified in our analysis for which no prior evidence was found. In some cases, suggested targets were fairly obvious, in others they were not. Figure 5 shows eight representative drug hubs with exemplary potential targets for which no direct DrugBank or ChEMBL annotations were available.

In Fig. 5a, vidarabine is shown that is an antiviral agent used to treat a variety of viral infections. It acts by terminating viral DNA replication through competitive inhibition of viral DNA polymerase. Vidarabine formed MMPs with 23 and 25 compounds active against adenosine receptor (ADR) A2a and A3, respectively. No reported activity of vidarabine for these receptors was detected in DrugBank, ChEMBL, PubChem, and Open PHACTS (32). The latter two databases were also used to assess selected drug-target interactions. However, adenosine, a stereoisomer of vidarabine, is well-known to be active against these two receptors. Hence, based on this information alone, vidarabine would likely interact with these targets, although there is no guarantee that stereoisomers are active against the same target (especially considering, in this case, that vidarabine is a known antiviral). In Fig. 5b, vidarabine and adenosine are shown. Different from vidarabine, adenosine is a natural metabolite and used as an anti-arrhythmia agent that primarily acts on four subtypes of adenosine receptors. These drugs both formed MMPs with 27 analogs active against

adenosylhomocysteinase (AHCY). Both drugs were reported to be active against this target in individual confirmatory assays in PubChem, with a reported K_i value of 30 μ M for vidarabine (assay ID: 199591) and a K_m value of 0.82 μ M for adenosine (assay ID: 199751). Open PHACTS also reported activity for vidarabine against AHCY. However, given the very low potency reported for vidarabine against AHCY, the target annotation should be considered with caution.

In Fig. 5c, modafinil, a stimulant marketed as a wakefulness promoting agent, is shown. It is used in the treatment of narcolepsy. It has been observed that modafinil was weakly selective for the dopamine transporter and might act by blocking the transporter (33). It has also been observed that modafinil has α 1 adrenergic receptor antagonist properties (34). However, a molecular mechanism-of-action is yet to be confirmed for this drug (35). Modafinil formed 9 MMPs with analogs active against the histamine H1 and serotonin 2a receptor that have thus far not been considered as primary targets. There was no reported activity of modafinil for these receptors in DrugBank, ChEMBL, PubChem, and Open PHACTS. However, the bioactive analogs mostly contained a tertiary amine that is considered a hallmark of classical histamine H1 receptor antagonists (36) but that was not present in modafinil. Hence, in this case, one would predict that modafinil would not be active against the histamine H1 receptor because of an obvious pharmacophore discrepancy compared to its bioactive analogs. However, a recent study has revealed that histamine H1 and H3 receptors were associated with wakefulness (37), i.e., suggesting a potential

Fig. 5. Drug hubs with alternative/potential targets. In **a–c**, exemplary drug hubs with non-confirmed potential targets are shown. Potential targets and the number of bioactive analogs of the drug suggesting these targets are reported. In each case, six representative bioactive analogs are shown. Structural differences between drugs and analogs are *highlighted in red*. Target abbreviations: *ADR*, adenosine receptor; *AHCY*, adenosylhomocysteinase; *HR*, histamine receptor; *AChE*, acetylcholinesterase; *BuChE*, butyrylcholinesterase

link to modafinil-related effects, despite the apparent pharmacophore inconsistency.

In Fig. 5d, four drug hubs, *i.e.*, promazine, methdilazine, promethazine and trimeprazine, sharing the same core

Fig. 5. (continued)

structure are shown, which nonetheless have at least partly different therapeutic indications including allergic reactions or neurologic disorders. Although these drugs were active against different numbers of targets, ranging from 1 to 18, their primary targets included 1 or more class A GPCRs.

They formed MMPs with 6 or 8 analogs active against acetylcholinesterase (AChE) and 22 or 27 analogs active against the related enzyme butyrylcholinesterase (BuChE), for which no evidence existed in the databases we queried. However, searches of the medicinal chemistry literature

Fig. 5. (continued)

revealed reported cholinesterase inactivation (without specifying subtypes) for promazine and trimeprazine with IC_{50} values of 0.27 ± 0.05 and 0.65 ± 0.18 μM , respectively (38), hence supporting these target hypotheses.

