



## COMMENTARY

# High-resolution view of compound promiscuity [v2; ref status: indexed, <http://f1000r.es/1ig>]

Ye Hu, Jürgen Bajorath

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

**v2** **First Published:** 27 Jun 2013, 2:144 (doi: 10.12688/f1000research.2-144.v1)  
**Latest Published:** 26 Jul 2013, 2:144 (doi: 10.12688/f1000research.2-144.v2)

## Abstract

Compound promiscuity is defined as the ability of a small molecule to specifically interact with multiple biological targets. So-defined promiscuity is relevant for drug discovery because it provides the molecular basis of polypharmacology, which is increasingly implicated in the therapeutic efficacy of drugs. Recent studies have analyzed different aspects of compound promiscuity on the basis of currently available activity data. In this commentary, we present take-home messages from these studies augmented with new results to generate a detailed picture of compound promiscuity that might serve as a reference for further discussions and research activities.

## Article Status Summary

### Referee Responses

Referees	1	2	3
<b>v1</b> published 27 Jun 2013	 report	 report	 report
	↓		↓
<b>v2</b> published 26 Jul 2013	 report		 report

**1 Stefan Laufer**, University of Tübingen  
Germany

**2 Hans Matter**, Sanofi-Aventis France

**3 Jeremy Jenkins**, Novartis Institutes for  
BioMedical Research USA

### Latest Comments

**Jürgen Bajorath**, Department of Life Science Informatics, B-IT and LIMES Institutes, Rheinische Friedrich-Wilhelms-Universität Bonn, Germany  
08 Jul 2013 (V1)

**Corresponding author:** Jürgen Bajorath ([bajorath@bit.uni-bonn.de](mailto:bajorath@bit.uni-bonn.de))

**How to cite this article:** Hu Y, Bajorath J (2013) High-resolution view of compound promiscuity [v2; ref status: indexed, <http://f1000r.es/1ig>] *F1000Research* 2013, 2:144 (doi: 10.12688/f1000research.2-144.v2)

**Copyright:** © 2013 Hu Y et al. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Grant information:** The author(s) declared that no grants were involved in supporting this work.

**Competing Interests:** No competing interests were disclosed.

**First Published:** 27 Jun 2013, 2:144 (doi: 10.12688/f1000research.2-144.v1)

**First Indexed:** 09 Jul 2013, 2:144 (doi: 10.12688/f1000research.2-144.v1)

**UPDATED** Changes from Version 1

In version 2, three references (9, 15, and 16) have been updated. In response to reviewer comments of Dr. Hans Matter, we now also report the results of compound promiscuity analysis for five well-known target families including G protein-coupled receptor (GPCR) class A, protein kinases, ion channels, proteases, and nuclear hormone receptors. In addition, we have determined promiscuity levels for compounds in different molecular weight ranges, as also suggested by Dr. Matter. Four new tables (3–6) have been added. Furthermore, in response to reviewer comments of Dr. Jeremy Jenkins, we report median promiscuity rates compared to average rates and briefly discuss a potential relationship between privileged structures and compounds displaying intra-family promiscuity.

See referee reports

## Introduction

Polypharmacology is an emerging theme in drug discovery<sup>1,2</sup>. It is generally accepted that drugs often elicit their therapeutic effects through interactions with different targets and the ensuing modulation of multiple signaling pathways. In some therapeutic areas such as oncology, polypharmacology is heavily exploited, for example, through the use of promiscuous ATP site-directed protein kinase inhibitors<sup>3</sup>. In other areas, such as the treatment of infectious or chronic inflammatory diseases, achieving a high degree of target selectivity of drug candidates plays a major role.

The study of drug polypharmacology has become an important topic in pharmaceutical research<sup>4,5</sup>, especially focusing on combined computational and experimental analysis<sup>5</sup>. On the basis of drug-target networks, it was estimated early on that a drug interacts on average with approximately two targets<sup>4</sup>. More recent estimates from computational data analysis suggest that drugs might bind on average to two to seven targets, depending on the primary target families, and that more than 50% of current drugs might interact with more than five targets<sup>6</sup>.

