Pharmacology & Therapeutics 175 (2017) 47–66

Contents lists available at ScienceDirect

Pharmacology & Therapeutics

journal homepage: www.elsevier.com/locate/pharmthera

Associate editor: N. Frossard

The impact of in silico screening in the discovery of novel and safer drug candidates

Didier Rognan

Laboratoire d'Innovation Thérapeutique, UMR 7200 CNRS-Université de Strasbourg, 74 route du Rhin, 67400 Illkirch, France

ARTICLE INFO

Available online 14 February 2017

Keywords:
Screening
Hit
Target
Profile
Safety
Pharmacokinetics

ABSTRACT

Drug discovery is a multidisciplinary and multivariate optimization endeavor. As such, in silico screening tools have gained considerable importance to archive, analyze and exploit the vast and ever-increasing amount of experimental data generated throughout the process. The current review will focus on the computer-aided prediction of the numerous properties that need to be controlled during the discovery of a preliminary hit and its promotion to a viable clinical candidate. It does not pretend to the almost impossible task of an exhaustive report but will highlight a few key points that need to be collectively addressed both by chemists and biologists to fuel the drug discovery pipeline with innovative and safe drug candidates.

Contents

1. Introduction

Drug discovery, as any other discipline, is accumulating experimental data at an exponential pace. Due to the sequential and multidisciplinary nature of drug discovery pipelines, archiving and efficient mining of key compound and target properties (e.g. structural, physicochemical, biochemical, pharmacological, toxicological) is crucial for a better understanding and prediction of the developability of a given compound. These good practices are supposed to reduce the overall attrition rates (Hay, Thomas, Craighead, Economides, & Rosenthal, 2014) and therefore lead to a significant decrease of the drug development costs (DiMasi, Grabowski, & Hansen, 2016).

Is therefore not surprising that in silico methods have gained so much importance in drug discovery. This trend can be simply illustrated by the herein reported survey of several key descriptors for chemical/biological space and computing power (Fig. 1). On the one hand, there are currently over 110 million chemicals registered by the Chemical Abstracts Service, out of which only 1.5% exhibit known biological activity (Gaulton et al., 2012). On the other hand, about 11,000 pharmacological targets are known up to date (Gaulton et al., 2012), giving rise to 125,000 different three-dimensional structures (Berman et al., 2000). Both compound and target counts experience an exponential growth that mirrors the growth in computing power. This, expressed by the number of transistors in microprocessors, follows the well-known Moore's law stating that the count of the integrated circuits doubles approximately every two years. It is therefore not surprising that the application of in silico technologies in drug discovery literature also experiences an exponential growth with 4–5 PubMed citations every day (Fig. 1).

In silico technologies may be applied at any of the numerous possible stages of drug discovery and the review their overall applicability

E-mail address: rognan@unistra.fr.

http://dx.doi.org/10.1016/j.pharmthera.2017.02.034
0163-7258/

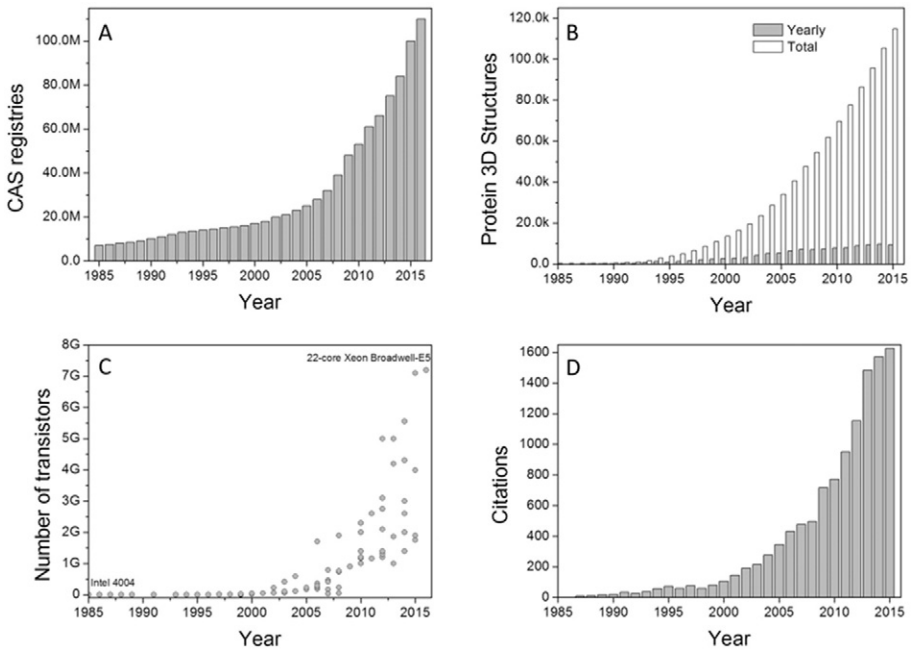

Fig. 1. Variation over time of a few key properties relevant for rational drug discovery. A) Registered substances in the Chemical Abstract Service (http://www.cas.org). B) Entries in the Protein Data Bank Berman et al. (2000); C) Transistor count in microprocessors (https://en.wikipedia.org/wiki/Transistor_count); D) Citations with the following combination of keywords "in silico" AND ("*drug discovery*" OR "*drug design*") in the PubMed resource (http://www.ncbi.nlm.nih.gov/pubmed).

