Original article

Many drugs contain unique scaffolds with varying structural relationships to scaffolds of currently available bioactive compounds

Ye Hu, Jürgen Bajorath*

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

ABSTRACT

Molecular scaffolds were systematically extracted from approved drugs and analyzed. The majority of drug scaffolds, 552 of 700, were found to represent only a single drug. Moreover, 221 drug scaffolds were not detected in currently available bioactive compounds, i.e., the pool from which drug candidates usually originate. These "drug-unique" scaffolds displayed a variety of structural relationships to currently known bioactive scaffolds, reflecting rather different degrees of relatedness. Many drug-unique scaffolds formed only very limited structural relationships to bioactive scaffolds. These drug scaffolds should represent promising candidates for further chemical exploration and drug repositioning efforts and are made freely available.

© 2014 Elsevier Masson SAS. All rights reserved.

1. Introduction

The scaffold concept is popular in medicinal chemistry because it enables the organization of compounds according to core structures, the association of cores with biological compound activities, the search for privileged substructures, and the design of new compound series [1,2]. In addition, the scaffold concept has also been applied to analyze and compare different drugs and better understand key structural features [3–5]. Scaffolds can be defined in different ways [2]. The scaffold definition most widely applied in medicinal chemistry was originally introduced by Bemis and Murcko [3]. This definition followed a molecular hierarchy by dividing compounds into R-groups, linkers, and rings. Accordingly, Bemis–Murcko (BM) scaffolds are obtained from compounds by removing R-groups but retaining aliphatic linker fragments between rings [3]. Thus, BM scaffolds generally represent cores consisting of single or multiple ring systems that can be connected in different ways. Thus, they account for molecular topology. From these scaffolds, one can further abstract by converting all heteroatoms to carbon and setting all bonds to single bonds (i.e., setting all bond orders to 1) [3,6]. These modifications generate so-called cyclic skeletons (CSKs) [6]. Hence, a given CSK covers a set of topologically equivalent scaffolds that are only distinguished by

heteroatom substitutions and/or bond orders. It follows that different CSKs represent topologically distinct scaffolds.

A primary focal point of scaffold analysis in medicinal chemistry has been — and continues to be — the association of scaffolds with biological activities of compounds they represent [7–11]. Different approaches have been introduced to systematically derive and organize scaffolds on the basis of retrosynthetic information [12], structural similarity criteria [13], structural rule-based scaffold decomposition [14], or compound-scaffold-CSK hierarchies [15]. Such structural organization schemes can substantially aid in the association of scaffolds with biological activities and the analysis of structure–activity relationships (SARs). For example, the Layered Skeleton-Scaffold Organization (LASSO) graph has been used to systematically explore SARs in compound data sets along molecular hierarchies [15]. Moreover, the Scaffold Tree that is based on structural rule-based decomposition [14] has not only been utilized in SAR analysis but also to generate virtual scaffolds within experimental scaffold hierarchies for the prediction of novel active compounds [11,16,17].

While scaffolds in active compounds and drugs have been analyzed in a variety of ways, as discussed above, only few investigations have thus far systematically compared drug scaffolds with scaffolds originating from other bioactive non-drug compounds. Such comparisons might help to better understand whether there are specific differences between core structures from drugs and bioactive compounds. In one study, it was shown

* Corresponding author.
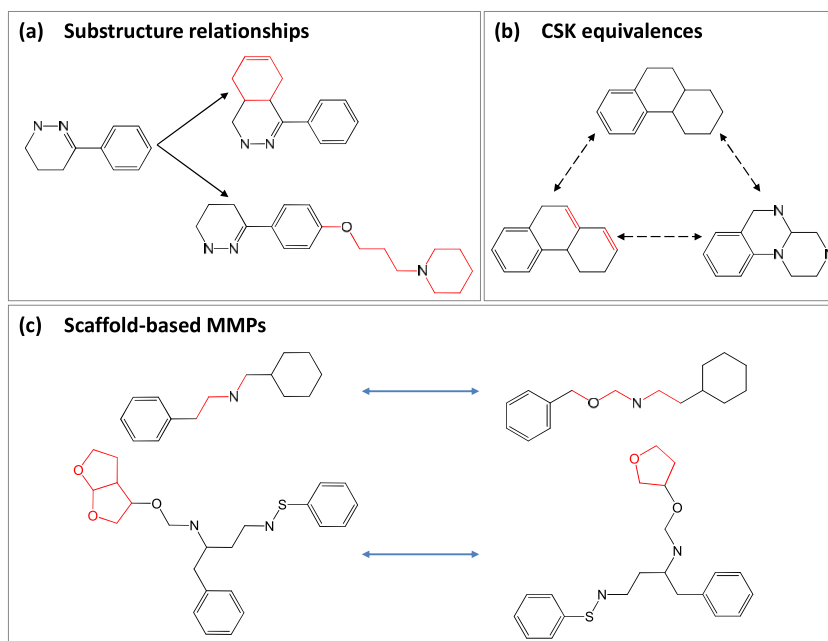E-mail address: bajorath@bit.uni-bonn.de (J. Bajorath).