Compound promiscuity as defined herein is the origin of polypharmacology. Promiscuity analysis can be extended from drugs to bioactive compounds through computational mining of currently available activity data. The results of activity data analysis are generally affected by data incompleteness<sup>7</sup>. This potential influence can only be eliminated by reaching the ultimate (and probably elusive) goal of chemogenomics<sup>8</sup>, i.e., testing all compounds against all targets. In the presence of data incompleteness, compound promiscuity rates are likely underestimated. However, it is not certain that further increasing amounts of assay data will indeed significantly alter the currently emerging view of compound promiscuity (*vide infra*).

Recent studies have generated a differentiated picture of compound promiscuity. The interested reader is also referred to comprehensive reviews of compound promiscuity analysis<sup>9</sup> and polypharmacology<sup>6</sup>. In this commentary, we summarize key messages from recent promiscuity analysis in a compact format. It is hoped that this summary might be helpful as a reference for further studies.

## Key results of compound promiscuity analysis

Public data sources for compound promiscuity analysis discussed herein have been ChEMBL<sup>10</sup>, the major repository of compound activity data from medicinal chemistry (currently in May 2013 containing 1,295,510 compounds with a total of 11,420,351 activity annotations), the PubChem BioAssay database<sup>11</sup>, the major repository of screening data (with more than 3300 confirmatory assays), and DrugBank<sup>12</sup>, which currently contains 1518 approved and 5080 experimental drugs.

It is important to note that collecting all activity annotations for a compound reported in the literature including, for example, reporter gene or other cell-based assays is at best providing a measure of assay promiscuity, but not of specific interactions with different targets<sup>9</sup>. Therefore, it is generally required to apply data confidence criteria such as the presence of well-defined activity measurements or evidence for direct ligand-target interactions<sup>9</sup> (as provided in ChEMBL as activity data filters).

## Activity measurement dependence

When monitoring the growth of compound activity data in ChEMBL over a period of more than two years from its original release (January 2010) to release 13 (May 2012), a significant increase in the number of promiscuous compounds was detected<sup>13</sup>. However, by quantifying compound-based target relationships, it was determined that the increase in compounds with activity against targets from different families was largely due to (assay-dependent)  $IC_{50}$  measurements, rather than (assay-independent) equilibrium constants ( $K_i$  values)<sup>13</sup>.  $IC_{50}$  values are easier to determine than  $K_i$  values and provide the read-out of most primary biochemical assays (except single-point screening assays), which might at least in part rationalize greater target coverage and the  $IC_{50}$ -dependent increase in compound promiscuity across different families. However, it can also not be excluded that apparent promiscuity in different assays is higher on the basis of  $IC_{50}$  measurements, given their assay dependence (and often limited accuracy). Regardless, the type of activity measurements that are taken into account influences the outcome of promiscuity analysis. Thus, clear specification of activity measurements and data selection criteria are required.

The subset of compounds with available  $K_i$  measurements from ChEMBL release 13 was further investigated. *On the basis of  $K_i$  measurements, approximately 62% of all compounds were only annotated with a single target, ~36% with two or more targets from the same family, and only ~2% of all active compounds with multiple targets from different families<sup>14</sup>. A promiscuous bioactive compound was found to interact on average with two to three targets.*

Accordingly, compounds that display intra-family promiscuity might also be considered as candidates for privileged structures/compounds that are preferentially active against targets from a particular family. Therefore, these compounds can be distinguished from those that are promiscuous across different target families.

## Activity data from different sources

One might anticipate that the degree of compound promiscuity would be particularly high in screening assays (even if frequent hitters and other non-specific compounds are excluded). Therefore,

1085 confirmatory bioassays from PubChem were systematically analyzed. It was found that ~77% of all confirmed active compounds were tested in more than 50 different assays<sup>15</sup>. Thus, these active PubChem compounds provided a sound basis for promiscuity assessment. These results were in part surprising. *An active PubChem compound displayed a ~50% probability to interact with two or more targets. The probability to interact with more than five targets was only ~8%. On average, a PubChem screening hit was active against 2.5 targets.* For comparison, *compounds from the IC<sub>50</sub>- and K<sub>i</sub>-based subsets of ChEMBL release 14 (August 2012) interacted on average with 1.4 and 1.7 targets, respectively<sup>15</sup>.* The comparably low ratios observed for both compound subsets indicated that IC<sub>50</sub> measurements did not systematically increase promiscuity rates (*vide supra*). The analysis of active compounds from PubChem confirmatory assays provided an upper level estimate of promiscuity, which was not significantly higher than that for ChEMBL compounds.