falls outside of the scope of the present article. Here, we will focus on any computerized method to assist chemists and biologists in the preclinical development of drug candidates, ranging from target validation, compound library design, hit identification, hit-to-lead optimization and preclinical candidate identification. To illustrate the integration of computational design in pharmaceutical companies, it is worth mentioning a recent report from Bayer HealthCare (Hillisch, Heinrich, & Wild, 2015) stating that half of the 20 new chemical entities (NCEs) currently being tested in phase I clinical trials really benefited from computer-aided design methods.

2. Target validation

Target-related safety issues have recently been shown to be the major cause of attrition in clinical trials at a big pharmaceutical company (Cook et al., 2014). It is therefore of utmost importance to carefully select the right target before entering costly compound screening processes. Considering validated targets as those to which FDA-approved drugs physically bind, we have progressively learned that: (i) targeting certain protein families (e.g. G protein-coupled receptors, protein kinases) reduces the probability of early closures (Hopkins & Groom, 2002; Rask-Andersen, Masuram, & Schioth, 2014), (ii) specific pockets to which launched drugs associate exhibit a well-defined range of physicochemical properties (e.g. hydrophobicity, accessibility, curvature) that are distinct from that of less druggable targets like protein-protein interfaces (Kuenemann, Bourbon, Labbe, Villoutreix, & Sperandio, 2014). However, there is still an urgent need for computational methods that would robustly reduce risks associated with a particular target selection. Of course, "druggability" is by far more complex than the simple propensity of a particular protein cavity to accommodate high-affinity bioavailable drug-like compounds. Other terms like "ligandability" (Edfeldt, Folmer, & Breeze, 2011) or "bindability" (Sheridan, Maiorov, Holloway, Cornell, & Gao, 2010) have recently been proposed since they better capture target property ranges (cavity volume, polarity and buriedness) known to be important for druggable targets (Cheng et al., 2007). The most conservative way to define druggable target space is to identify those targets that do physically associate with approved small molecular-weight drugs. One of the most recent surveys (Rask-Andersen, Almen, & Schioth, 2011) identified 989 small molecular-weight drugs acting on 435 therapeutic effect-mediated human targets. In addition, drug-target interactions (Rask-Andersen et al., 2014) suggest 475 potentially novel drug targets in addition those previously identified.

Altogether, three kinds of methods for predicting target druggability can be distinguished: methods based on the target's sequence, its three-dimensional structure, or its integration in more complex systems biology networks. Whatever the method, the first step is to define the instances (targets, drugs, networks) to which usually machine learning algorithms (Jordan & Mitchell, 2015) are applied in order to establish non-linear relationships between descriptors and the property to predict (Fig. 2). Many specialized databases storing this information are freely accessible (Table 1).

The most straightforward method to estimate target druggability relies on different amino acid sequence descriptors (e.g. amino acid composition, physicochemical properties) of known drug targets and putative non-drug targets (or targets still awaiting approved drugs). Such models usually report accuracies of 85–95% (Bakheet & Doig, 2009; Li & Lai, 2007), but are optimistic because of an oversimplified definition of the large non-druggable target space (any target not explicitly defined as a drug target). As a consequence, sequence-based classification tends to reward entire protein subfamilies as potentially druggable (Li & Lai, 2007) although experimental screening data usually indicates the opposite. Moreover, sequence-based models are hard to interpret and are not linked with any particular domain or pocket on which to focus hit identification efforts.

Structure-based methods are therefore much more popular to predict target druggability. They rely on 3D structural descriptors (polarity, hydrophobicity, buriedness, volume, curvature) of ligand-bound cavities in both druggable and undruggable targets to learn rules able to optimally distinguish both categories in a binary manner. Current state-of-the-art tools (Borrel, Regad, Xhaard, Petitjean, & Camproux, 2015; Desaphy, Azdimousa, Kellenberger, & Rognan, 2012; Krasowski, Muthas, Sarkar, Schmitt, & Brenk, 2011; Schmidtke & Barril, 2010; Volkamer, Kuhn, Grombacher, Rippmann, & Rarey, 2012) exhibit an accuracy of approximately 85% for conventional targets (GPCRs, kinases). The main advantage of such methods is their high interpretability in