Research Article

Theme: New Paradigms in Pharmaceutical Sciences: In Silico Drug Discovery

Guest Editor: Xiang-Qun Xie

Rationalizing Structure and Target Relationships between Current Drugs

Ye Hu¹ and Jürgen Bajorath^{1,2}

Received 14 March 2012; accepted 10 July 2012; published online 24 July 2012

Abstract. A recent analysis of structure and target relationships between current drugs and bioactive compounds has revealed that only a small fraction of drugs that are active against the same or overlapping targets are involved in substructure relationships and/or share the same topology. By contrast, structurally related drugs displayed a tendency to preferentially act against different targets. For bioactive compounds, opposite trends were observed. These surprising findings arising from the global analysis have now been examined in detail by analyzing structure and target relationships between drugs at the level of individual targets and individual drugs and by comparing the results of local (target- or drug-based) and global relationship analysis. On the basis of target-based analysis, on average, only 14% of drugs active against a given target form well-defined structural relationships. In addition, drug-based analysis revealed that on average 72% of all structurally related drugs have no or at most 20% target overlap. Taken together, the results of our current analysis at the level of single targets and drugs rationalize their unexpected structure and target relationships in a consistent manner. These findings also have implications for ligand binding characteristics of popular drug targets and for frequently observed polypharmacological drug behavior.

KEY WORDS: bioactive compounds; drugs; drug–target interactions; pharmaceutical targets; structural relationships.

INTRODUCTION

The study of molecular features and properties that might be characteristic of drugs is an intensely investigated topic in pharmaceutical research (1). Related to this topic is the study of polypharmacological drug behavior, an emerging theme in drug discovery (2). Different types of indicators or measures of putative drug likeness have been introduced, ranging from statistically derived property rules for oral bioavailability (3) to quantitative numerical estimates (4). Nevertheless, the concept of drug likeness remains controversial (1). For bioactive compounds, relationships between chemical structure and target activity profiles have been studied from different points of view, revealing many aspects of structure–activity relationships, molecular selectivity, or promiscuity (5). Recently, a global analysis has been carried out for currently available drugs, bioactive compounds, and their targets in order to systematically

explore structure and target relationships among them (6). “Global” means that all drugs, bioactive compounds, or targets were analyzed together (forming data pools), without considering individual drugs, compounds, or targets. In this study, rather than attempting to assess drug likeness, substructure and topological relationships among approved and experimental drugs were systematically studied taking target annotations into account and compared to bioactive compounds. The results were surprising, if not provocative: drugs that were topologically equivalent and/or involved in substructure relationships were found to preferentially bind to different targets, whereas structurally dissimilar drugs were often active against the same or overlapping targets (6). For bioactive compounds, essentially opposing trends were observed (6), which were much more in line with intuitive and widely accepted expectations that similar compounds should display similar biological activity (7). Given these unexpected observations made for drugs, we have been interested in further elucidating these findings. Therefore, we have complemented the global analysis of structure and target relationships among drugs and bioactive compounds with a two-component “local” analysis scheme to investigate structure and target relationships at the level of individual targets, drugs, and bioactive compounds. The results of this analysis are reported herein and rationalize observed relationships in detail.

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

²To whom correspondence should be addressed. (e-mail: bajorath@bit.uni-bonn.de)

Fig. 1. Structural hierarchy. The organization scheme underlying the analysis of substructure and topological relationships between drugs or bioactive compounds is illustrated. Following this hierarchy, BM scaffolds (red) abstract from compounds and CSKs further abstract from scaffolds. If one scaffold is contained in another, a substructure relationship exists. If scaffolds yield the same CSK, they are topologically equivalent

MATERIALS AND METHODS

Datasets

From DrugBank (version 3.0) (8), approved (a-) and experimental (e-) small molecule drugs were selected (e-drugs are in late-stage preclinical or in clinical evaluation). Drugs with available structure and target annotations were retained for further analysis and organized into individual target sets. A target set contained all available drugs having a specific target annotation. Thus, drugs with multiple target annotations belonged to different target sets. In addition, bioactive compounds active against human targets with at least 10 μ M potency were assembled from ChEMBL (release 10) (9,10). Only compounds with explicit K_i or IC_{50} measurements and reported direct target interactions (*i.e.*, target relationship type “D”) with highest confidence level

(target confidence score 9) were selected. The target confidence score quantifies the level of confidence that activity against a given target is evaluated in a relevant assay system. It ranges from 0 for noncurated data entries to 9 (*i.e.*, highest level of confidence for a given target). Qualifying compounds were also grouped into individual target sets, in analogy to drugs.