### Prevalent promiscuity profile

Detailed analysis of compound activity data from ChEMBL release 14 (August 2012) has made it possible to derive a *promiscuity profile* that is most characteristic of bioactive compounds from medicinal chemistry sources. *The majority of currently available promiscuous compounds is active in the sub- $\mu$ M range against two to five targets from the same family and displays potency differences against these targets within one or two orders of magnitude<sup>16</sup>.* An important aspect of this representative profile is that promiscuity does not imply low potency. Furthermore, compounds that are highly potent against a (primary) target and weakly potent against others are not frequently found<sup>16</sup>.

### Up-to-date promiscuity rates

In [Table 1](#), current average promiscuity rates are summarized for compounds from ChEMBL, PubChem, and DrugBank. For promiscuity assessment of drugs, all targets reported in DrugBank were considered.

If all compounds with single or multiple target annotations are analyzed, ChEMBL compounds interact on average with one to two targets and PubChem compounds with two to three. However, approved drugs have on average close to six targets. In contrast, the degree of promiscuity of experimental drugs is considerably lower, with less than two targets per drug candidate. If only promiscuous compounds or drugs are taken into account (i.e., if compounds with single target annotations are excluded), promiscuity rates only slightly increase by about one target per compound, the exception being experimental drugs whose average number of targets increases from 1.8 to 4.7. Furthermore, median promiscuity rates were also calculated for promiscuous compounds from different sources, i.e., ChEMBL compounds with activity against at least two targets (K<sub>i</sub> and IC<sub>50</sub>), approved and experimental drugs annotated with more than four or at least two targets, respectively, and PubChem compounds active against at least three targets. Compared to the average promiscuity rates reported in [Table 1](#), the median rates were consistently lower. However, the differences between the average and median rates were small, i.e., less than one for ChEMBL and PubChem compounds. By contrast, differences were larger than one for approved and experimental drugs, i.e., on the basis of median rates, drug target numbers were reduced by 1.9 and 2.7, respectively. Hence, average promiscuity rates for drugs were likely biased by highly promiscuous drugs.

In [Table 2](#), the probability of promiscuity is reported for compounds from different sources (calculated from target distributions of compounds). For a ChEMBL compound with available IC<sub>50</sub> and K<sub>i</sub> measurements, the current probability of activity against two or more targets is ~25% and ~38%, respectively (if both IC<sub>50</sub> and K<sub>i</sub> measurements were available for a compound, they were separately considered). However, for activity against more than five targets, the probabilities are reduced to only ~1%. Similar observations are made for confirmed PubChem screening hits (providing an upper-limit promiscuity assessment for bioactive compounds, *vide supra*). In this case, the probability of activity against two or more, or against

**Table 1. Average promiscuity of different compound categories.**

Compound categories		Avg. # targets/compound
ChEMBL 14/all bioactive compounds	K <sub>i</sub>	1.7
	IC <sub>50</sub>	1.4
DrugBank/drugs	Approved	5.9
	Experimental	1.8
PubChem/active compounds		2.5
ChEMBL 14/promiscuous compounds	K <sub>i</sub>	2.9
	IC <sub>50</sub>	2.7
DrugBank/promiscuous drugs	Approved	6.9
	Experimental	4.7
PubChem/promiscuous active compounds		3.7

The average number of targets is reported for compounds from ChEMBL release 14 (divided into K<sub>i</sub> and IC<sub>50</sub> value-based subsets), approved or experimental drugs from DrugBank 3.0, and active compounds from PubChem confirmatory bioassays. Corresponding statistics are provided in italics for promiscuous compounds (having two or more target annotations). For compounds from ChEMBL, only high-confidence activity annotations were taken into account (i.e., explicit activity measurements with the highest confidence level of direct ligand-target interactions). For calculations on drugs, all DrugBank target categories were taken into account.