In Fig. 5e, cefmenoxime, an antibiotic, is shown, which was annotated with two subtypes of penicillin-binding protein in DrugBank. Cefmenoxime formed MMPs with a total of 11 bioactive analogs, 7 of which were active against members of solute carrier (protein) family that were not related to known primary drug targets. There was no high-confidence activity data available for this drug hub in ChEMBL, Open PHACTS did not contain pharmacology records for cefmenoxime, and no literature evidence was found for potential activity against other than the primary targets. PubChem revealed that cefmenoxime was tested in 28 assays not including solute carrier family members. Hence, in this case, bioactive analogs of cefmenoxime suggest an additional target hypothesis for this drug that is currently not supported by other available information, but can also not be disregarded on the basis of prior knowledge.

CONCLUSIONS

Herein, we have computationally screened chemical and associated target space proximal to approved drugs. To these ends, drug-to-bioactive compound MMPs were systematically generated to identify bioactive analogs of approved drugs. Then, target information was assembled for drugs and their analogs and compared. The analysis was primarily focused on approved drugs for which many structural relationships were detected (designated drug hubs). For reasons detailed in the introductory section, we anticipated that many analogs of drug hubs might be active against subsets of known drug targets. However, the results of our

analysis clearly departed from these expectations. In fact, we found that narrowly confined chemical space around a subset of >600 approved drugs was populated with analogs reported to be active against different targets, for which drug target evidence was lacking. For drug hubs, nearly 900 interactions with alternative/potential targets were detected that were not supported by DrugBank. When approved drugs were mapped to ChEMBL, 173 of these interactions could be confirmed on the basis of high-confidence activity data. These findings demonstrated, first, that close structural analogs are often active against the same target(s), an intuitive and widely accepted premise, and, second, that drug target information was frequently inconsistent in DrugBank and ChEMBL. In light of the increasing complexity and heterogeneity of compound activity data (39) as well as intrinsic difference between database organization and data curation, such inconsistencies might be anticipated. However, ~700 hypothetical target interactions for approved drugs remained for which no supporting database evidence was detected. These were too many interactions to attribute their existence entirely to data inconsistency issues. Bioactive analogs of approved drugs were indeed often annotated with a significant number of targets that differed from (and often outnumbered) known drug targets. This was the case despite the fact that approved drugs are among the most intensely studied and well-characterized small molecules. How might this be rationalized? Clearly, in some cases, additional targets suggested on the basis of close structural relationships might be rather obvious, as illustrated by the vidarabine vs. adenosine example discussed above. From a drug discovery perspective, such relationships might perhaps not be considered worth exploring, which might explain the lack of drug target annotations; yet, it remains surprising that rather obvious target assignments are not confirmed or explicitly

stated for approved drugs. Indeed, it is also possible that obvious structure/analog–target relationships are often not considered –or overlooked– given the strong initial focus of compound development of primary target(s). In other cases, as also discussed above, systematically generated analogs might lack critical pharmacophore requirements for a given target (which can often be rather subtle) and hence structural relationships might not translate into target relationships. However, many suggestions for potential as of yet unconfirmed targets remained for drug hubs on the basis of our analysis that could not be easily discarded on such grounds. The drug examples discussed above also illustrate the complexity of studying target annotations on the basis of available database and literature information. However, it is evident that sophisticated computational approaches were not required to propose novel drug–target interactions for ~70 approved drugs. Rather, examining bioactive analogs of drug hubs, which is a straightforward exercise, yielded many different hypotheses. These findings have implications for drug repositioning efforts. Regardless of data inconsistency issues and complications involved in assessing target annotations, exploring analog space around drugs might have thus far been insufficiently considered, for possible reasons discussed above, or might not have been considered at all (at least in the scientific literature). However, on the basis of our analysis, we would conclude that bioactive analog–derived target hypotheses for drugs should probably be taken into consideration, especially in light of the large number of unique analog–target interactions detected for drug hubs in our study. It should also be emphasized that these hypothetical target interactions were derived from incomplete bioactivity data because structural analogs of drugs have certainly not been systematically profiled against all current targets. Hence, data incompleteness must principally be taken into consideration in the interpretation of our results. Importantly, however, the data incompleteness issue further increases the relevance of the findings reported herein. This is the case because increasing volumes of profiling data for the compounds reported herein would most likely lead to additional target annotations of close analogs of approved drugs, even more than we have already been able to identify.

Taken together, our findings provide a large number of experimentally testable hypotheses in the context of drug repositioning efforts and a sound basis for experimental follow-up. Only experimental assessment will provide ultimate proof (or disproof) of target hypotheses revealed by our chemoinformatic analysis. As a basis for further investigations, all our data and results are made freely available as a part of this study.

