COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

Volume No: 9, Issue: 13, e201401003, http://dx.doi.org/10.5936/csbj.201401003

CSBJ

Exploring Compound Promiscuity Patterns and Multi-Target Activity Spaces

Ye Hu [a,†], Disha Gupta-Ostermann [a,†], Jürgen Bajorath [a,*]

Abstract: Compound promiscuity is rationalized as the specific interaction of a small molecule with multiple biological targets (as opposed to non-specific binding events) and represents the molecular basis of polypharmacology, an emerging theme in drug discovery and chemical biology. This concise review focuses on recent studies that have provided a detailed picture of the degree of promiscuity among different categories of small molecules. In addition, an exemplary computational approach is discussed that is designed to navigate multi-target activity spaces populated with various compounds.

MINI REVIEW ARTICLE

Introduction

Over the past decade it has been increasingly recognized that many pharmaceutically relevant compounds are promiscuous in nature [1-3] and that many drugs elicit their therapeutic effects -and undesired side effects- through polypharmacology [4,5]. For a number of drugs that were originally considered to be target-selective or -specific, high degrees of promiscuity and ensuing polypharmacology have been shown to be responsible for their efficacy, with protein kinase inhibitors applied in oncology being a prime example [6]. In addition, polypharmacology also provides the basis for drug repurposing [7-9], another current topic of high interest in pharmaceutical research.

Given that compound promiscuity represents the molecular basis of polypharmacological effects, a detailed assessment of the degree of promiscuity among compounds at different stages of the drug development pathway is of considerable interest. The unprecedented recent growth of compound activity data in the public domain has made it possible to approach this question through data mining. This is illustrated in **Figure 1**, which shows a drug-target network generated on the basis of known target annotations of approved drugs, reflecting a generally high degree of drug promiscuity. In promiscuity analysis, most efforts have thus far concentrated on elucidating the promiscuous nature of drugs, often by database analyses combined with computational predictions. Recent estimates have been that a drug might on average interact with ~3-6 targets and that 50% of all drugs might exhibit activity against more than five targets [5,10].

Results of data mining efforts are generally affected by data incompleteness [10], i.e., not all compounds have been tested against all targets (and probably will never be). However, given increasingly large amounts of compound activity data that become available at present (much more than one could have imagined just a few years ago), reliable trends can already be detected and some meaningful conclusions drawn from them [11].

Herein, we review recent insights into promiscuity of screening hits, bioactive compounds, and drugs obtained through systematic mining of compound activity data. All currently investigated aspects of promiscuity are discussed. In addition, we introduce a computational and graphical framework for the analysis of multi-target activity spaces and compound promiscuity patterns.** The interested reader is also referred to other recent reviews of compound promiscuity [11,12].

**This review is based upon the presentation *'Compound Data Mining in Systems Chemical Biology: Exploring Multi-target Activity Spaces and Compound Promiscuity Patterns'* given by one of us (J.B.) at the 2013 Meeting of the International Chemical Biology Society (ICBS2013). The program of ICBS2013 can be viewed via the following URL: http://www.chemical-biology.org/?page=ICBS2013Schedule).

Activity data of compounds from different sources

In order to comprehensively assess compound promiscuity, various types of compounds at different pharmaceutical development stages should be considered. A large number of relevant compounds and associated activity data can currently be collected from several public repositories.

Screening hits

The PubChem BioAssay database [13] contains bioactivity information from confirmatory high-throughput screens including confirmed active and inactive compounds. To ensure high data confidence, a pre-requisite for meaningful data mining efforts [11], a total of 1085 confirmatory assays with reported activity against a single protein target and dose-response data were extracted from PubChem in January 2013 [14]. These assays involved 437,288 compounds and 439 targets.

A subset of 140,112 compounds was confirmed to be active in one or more assays, representing screening hits at the early stages of drug discovery. More than 77% of these hits were tested in more than 50 assays, hence providing a sound basis for promiscuity analysis [14], as discussed below.

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

[†]These authors contributed equally to this work

* Corresponding author. Tel.: +49 2282699306; Fax: +49 228269934l
E-mail address: bajorath@bit.uni-bonn.de (Jürgen Bajorath)