**Table 2.** Probability of promiscuity.

Compound categories		# Targets	Probability (%)
ChEMBL 14/all bioactive compounds	$K_i$	$\geq 2$	37.9
		$> 5$	1.2
	$IC_{50}$	$\geq 2$	24.7
		$> 5$	0.8
DrugBank/drugs	Approved	$\geq 2$	84.1
		$> 5$	37.4
	Experimental	$\geq 2$	23.6
		$> 5$	3.4
PubChem/active compounds	$\geq 2$	50.9	
	$> 5$	7.6	

For different compound categories and activity measurements, the probability of a compound to be active against two or more targets or more than five targets is reported.

**Table 3.** Average promiscuity of compounds active against prominent target families.

Family	# Targets	# Compounds	$K_i$	
			Avg. # targets/compound	
			All	<i>Promiscuous</i>
GPCR class A	121	21,754	1.7	<i>2.8</i>
Kinases	74	1151	1.4	<i>2.4</i>
Ion channels	22	1086	1.2	<i>3.9</i>
Proteases	90	4488	1.5	<i>2.8</i>
Nuclear receptors	13	901	1.4	<i>2.6</i>
Family	# Targets	# Compounds	$IC_{50}$	
			Avg. # targets/compound	
			All	<i>Promiscuous</i>
GPCR class A	135	16,968	1.3	<i>3.0</i>
Kinases	192	13,316	1.3	<i>2.7</i>
Ion channels	52	4150	1.1	<i>2.2</i>
Proteases	108	11,833	1.6	<i>3.0</i>
Nuclear receptors	26	3782	1.4	<i>2.1</i>

From ChEMBL release 14,  $K_i$  and  $IC_{50}$  value-based compound subsets active against targets belonging to five prominent families were collected. The average number of targets is reported for compounds from individual target families. In addition, corresponding statistics are provided (in italics) for promiscuous compounds only (i.e., compounds having two or more target annotations within the family).

more than five targets is ~51% and ~8%, respectively. Furthermore, the probability of promiscuity of approved drugs from DrugBank is ~84% and the probability to interact with more than five targets still ~37%. For experimental drugs, the corresponding probabilities are much lower, with only ~24% and ~3%, respectively.

### Compound promiscuity for different target families

All available compounds active against targets belonging to the five target families, including G protein-coupled receptor (GPCR) class A, protein kinases, ion channels, proteases, and nuclear hormone receptors, were assembled from ChEMBL release 14 and separated into  $K_i$  and  $IC_{50}$  value-based subsets, as described above. Average promiscuity rates were calculated for all compounds active

against a given family as well as compounds active against multiple targets within the family, as reported in Table 3. With the exception of the  $K_i$  subset of the ion channel family, promiscuity degrees for compounds active against these target families were similar to those reported in Table 1. In Table 4, the probability of promiscuity (i.e., activity against at least two or more than five targets) is reported for compounds active against these families (according to Table 2). Similar observations were made. A significant relative increase (~10%) in probability of promiscuity was only observed for compounds active against two or more targets from the nuclear receptor family on the basis of the  $IC_{50}$  subset. Thus, for prominent target families, no above-average compound promiscuity rates were detected.

**Table 4. Probability of promiscuity for compounds active against prominent target families.**

Family	Probability (%)			
	$K_i$		$IC_{50}$	
	$\geq 2$ Targets	$> 5$ Targets	$\geq 2$ Targets	$> 5$ Targets
<b>GPCR class A</b>	39.1	0.8	16.0	0.8
<b>Kinases</b>	26.3	0.4	20.1	1.0
<b>Ion channels</b>	8.4	0.1	5.3	0.02
<b>Proteases</b>	27.5	0.5	27.2	1.0
<b>Nuclear receptors</b>	24.9	1.2	34.8	0.03

For compounds active against different target families, the probability of promiscuity (activity against two or more targets and activity against more than five targets) is reported.