Substructure and Topological Relationships

From all target sets consisting of a-drugs, e-drugs, or bioactive compounds, heteroatom-containing scaffolds were extracted according to Bemis and Murcko (BM scaffolds) by removal of all R groups (11). BM scaffolds were further transformed into cyclic skeletons (CSKs) by converting all heteroatoms to carbon and all bond orders to one (12). The structural hierarchy defined by compounds, BM scaffolds, and CSKs is illustrated in Fig. 1. On the basis of BM scaffolds and CSKs, two types of structural relationships were systematically explored for drugs or bioactive compounds comprising each target set: (1) a BM scaffold is a substructure of another and (2) two different BM scaffolds yield the same CSK. Here, single-ring scaffolds were excluded from the identification of structural relationships, due to their generic character.

Target Relationships

For all a-drugs, e-drugs, and bioactive compounds involved in structural relationships, the average target overlap was determined by calculating an activity overlap factor (AOF; 6) for individual drugs or bioactive compounds, defined here as follows:

$$AOF(i) = \frac{\sum_{j=1}^N \frac{\#Targets(i,j)}{\#Targets(i)}}{N}$$

In this equation, $\#Targets(i)$ is the number of targets that drug (or bioactive compound) i is active against, $\#Targets(i, j)$

Fig. 2. Compound distribution. The distributions of a-drugs (light orange), e-drugs (gray), and bioactive compounds (light blue) over target sets are shown in a histogram format

Table I. Structural Relationships in Target Sets

Type	Compound ratio		# Target sets (ratio) without structural relationships
	Global (%)	Target sets (avg; %)	
Approved drugs	71	16	660 (76%)
Experimental drugs	63	12	2397 (83%)
Bioactive compounds	97	71	146 (17%)

The ratio of all drugs/compounds involved in structural relationships and the average ratio of such drugs/compounds active against a given target are reported. In addition, the number (ratio) of target sets with no structural relationships is given

is the number of targets that drug (compound) i and drug (compound) j share, and N the number of drugs (compounds) involved in structural relationships with i . AOF values range from 0 to 1; 0 indicates distinct target activity (no target overlap) for i and all drugs (compounds) with structural relationships and 1 identical activity (complete target overlap).

RESULTS

Drug Hubs versus Target Sets

For the initial global analysis of drug structure and target relationships, the concept of “drug hubs” was introduced (6).

Fig. 3. Structural relationships. For each target set, the ratio of compounds involved in substructure/topological relationships was determined. The histograms in **a**, **c**, and **e** summarize the results for a-drugs, e-drugs, and bioactive compounds, respectively. In addition, the correlation between the number of drugs/compounds per target set and the percentage of compounds involved in structural relationships is reported in scatter plots for a-drugs **b**, e-drugs **d**, and bioactive compounds **f**. In these plots, each data point represents a different target set

Fig. 4. Target overlap. Reported is the target overlap for a-drug **a**, e-drugs **b**, and bioactive compounds **c** involved in structural relationships

From drug–target interactions, drug–drug relationships were deduced by assembling pairs of drugs that had one or more targets in common. Accordingly, “activity-based drug hubs” were then defined as drugs that shared target activities with 50 or more other drugs. In addition, other drug–drug relationships were also accounted for by collecting drug pairs that were involved in one or two types of structural relationships. Therefore, so-called “structural relationship-based drug hubs” were defined as drugs that were involved in substructure and/or topological relationships with 50 or more other drugs. The analysis of structure and target relationships was then carried out on these drug hubs (6), hence providing a global view of these relationships. By contrast, our current

analysis focuses on all individual targets and drugs/compounds, thereby transforming the global into a two-component local analysis scheme of higher resolution; a key aspect of our investigation, as detailed in the following.