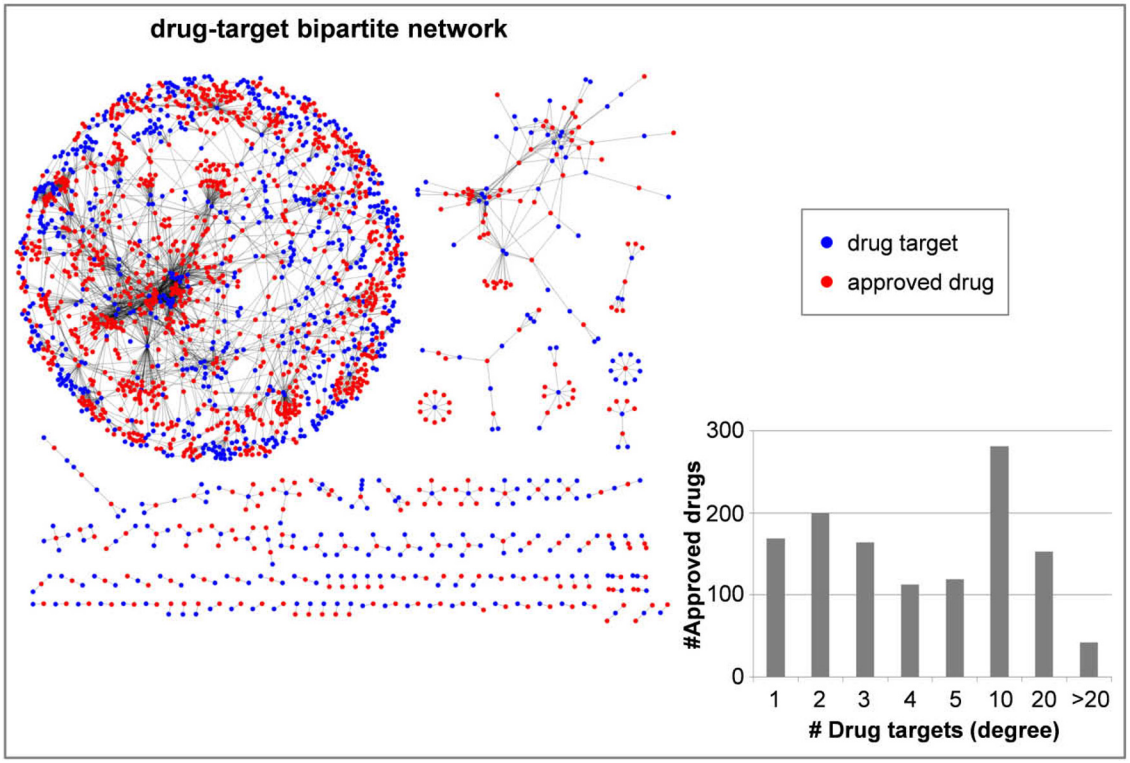

Figure 1. Drug-target interactions. Shown is an approved drug-target bipartite network. Red nodes represent approved drugs from DrugBank 3.0 and blue nodes drug targets. Edges between red and blue nodes indicate known drug-target interactions. In total, there are 3776 drug-target interactions between 1226 approved drugs and 881 targets. Similar yet distinct drug-based target networks have earlier been introduced by Yildirim et al. [29]. The insert reports the distribution of the degree of approved drug nodes, indicating the number of targets they were active against.

Bioactive compounds

The rapidly growing ChEMBL database [15] has become a major public repository of compound activity data obtained from medicinal chemistry sources. Currently, ChEMBL release 17 contains 1,324,941 distinct compounds with 12,077,491 activity annotations. It should be noted that the original investigations reviewed herein were carried out over time on different versions of ChEMBL (the versions were specified in each case).

To obtain high-confidence activity data from ChEMBL, only compounds with direct interaction against human targets at highest confidence level were extracted. Two types of potency measurements were separately considered, equilibrium constants (K_i) and assay-dependent IC_{50} values. Compounds with approximate potency annotations (i.e., ">", "<", "~") were excluded. From ChEMBL release 14, 36,542 compounds active against 579 targets were collected that yielded 62,913 explicit K_i values, comprising the K_i subset. In the IC_{50} subset, there were 80,522 compounds active against 1129 targets with 114,092 IC_{50} measurements [16]. These bioactive molecules, especially those from the K_i subset, were predominantly taken from medicinal chemistry literature and patent sources and hence mostly represented compounds at the hit-to-lead and lead optimization stages.

Experimental and approved drugs

The DrugBank database [17] is a public resource that contains drug entries, including approved small molecule drugs, approved biologicals, nutraceuticals, and experimental drugs (including compounds in clinical trials), with associated drug target information. For promiscuity analysis, 1274 approved small molecule drugs and 4931 experimental drugs with available structures were assembled from DrugBank 3.0. These approved drugs and drug candidates represented compounds at the late drug development stages.

Compound promiscuity rates

From these different data repositories, promiscuous compounds were extracted and promiscuity rates calculated as the average number of targets compounds were active against. In all cases reported herein, promiscuity rates were determined for compounds active against multiple targets, i.e., excluding compounds with reported single-target activity. Taking compounds with single-target activity into account would have reduced average promiscuity rates.

From 140,112 PubChem screening hits, 71,303 compounds (~50.9%) were identified to be active against two or more targets [14]. In addition, for the K_i and IC_{50} subsets of ChEMBL version 14, 13,842 (~37.9%) and 19,898 compounds (~24.7%) were identified to be promiscuous, respectively [16]. These compounds were active against a total of 459 and 867 human targets in the K_i and IC_{50} subsets, respectively. Furthermore, compound overlap between these two subsets was established on the basis of database IDs. There were 1025 promiscuous compounds conserved in both subsets. The remaining 12,817 and 18,873 promiscuous compounds were exclusively found in the K_i and IC_{50} subsets, respectively. In general, the IC_{50} subset contained > 6000 more promiscuous compounds than the K_i subset. Furthermore, 1072 approved (~84.1%) and 1113 experimental (~23.6%) drugs from DrugBank had multiple target annotations. For compounds from different sources, promiscuity rates are reported in **Figure 2a**. On average, promiscuous compounds from PubChem confirmatory assays were active against 3.7 targets. Bioactive compounds from the K_i and IC_{50} subsets of ChEMBL