**Table 5. Average promiscuity of compounds with different molecular weight.**

MW range	#Compounds	$K_i$	
		Avg. # targets/compound	
		All	Promiscuous
$\leq 200$	786	2.7	4.1
<b>(200, 300)</b>	3949	2.0	3.2
<b>(300, 400)</b>	10,913	1.8	2.9
<b>(400, 500)</b>	11,501	1.6	2.7
<b>(500, 600)</b>	6015	1.6	2.8
<b>(600, 700)</b>	1878	1.6	3.0
<b>&gt; 700</b>	1497	1.7	2.8

  

MW range	#Compounds	$IC_{50}$	
		Avg. # targets/compound	
		All	Promiscuous
$\leq 200$	1022	1.6	2.8
<b>(200, 300)</b>	9627	1.6	2.8
<b>(300, 400)</b>	25,190	1.4	2.7
<b>(400, 500)</b>	26,358	1.4	2.6
<b>(500, 600)</b>	12,534	1.4	2.6
<b>(600, 700)</b>	3247	1.3	2.8
<b>&gt; 700</b>	2495	1.4	2.9

From ChEMBL release 14, compounds were selected and divided into seven subsets with increasing MW. The average number of targets is reported for compounds in all MW ranges. In addition, corresponding statistics are provided (in italics) for promiscuous compounds only (i.e., compounds having two or more target annotations).

### Promiscuity of compounds with increasing molecular weight

The degree of promiscuity was also determined for compounds with different sizes, i.e., molecular weight (MW). Seven subsets of compounds with increasing MW were collected from ChEMBL release 14 and organized into  $K_i$  and  $IC_{50}$  value-based subsets, as reported in Table 5. Average promiscuity rates of compounds with increasing

**Table 6. Probability of promiscuity of compounds with different molecular weight.**

MW range	Probability (%)			
	$K_i$		$IC_{50}$	
	$\geq 2$ Targets	$> 5$ Targets	$\geq 2$ Targets	$> 5$ Targets
$\leq 200$	53.8	9.9	36.0	2.4
<b>(200, 300)</b>	44.9	2.4	32.7	0.9
<b>(300, 400)</b>	39.4	1.2	26.0	0.9
<b>(400, 500)</b>	37.3	0.7	22.8	0.7
<b>(500, 600)</b>	31.5	0.6	22.1	0.7
<b>(600, 700)</b>	30.7	0.9	17.6	0.7
<b>&gt; 700</b>	38.3	1.2	18.9	0.3

For compounds with different molecular weight, the probability of promiscuity (activity against two or more targets and activity against more than five targets) is reported.

MW were found to be comparable to the global rates. However, a significant relative increase in promiscuity was observed for the smallest compounds with MW  $\leq 200$  in  $K_i$  subset. Furthermore, the probability of activity against two or more targets also increased by more than 10% for the smallest compounds in both subsets, as reported in Table 6. For larger compounds across all MW ranges, no significant increases in promiscuity were observed compared to the global degree and probability of compound promiscuity reported in Table 1 and Table 2, respectively.

### Conclusions

Herein, we have provided a detailed and up-to-date view of compound promiscuity, the molecular basis of polypharmacology. For active compounds from medicinal chemistry and biological screening sources, the degree of promiscuity is lower than for drugs. There is a notable increase in promiscuity from bioactive compounds over drug candidates to approved drugs. The exploration of possible reasons for this apparent “promiscuity enrichment” along the drug discovery pathway should provide interesting opportunities for future research. On the basis of currently available high-confidence activity data, promiscuity of bioactive compounds is limited (and very low across different target families). However, if compounds are promiscuous, they typically bind to their targets with relatively high potency. Given



the overall low degree of promiscuity of bioactive compounds including screening hits in the presence of nearly exponential data growth in recent years, it remains an open question if future chemogenomics efforts might substantially change the current picture of compound promiscuity (*vide supra*). The majority of available bioactive compounds have single target annotations and we believe it is unlikely that most of them will display a high degree of currently undiscovered promiscuity. Hence, we would also conclude that the target specificity paradigm that has long dominated small molecule discovery efforts should continue to play a major role, despite emerging “anti-reductionism” and the increasing focus on phenotypic readouts.