Local Relationships

Structural Relationships

Following the structural hierarchy displayed in Fig. 1, two types of structural relationships were systematically determined on a per-target set basis: (1) if one BM scaffold is completely contained in another, a substructure relationship exists between these scaffolds and all drugs or bioactive compounds represented by them and (2) different scaffolds yielding the same CSK and all compounds represented by these scaffolds are topologically equivalent.

A total of 1,216 a-drugs active against 873 targets (*i.e.*, 873 target sets), 4,525 e-drugs active against 2,885 targets, and 71,678 bioactive compounds associated with 864 targets were selected. Figure 2 shows the distribution of a-drugs, e-drugs, and bioactive compounds over all target sets. For a- and e-drugs, the majority of the target sets consisted of only one to five drugs. However, four target sets were found to contain more than 60 approved drugs, *i.e.*, muscarinic acetylcholine receptor M1, histamine H1 receptor, alpha-1A adrenergic receptor, and DNA with 70, 73, 66, and 73 assigned a-drugs, respectively. For bioactive compounds, the distribution significantly differed from drugs, as expected. In this case, many target sets were found to comprise of more than 100 compounds.

Substructure and/or topological relationships were detected for 213, 488, and 718 target sets of a-drugs, e-drugs and bioactive compounds, respectively. On average, for each target set, ~16% of all a-drugs, ~12% of e-drugs, and ~71% of bioactive compounds were found to be involved in one or both types of structural relationships (Table I). For each target set, the ratio of compounds involved in structural relationships was determined and reported in Fig. 3a, c, and e. For a- and e-drugs, the majority of target sets contained only very few or no structural relationships. By contrast, target sets of bioactive compounds often contained many structural relationships. In addition, scatter plots reflecting the correlation between the number of compounds per target set and the percentage of the compounds involved in structural relationships are shown in Fig. 3b, d, and f. Essentially, no correlation was observed for a- and e-drugs. However, the more bioactive compounds a target set contained, the more structural relationships tended to be formed, hence reflecting

Table II. Target Overlap in the Presence of Structural Relationships

Type	Number of compounds (ratio)			
	In structural relationships	TO=0%	TO≤20%	TO=100%
Approved drugs	865	272 (31%)	528 (67%)	45 (5%)
Experimental drugs	2,839	1,183 (42%)	2,171 (76%)	302 (11%)
Bioactive compounds	69,604	3958 (6%)	22,965 (33%)	17,780 (26%)

The number of individual drugs/compounds involved in structural relationships and their average target overlap (TO) are reported

Fig. 5. Target set without structural relationships. Shown are 13 and 9 a-drugs for cholinesterase **a** and neuronal acetylcholine receptor subunit alpha-2 **b**, respectively, which are not involved in structural relationships. These examples are representative for many different target sets

another systematic difference between drugs and bioactive compounds.

Table I also reveals a striking difference between globally and locally assessed structural relationships. The present analysis focused on each target set independently. Globally, 71% and 63% of a-drugs and e-drugs, respectively, were involved in substructure relationships or were topologically equivalent. By contrast, for a given target set, on average only 16% and 12% of a- and e-drugs, respectively, were involved in structural relationships. Moreover, for 76% and 83% of the target sets of a- and e-drugs, respectively, no structural relationships were detectable. Thus, drugs in these target sets were structurally diverse, although they shared the same target activity. For bioactive compounds, different

results were obtained. About 97% of all bioactive compounds were globally involved in structural relationships. At the target set level, structural relationships were still formed by on average 71% of all bioactive compounds. Only 17% of all target sets contained no structural relationships. Thus, there was a much higher degree of structural diversity among drugs active against a given target than among bioactive compounds.