Fig. 1. Structural relationships. Three different types of structural relationships are illustrated including (**a**) substructure, (**b**) topological (CSK equivalences), and (**c**) MMP-based relationships. For each pair of scaffolds, the structural differences are highlighted in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that there was only limited overlap between BM scaffolds isolated from sets of compounds at different pharmaceutical development stages including active compounds, compound in clinical trials, and drugs [17]. In another analysis, ~13,000 BM scaffolds were extracted from compounds active against ~450 targets belonging to 19 different families and their activity profiles were determined [18]. More than 400 scaffolds were identified that were active against targets from at least two target families and a subset of 83 scaffolds was active against targets from three to 13 families. These 83 scaffolds yielded 33 distinct CSKs, 17 of which were detected in more than 200 approved drugs. Hence, this analysis demonstrated that scaffolds with multi-target activities were well represented in current drugs [19], consistent with the notion of drug poly-pharmacology [20,21].

The collection of scaffolds extracted from available approved drugs can be rationalized as a basic structural representation of known drug space [22], which is distinct from drug-like chemical space. Current known drug space has many origins reflecting the history of drug discovery and development (including classical pharmacological testing, molecular approaches, rational design, etc.). To further extend the exploration of drug scaffolds and analysis of known drug space, we have carried out a systematic structural comparison of scaffolds from approved drugs and from the large pool of currently available bioactive compounds that addresses three previously unexplored questions. First, how are scaffolds distributed across approved drugs? Second, to what extent are drug scaffolds represented in bioactive compounds? Third, which structural relationships exist between drug scaffolds and scaffolds from bioactive compounds? Our analysis has yielded a number of rather unexpected findings that are reported herein.

2. Experimental

2.1. Scaffolds from bioactive compounds and drugs

From the latest version of ChEMBL (release 17) [23], compounds with direct interactions (i.e., target relationship type "D") against human targets at the highest confidence level (i.e., target confidence score 9) and available equilibrium constants (K_i values) as activity measurements were extracted. From DrugBank 3.0 [24], approved small molecule drugs with available structures and activity information were collected. From all bioactive compounds and approved drugs, BM scaffolds [3] and CSKs [6] were isolated. As a control, scaffolds were also extracted from bioactive compounds for which IC_{50} measurements were available. In the following, BM scaffolds are simply referred to as scaffolds and scaffolds extracted from bioactive compounds and drugs are designated bioactive scaffolds and drug scaffolds, respectively. Bioactive scaffolds and drug scaffolds were systematically compared. Initially, the overlap between these two scaffold sets was determined. Then, different types of structural relationships between a subset of drug scaffolds and bioactive scaffolds were systematically explored.

2.2. Structural relationships

Three types of structural relationships between drug scaffolds and bioactive scaffolds were analyzed:

(1) Substructure relationship: a scaffold is contained as a sub-structure in another. Benzene, the most generic scaffold, was excluded from the assessment of substructure relationships (to avoid an inflation of substructure matches for this scaffold). The size of scaffolds with substructure relationships was compared by determining the difference in the number of rings they contained.

(2) Topology relationship: if two scaffolds share the same topology, they yield the same CSK. Cyclohexane, the CSK of benzene, was excluded from the assessment of CSK equivalence.

(3) Matched molecular pair (MMP) relationship: an MMP is formed by compounds that differ only at a single site by the exchange of a pair of substructures [25], termed a chemical transformation [26]. Transformation size-restricted MMPs [25,27] were calculated for drug scaffolds vs. bioactive scaffolds using our implementation of the algorithm by Hussain and Rea [27] that utilizes the OpenEye toolkit [28]. Size-restricted MMPs limit chemical transformations to