### Author contributions

JB conceived the study, YH collected and organized the data and information, YH and JB wrote the manuscript.

### Competing interests

No competing interests were disclosed.

### Grant information

The author(s) declared that no grants were involved in supporting this work.

## References

- Paolini GV, Shapland RH, van Hoon WP, *et al.*: **Global mapping of pharmacological space.** *Nat Biotechnol.* 2006; **24**(7): 805–815.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Boran AD, Iyengar R: **Systems approaches to polypharmacology and drug discovery.** *Curr Opin Drug Discov Devel.* 2010; **13**(3): 297–309.  
[PubMed Abstract](#) | [Free Full Text](#)
- Knight ZA, Lin H, Shokat KM: **Targeting the cancer kinome through polypharmacology.** *Nat Rev Cancer.* 2010; **10**(2): 130–137.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Yildirim MA, Goh KI, Cusick ME, *et al.*: **Drug-target network.** *Nat Biotechnol.* 2007; **25**(10): 1119–1126.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lounkine E, Keiser MJ, Whitebread S, *et al.*: **Large-scale prediction and testing of drug activity on side-effect targets.** *Nature.* 2012; **486**(7403): 361–367.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jalencas X, Mestres J: **On the origins of drug polypharmacology.** *Med Chem Comm.* 2013; **4**(1): 80–87.  
[Publisher Full Text](#)
- Mestres J, Gregori-Puigjané E, Valverde S, *et al.*: **Data completeness—the Achilles heel of drug-target networks.** *Nat Biotechnol.* 2008; **26**(9): 983–984.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Rognan D: **Chemogenomic approaches to rational drug design.** *Br J Pharmacol.* 2007; **152**(1): 38–52.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hu Y, Bajorath J: **Compound promiscuity: what can we learn from current data?** *Drug Discov Today.* 2013; **18**(13–14): 644–650.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Gaulton A, Bellis LJ, Bento AP, *et al.*: **ChEMBL: a large-scale bioactivity database for drug discovery.** *Nucleic Acids Res.* 2012; **40**(Database issue): D1100–D1107.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wang Y, Xiao J, Suzek TO, *et al.*: **PubChem's bioassay database.** *Nucleic Acids Res.* 2012; **40**(Database issue): D400–D412.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Knox C, Law V, Jewison T, *et al.*: **DrugBank 3.0: a comprehensive resource for 'omics' research on drugs.** *Nucleic Acids Res.* 2011; **39**(Database issue): D1035–41.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hu Y, Bajorath J: **Growth of ligand-target interaction data in ChEMBL is associated with increasing and activity measurement-dependent compound promiscuity.** *J Chem Inf Model.* 2012; **52**(10): 2550–2558.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hu Y, Bajorath J: **How promiscuous are pharmaceutically relevant compounds? A data-driven assessment.** *AAPS J.* 2013; **15**(1): 104–111.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hu Y, Bajorath J: **What is the likelihood of an active compound to be promiscuous? Systematic assessment of compound promiscuity on the basis of PubChem confirmatory bioassay data.** *AAPS J.* 2013; **15**(3): 808–15.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hu Y, Bajorath J: **Promiscuity profiles of bioactive compounds: potency range and difference distributions and the relation to target numbers and families.** *Med Chem Commun.* 2013; **4**: 1196–1201, in press.  
[Publisher Full Text](#)

## Current Referee Status:

---

### Referee Responses for Version 2



**Stefan Laufer**

University of Tübingen, Tübingen, Germany

**Approved: 01 August 2013**

**Referee Report:** 01 August 2013

as already said, a very valid approach for my (kinase) community to deal with multi-target compounds

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

**Competing Interests:** No competing interests were disclosed.



**Jeremy Jenkins**

Centre for Protein Chemistry, Novartis Institutes for BioMedical Research, MA, USA

**Approved: 29 July 2013**

**Referee Report:** 29 July 2013

My questions from v1 were sufficiently answered.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