Target Relationships

In the second step of our local analysis, we quantified the target overlap for individual drugs involved in structural relationships on the basis of AOF calculations. Figure 4a–c

Fig. 6. Structurally related drugs with distinct target activities. Shown is the a-drug sumatriptan (*top*) that is involved in structural relationships with other 78 a-drugs. Eight of these drugs are shown that form substructure relationships (*left*) or share the same topology (*right*). For each drug, the number of its currently known targets is reported. Sumatriptan is active against four targets, none of which is shared with any of the other structurally related drugs

report the target overlap for structurally related a-drugs, e-drugs, and bioactive compounds, respectively. The distributions are significantly different for drugs and bioactive compounds. The majority of structurally related a- and e-drugs had no or only very limited target overlap. By contrast, the majority of bioactive compounds displayed extensive target overlap. Table II further quantifies these findings. Approximately 31% (*i.e.*, 272) and 42% (*i.e.*, 1,183) of structurally related a- and e-drugs, respectively, but only 6% of structurally related bioactive compounds had no targets in common. In addition, ~67% and ~76% of a- and e-drugs, respectively, displayed at most 20% target overlap. However, only ~33% of structurally related bioactive compounds fell into the same category. Hence, there was significantly higher degree of activity diversity among drugs involved in structural relationships with other drugs than among bioactive compounds.

Relationship Patterns

Among a- and e-drugs, two dominant types of structure and target relationships emerged, consistent with the findings reported in Tables I and II. First, drugs active against the same target were structurally diverse, *i.e.*, only few structural

relationships were detectable among drugs having the same target activity. Second, structurally related drugs mostly had distinct target annotations. In the following, examples are provided.

Structurally Distinct Drugs Active against a Given Target

Cholinesterase is the target of 13 approved drugs (that form its target set), as shown in Fig. 5a. These drugs are characterized by a high degree of structural diversity and the absence of any substructure relationships or topological equivalence, which was also observed for many other target sets. In such cases, drug targets might be highly small molecule-permissive (*i.e.*, tolerant to different scaffolds and/or mechanisms of action) or characteristic functional groups might be activity determinants. In Fig. 5b, nine approved drugs for neuronal acetylcholine receptor subunit alpha-2 are shown, providing a similar picture.

Structurally Related Drugs with Distinct Target Activities

Figure 6 illustrates the other highly recurrent case of structurally related drugs having distinct target annotations. In this example, sumatriptan is used as a reference drug.

Sumatriptan is involved in structural relationships with a total of 78 other a-drugs. On the left in Fig. 6, four drugs are shown that form substructure relationships with sumatriptan and on the right, four other drugs that are topologically equivalent. Sumatriptan is active against four different targets, but all structurally related drugs have distinct activities (*i.e.*, there is no target overlap).

DISCUSSION

Previous findings have indicated that structure and target relationships between drugs depart from the intuitive paradigm that similar compounds should have similar (target) activity (7), which is widely accepted in the study of bioactive compounds, despite its generality and notable exceptions (13). Global analysis of structure and target relationships indicated that structurally diverse drugs were frequently active against the same or overlapping targets, whereas structurally related drugs preferentially acted on different targets. Because these findings were rather unexpected, we have explored different ways to further detail them and elucidate their origin. We reasoned that one should best transform the previously applied global analysis scheme into a local one, by directly focusing on individual targets and drugs. The results of our systematic target set- and drug-based analyses are reported herein. They clearly rationalize surprising structure and target relationships among drugs and provide further details. The results were very similar for a- and e-drugs, but significantly differed from bioactive compounds. One of two key findings has been that substructure relationships and/or topological equivalence between drugs acting against the same target are extremely rare. On average, only ~14% of drugs within a target set were involved in any structural relationships. The examples in Fig. 5 illustrate that in many instances no structural relationships were observed in target sets. The second key finding has been that the majority of structurally related drugs indeed act against different targets, as illustrated by the sumatriptan example in Fig. 6. We have been able to quantify the target overlap for each drug involved in structural relationships. On average, ~72% of all of these drugs displayed no or very little target overlap of maximally 20%, which again contrasts the observations made for bioactive compounds. Taken together, these findings establish the presence of characteristic structure and target relationships among drugs that set them apart from bioactive compounds (although drugs originate from a pool of bioactive molecules). However, our findings can at least in part be rationalized from a chemical or pharmacological perspective. In the pharmaceutical industry, there is pressure to establish intellectual property positions and structural novelty as a prerequisite for the approval of new drugs against existing targets. This might at least in part explain the observation that many structurally unrelated compounds were active against a given target. Furthermore, the finding that structurally related drugs act against many different targets is consistent with the increasing notion of drug polypharmacology (14). This observation also suggests that chemotypes are frequently utilized in medicinal chemistry having a high probability of activity against different targets. Preferred chemotypes have often been considered as target class- or family-directed (privileged) structural motifs

