Methods 71 (2015) 98–103

Contents lists available at ScienceDirect

Methods

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

Tools for *in silico* target fishing

Adrià Cereto-Massagué [a], María José Ojeda [a], Cristina Valls [a], Miquel Mulero [a], Gerard Pujadas [a,b], Santiago Garcia-Vallve [a,b,*]

[a] Group of Cheminformatics & Nutrition, Biochemistry and Biotechnology Department, Universitat Rovira i Virgili (URV), Tarragona, Catalonia, Spain
[b] Centre Tecnològic de Nutrició i Salut (CTNS), TECNIO, Reus, Spain

ARTICLE INFO

Article history:
Received 28 June 2014
Revised 18 September 2014
Accepted 19 September 2014
Available online 30 September 2014

Keywords:
Computational target fishing
Reverse screening
Drug repositioning
Polypharmacology
Drug research
Functional foods

ABSTRACT

Computational target fishing methods are designed to identify the most probable target of a query molecule. This process may allow the prediction of the bioactivity of a compound, the identification of the mode of action of known drugs, the detection of drug polypharmacology, drug repositioning or the prediction of the adverse effects of a compound. The large amount of information regarding the bioactivity of thousands of small molecules now allows the development of these types of methods. In recent years, we have witnessed the emergence of many methods for *in silico* target fishing. Most of these methods are based on the similarity principle, i.e., that similar molecules might bind to the same targets and have similar bioactivities. However, the difficult validation of target fishing methods hinders comparisons of the performance of each method. In this review, we describe the different methods developed for target prediction, the bioactivity databases most frequently used by these methods, and the publicly available programs and servers that enable non-specialist users to obtain these types of predictions. It is expected that target prediction will have a large impact on drug development and on the functional food industry.

1. Introduction

In contrast to virtual screening, which is used to search large libraries of compounds for molecules that are most likely to bind a specific target, the aim of reverse screening, also known as *in silico* or computational target fishing [1,2] or reverse pharmacognosy [3], is to identify the most likely targets of a query molecule. This approach allows the prediction of the bioactivity of the query molecule or its mechanism of action. In addition, these techniques can be used to predict the adverse effects of a compound [4,5], to detect drug polypharmacology [6–8], or to reposition drugs [7,9–13].

Known drugs have, on average, six molecular targets on which they exhibit activity [14]. Polypharmacology, the ability of small molecules to interact with multiple proteins, is of particular interest for rationally designing more effective and less toxic drugs. Drug repositioning, the process of finding new uses for known drugs, is a promising way to explore alternative indications for existing drugs [13]. Because the successful launch of a single new drug is estimated to cost approximately U.S. $800 million and takes a staggering 15 years, and because very few compounds that start a clinical trial emerge to the market [10], finding new uses for old drugs could be economically advantageous.

Taking into account that several databases, such as ChEMBL, contain millions of molecules and information about their bioactivity, it is now becoming feasible to merge the known "chemical space" and "biological space" into models that will enable us to generate biological "spectra" to predict the phenotypic activity of new molecules based on their chemical structures and the known bioactivities of structurally similar compounds [15]. Although the current methods of virtual screening could be successfully adopted for target fishing, the differences in the general tasks of these methods justify the independent development of new *in silico* techniques for target fishing.

2. Computational methods for target fishing

Various computational methods have been developed to predict the molecular targets of a compound [1,16]. These methods were initially classified into four groups: chemical similarity searching, data mining/machine learning, panel docking, and the analysis of bioactivity spectra [16,17]. Recently, other classes, such as protein-structure-based methods, have been proposed [18]. Below, we summarize the main characteristics of some of these methods.