**Competing Interests:** No competing interests were disclosed.

---

### Referee Responses for Version 1



**Jeremy Jenkins**

Centre for Protein Chemistry, Novartis Institutes for BioMedical Research, MA, USA

**Approved: 16 July 2013**

**Referee Report:** 16 July 2013

Hu and Bajorath's update on compound promiscuity in public compound bioactivity databases is timely with the increasing cognizance of polypharmacology and its role in the efficacy and safety of drugs. The article aims to raise reader awareness and present new questions to be answered, and so the title, abstract, and content are appropriate. All data are freely available for download for primary sources



mentioned.

The conclusions are fair and unbiased.

Additional questions do arise from this survey. First, does the average number of targets per compound differ from the median (or do highly promiscuous compounds skew the average?).

Second, is it reasonable to begin distinguishing promiscuous from privileged compounds? For example, by incorporating target class information, staurosporine might be viewed differently from quercetin, where the former represents a highly privileged scaffold among kinases and the latter displays IC50 values against an abundance of target types.

Third, the drug discovery field needs to understand if the "promiscuity enrichment" that occurs between the screening hits phase to the marketed drug phase largely reflects the depth of bioactivity data coverage for drugs, as drugs are highly profiled globally. The hit rates of drugs and medchem compounds across the same set of assays and targets would be needed to definitively conclude that drugs are more promiscuous.

However, the apparent increased promiscuity of drugs supports the growing resurgence of phenotypic screening, the impetus for exploring compound combinations in the context of multiple genotypes, and begs the question of how medchem optimization of multiple targets should be attacked.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

**Competing Interests:** No competing interests were disclosed.



**Hans Matter**

Sanofi-Aventis, Paris, France

**Approved: 09 July 2013**

**Referee Report:** 09 July 2013

This interesting manuscript presents a view on compound promiscuity based on in-vitro data and the number of potential targets per compound in public databases such as ChEMBL, PubChem and DrugBank. In particular the authors investigate and challenge the notion that most compounds today in lead findings are active on a large multitude of biological data. The title is appropriate for this contribution and the abstract sufficiently summarizes this study. The conclusions are balanced and justified on the basis of the data analysis; this is therefore an essential view on the number of targets.

It is an interesting observation from this study that DrugBank annotated drugs appear to interact with a higher number of molecular targets compared to early phase compounds in ChEMBL or PubChem. Any interpretation of this finding should be treated with caution, but it is tempting to discuss from a partially historical view as DrugBank may be enriched with older drugs that would have been subjected to less strict requirements for in-vitro selectivity than in today's drug discovery. In addition during and after approval, drugs may have been tested in more profiling assays as is the case with earlier screening-type substances.

Following the authors, this interesting argument also supports target-specific drug discovery paradigms used in past years. However, working with public databases leads to many caveats, all of which have been pointed out earlier, e.g. incompleteness of the data matrix and differences of data from different sources. It might be interesting for future investigations to cross-check this conclusion for compounds

targeting families like kinases or GPCRs. Due to the challenges of inherent selectivity in those families one could expect a larger percentage of promiscuous compounds. The same discrimination might possibly be true for smaller versus larger compounds.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

**Competing Interests:** No competing interests were disclosed.



**Stefan Laufer**

University of Tübingen, Tübingen, Germany

**Approved: 01 July 2013**

**Referee Report:** 01 July 2013

Excellent work. My field ("the kinase community") will benefit a lot from this commentary as compound promiscuity is an issue.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

**Competing Interests:** No competing interests were disclosed.

---

## Article Comments

### Comments for Version 1

#### Author Response

**Jürgen Bajorath**, Department of Life Science Informatics, B-IT and LIMES Institutes, Rheinische Friedrich-Wilhelms-Universität Bonn, Germany

Posted: 08 Jul 2013

Updates are available for the following references:

9. *Drug Discov Today* (2013); **18** (13-14): 644-650.

15. *AAPS J.* (2013); **15** (3): 808-815

16. *Med. Chem. Commun.* (2013), in press.

**Competing Interests:** No competing interests were disclosed.