(15,16). However, scaffolds have also been identified that display a strong tendency to be active across different families (17). Going beyond chemical considerations, the characteristic drug structure and target relationships also have interesting implications concerning current drug targets. Importantly, the observation that many structurally distinct drugs act against individual targets directly implies that many current drug targets are particularly permissive small molecule targets. In addition, the finding that structurally related drugs act on a variety of different targets also points at a high degree of permissiveness. The presence of many permissive small molecule drug targets is also consistent with increasingly observed polypharmacological drug behavior.

CONCLUSIONS

Structure and target relationships between current approved and experimental drugs have been explored in detail at the level of individual targets and drugs. The results provide further evidence for previous observations made on the basis of a global statistical analysis of drug structure and target annotations. Herein, structure and target relationships have been quantified for individual drugs and targets. On average, only ~14% of drugs acting against a specific target are structurally related, and ~72% of all structurally related drugs have no or at most 20% target overlap. In summary, these results indicate that there should still be ample opportunities to consider new chemical entities for established drug targets and, in addition, that polypharmacological effects among structurally related drugs might even be of larger magnitude than currently assumed.

REFERENCES

1. Ursu O, Rayan A, Goldblum A, Oprea TI. Understanding drug-likeness. *WIREs Comp Mol Sci*. 2011;1(5):760–81. doi:10.1002/wcms.52.
2. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol*. 2008;4(11):682–90. doi:10.1038/nchembio.118.
3. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 1997;23(1–3):3–25. doi:10.1016/S0169-409X(96)00423-1.
4. Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL. Quantifying the chemical beauty of drugs. *Nat Chem*. 2012;4(2):90–8. doi:10.1038/nchem.1243.
5. Hu Y, Stumpfe D, Bajorath J. Lessons learned from molecular scaffold analysis. *J Chem Inf Model*. 2011;51(8):1742–53. doi:10.1021/ci200179y.
6. 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.
7. Johnson MA, Maggiora GM, editors. *Concepts and applications of molecular similarity*. Wiley: New York; 1990.
8. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, Djoumbou Y, Eisner R, Guo AC, Wishart DS. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res*. 2011;39(Database issue):D1035–41. doi:10.1093/nar/gkq1126.
9. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res*. 2012;40(Database issue):D1100–7. doi:10.1093/nar/gkr777.

10. ChEMBL; European Bioinformatics Institute (EBI): Cambridge, 2011. <http://www.ebi.ac.uk/chembl/> (accessed September 1, 2011).
11. Bemis GW, Murcko MA. The properties of known drugs. 1. Molecular frameworks. *J Med Chem.* 1996;39(15):2887–93. doi:10.1021/jm9602928.
12. Xu YJ, Johnson M. Using molecular equivalence numbers to visually explore structure features that distinguish chemical libraries. *J Chem Inf Comput Sci.* 2002;42(4):912–26. doi:10.1021/ci025535l.
13. Maggiora GM. On outliers and activity cliffs—why QSAR often disappoints. *J Chem Inf Model.* 2006;46(4):1535. doi:10.1021/ci060117s.
14. Keiser MJ, Irwin JJ, Shoichet BK. The chemical basis of pharmacology. *Biochemistry.* 2010;49(48):10267–76. doi:10.1021/bi101540g.
15. 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.
16. Hu Y, Wassermann AM, Lounkine E, Bajorath J. Systematic analysis of public domain compound potency data identifies selective molecular scaffolds across druggable target families. *J Med Chem.* 2010;53(2):752–8. doi:10.1021/jm9014229.
17. 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.