ACKNOWLEDGEMENTS

We thank OpenEye Scientific Software, Inc., for the free academic license of the OpenEye Toolkits.

REFERENCES

- Paolini GV, Shapland RHB, van Hoorn WP, Mason JS, Hopkins AL. Global mapping of pharmacological space. *Nat Biotechnol*. 2006;24(7):805–15. doi:10.1038/nbt1228.
- Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein pharmacology by ligand chemistry. *Nat Biotechnol*. 2007;25(2):197–206. doi:10.1038/nbt1284.
- Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol*. 2008;4(11):682–90. doi:10.1038/nchembio.118.
- Boran ADW, Iyengar R. Systems approaches to polypharmacology and drug discovery. *Curr Opin Drug Discov Devel*. 2010;13(3):297–309.
- Knight ZA, Lin H, Shokat KM. Targeting the cancer kinome through polypharmacology. *Nat Rev Cancer*. 2010;10(2):130–7. doi:10.1038/nrc2787.
- Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P. Drug target identification using side-effect similarity. *Science*. 2008;321(5886):263–6. doi:10.1126/science.1158140.
- Lounkine E, Keiser MJ, Whitebread S, Mikhailov D, Hamon J, Jenkins JL, *et al.* Large-scale prediction and testing of drug activity on side-effect targets. *Nature*. 2012;486(7403):361–7. doi:10.1038/nature11159.
- Chong CR, Sullivan DJ. New uses for old drugs. *Nature*. 2006;448(7154):645–6. doi:10.1038/448645a.
- Xie L, Xie L, Kinnings SL, Bourne PE. Novel computational approaches to polypharmacology as a means to define responses to individual drugs. *Annu Rev Pharmacol Toxicol*. 2012;52:361–79. doi:10.1146/annurev-pharmtox-010611-134630.
- Jalencas X, Mestres J. On the origins of drug polypharmacology. *MedChemComm*. 2013;4(1):80–7. doi:10.1039/C2MD20242E.
- Hu Y, Bajorath J. How promiscuous are pharmaceutically relevant compounds? A data-driven assessment. *AAPS J*. 2013;15(1):104–11. doi:10.1208/s12248-012-9421-y.
- Hu Y, Bajorath J. Compound promiscuity: what can we learn from current data? *Drug Discov Today*. 2013;18(13–14):644–50. doi:10.1016/j.drudis.2013.03.002.
- Hu Y, Bajorath J. High-resolution view of compound promiscuity. *F1000Res*. 2013;2:144. doi:10.12688/f1000research.2-144.v2 [v2; ref status: indexed, <http://f1000r.es/lig>].
- Mestres J, Gregori-Puigjané E, Valverde S, Solé RV. Data completeness—the Achilles heel of drug–target networks. *Nat Biotechnol*. 2008;26(9):983–4. doi:10.1038/nbt0908-983.
- Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res*. 2012;40(Database issue):D1100–7. doi:10.1093/nar/gkr777.
- Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Zhou Z, *et al.* PubChem's BioAssay database. *Nucleic Acids Res*. 2012;40(Database issue):D400–12. doi:10.1093/nar/gkr1132.
- Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, *et al.* DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res*. 2011;39(Database issue):D1035–41.
- Hu Y, Bajorath J. Many structurally related drugs bind different targets whereas distinct drugs display significant target overlap. *RSC Adv*. 2012;2(8):3481–9. doi:10.1039/C2RA01345B.
- Leeson PD, Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nature Rev Drug Discovery*. 2007;6(11):881–90. doi:10.1038/nrd2445.
- Wang J, Hou T. Drug and drug candidate building block analysis. *J Chem Inf Model*. 2010;50(1):55–67. doi:10.1021/ci900398f.
- Müller G. Medicinal chemistry of target family-directed masterkeys. *Drug Discov Today*. 2003;8(15):681–91. doi:10.1016/S1359-6446(03)02781-8.
- Sutherland JJ, Higgs RE, Watson I, Vieth M. Chemical fragments as foundations for understanding target space and activity prediction. *J Med Chem*. 2008;51(9):2689–700. doi:10.1021/jm701399f.
- Renner S, van Otterlo WA, Dominguez Seoane M, Möcklinghoff S, Hofmann B, Wetzel S, *et al.* Bioactivity-guided mapping and navigation of chemical space. *Nat Chem Biol*. 2009;5(8):585–92. doi:10.1038/nchembio.188.
- Bemis GW, Murcko MA. The properties of known drugs. 1. Molecular frameworks. *J Med Chem*. 1996;39(15):2887–93.
- Hu Y, Bajorath J. Scaffold distributions in bioactive molecules, clinical trials compounds, and drugs. *ChemMedChem*. 2010;5(2):187–90. doi:10.1002/cmdc.200900419.
- Hu Y, Bajorath J. Many drugs contain unique scaffolds with varying structural relationships to scaffolds of currently available bioactive compounds. *Eur J Med Chem*. 2014;76:427–34. doi:10.1016/j.ejmech.2014.02.040.
- Hu Y, Bajorath J. Polypharmacology directed compound data mining: identification of promiscuous chemotypes with different

- activity profiles and comparison to approved drugs. *J Chem Inf Model.* 2010;50(12):2112–8. doi:[10.1021/ci1003637](https://doi.org/10.1021/ci1003637).
28. Kenny PW, Sadowski J. Structure modification in chemical databases. In *Chemoinformatics in Drug Discovery*; Oprea TI, Ed.; Wiley-VCH: Weinheim, Germany, 2004; pp 271–285.
29. Hussain J, Rea C. Computationally efficient algorithm to identify matched molecular pairs (MMPs) in large data sets. *J Chem Inf Model.* 2010;50(3):339–48. doi:[10.1021/ci900450m](https://doi.org/10.1021/ci900450m).
30. Hu X, Hu Y, Vogt M, Stumpfe D, Bajorath J. MMP-Cliffs: systematic identification of activity cliffs on the basis of matched molecular pairs. *J Chem Inf Model.* 2012;52(5):1138–45. doi:[10.1021/ci3001138](https://doi.org/10.1021/ci3001138).
31. OEChem, version 1.7.7, OpenEye Scientific Software, Inc., Santa Fe, NM, USA. 2012. <http://www.eyesopen.com>.
32. Williams AJ, Harland L, Groth P, Pettifer S, Chichester C, Willighagen EL, *et al.* Open PHACTS: semantic interoperability for drug discovery. *Drug Discov Today.* 2012;17(21–22):1188–98. doi:[10.1016/j.drudis.2012.05.016](https://doi.org/10.1016/j.drudis.2012.05.016).
33. Mignot E, Nishino S, Guilleminault C, Dement WC. Modafinil binds to the dopamine uptake carrier site with low affinity. *Sleep.* 1994;17(5):436–7.
34. Sebban C, Tesolin-Decros B, Millan MJ, Spedding M. Contrasting EEG profiles elicited by antipsychotic agents in the prefrontal cortex of the conscious rat: antagonism of the effects of clozapine by modafinil. *Br J Pharmacol.* 1999;128(5):1055–63. doi:[10.1038/sj.bjp.0702893](https://doi.org/10.1038/sj.bjp.0702893).
35. Gerrard P, Malcolm R. Mechanisms of modafinil: a review of current research. *Neuropsychiatr Dis Treat.* 2007;3(3):349–64. doi:[10.2147/NDT.S](https://doi.org/10.2147/NDT.S).
36. Sharma A, Hamelin BA. Classic histamine H1 receptor antagonists: a critical review of their metabolic and pharmacokinetic fate from a bird's eye view. *Curr Drug Metab.* 2003;4(2):105–29. doi:[10.2174/1389200033489523](https://doi.org/10.2174/1389200033489523).
37. Sundvik M, Kudo H, Toivonen P, Rozov S, Chen YC, Panula P. The histaminergic system regulates wakefulness and orexin/hypocretin neuron development via histamine receptor H1 in zebrafish. *FASEB J.* 2011;25(12):4338–47. doi:[10.1096/fj.11-188268](https://doi.org/10.1096/fj.11-188268).
38. Muraoka S, Miura T. Inactivation of cholinesterase induced by chlorpromazine cation radicals. *Pharmacol Toxicol.* 2003;92(2):100–4. doi:[10.1034/j.1600-0773.2003.920207.x](https://doi.org/10.1034/j.1600-0773.2003.920207.x).
39. Hu Y, Bajorath J. Learning from 'big data': compounds and targets. *Drug Discov Today.* 2014;19(4):357–60. doi:[10.1016/j.drudis.2014.02.004](https://doi.org/10.1016/j.drudis.2014.02.004).