* Corresponding author at: Group of Cheminformatics & Nutrition, Biochemistry and Biotechnology Department, Universitat Rovira i Virgili (URV), Tarragona, Catalonia, Spain.
E-mail address: santi.garcia-vallve@urv.cat (S. Garcia-Vallve).

http://dx.doi.org/10.1016/j.ymeth.2014.09.006
1046-2023/

2.1. Molecular similarity methods

This section describes chemical similarity methods and shape-based similarity methods. The simplest methods for target prediction are based on chemical similarity and the use of current knowledge about the bioactivity of millions of small molecules. These methods are based on the "chemical similarity principle," which states that similar molecules are likely to have similar properties [19,20]. Thus, the targets of a molecule can be predicted by identifying proteins with known ligands that are highly similar to the query molecule [16]. The advantage of these methods is that they only require the computation of the similarity between compounds [19,21]. An outline of a chemical similarity method is shown in Fig. 1. In this method, a small molecule is represented as a chemical fingerprint. Fingerprints are a way of encoding the structure of a molecule. The most common type of fingerprint is a series of binary digits (bits) that represent the presence or absence of particular substructures in the molecule. The interested reader is referred to [22] for a review about fingerprints. To compare the fingerprints of two molecules, the Tanimoto coefficient or any other similarity criterion can be used. The more similar two compounds are, the closer the Tanimoto coefficient will be to 1. Several databases describing the bioactivities of thousands or millions of small molecules or the activities of known drugs can be used for target prediction (see Table 1 and reference [1]).

Keiser et al. [23] used a similarity ensemble approach to compare protein targets by the 2D similarity of the ligands that they are known to bind. The authors screened a dataset of 3665 drugs, including drugs approved by the FDA and investigational drugs, against a database of 65,241 ligands organized into 246 protein targets taken from the MDL Drug Data Report database. Their study revealed unanticipated associations between thousands of drugs and ligand sets [23]. Of the 30 most promising drug-target associations that were tested experimentally, 23 were confirmed, and 5 of the 23 were shown to be potent (<100 nM) modulators of their predicted target [23]. Thus, their study demonstrated the power of using simple ligand-based similarity searches.

Because they can be calculated quickly, 2D fingerprints have been widely used for similarity searching in target fishing. However, 3D chemical descriptors can also be used [17], although calculating them is computationally more expensive. Because they contain more information, the predictions based on 3D fingerprints would be expected to be better than those based on 2D fingerprints. However, in some cases, methods that use 2D fingerprints outperform those methods that use 3D fingerprints in correct target prediction [24]. 3D descriptors work better in cases of low structural similarity [24].

A known limitation of chemical similarity approaches is that inactive compounds can sometimes exhibit good similarity with active molecules if they have been obtained by modifying an active compound at some key position that was crucial for its interactions [25]. These inactive compounds can be false positive predictions of target fishing methods. In addition, in some cases, a large group of false negatives is also expected, because not all types of active compounds for a specific target have been identified.

Shape-based similarity methods use 3D shape comparisons between molecules, usually comparing the shape of the molecular volume, but other "shapes" can be compared, like the electrochemical surface. This can be done with software such as ROCS [26], Phase Shape [27], ESHAPE3D [28], PARAFIT [29], ShaEP [30] and USR [31] as some examples. Shape-based methods have the potential of detecting similarities between molecules with different atomic structures, thus making them specially useful for scaffold-hopping. Pharmacophores and some molecular fingerprints (like spectrophores [32] and many pharmacophore-based fingerprints [33]) can also include 3D information [22,33]. All these 3D methods require ligand conformations. In many cases (where there is no known biologically active conformation for the molecule), a single low-energy conformer is used, although it can be biologically irrelevant. Another approach is to get the conformation of the molecules by aligning them to a known bioactive conformation of a known ligand. However, 2D fingerprint-based methods give better performance than 3D shape-based methods in virtual screenings [34]. In other cases, combining chemical and shape similarity measures significantly increases the target prediction accuracy [35].

After obtaining the highest similarity coefficient between a query compound and the compounds in an annotated database, it is important to assess the statistical significance of the similarity. Two structures are usually considered similar if the Tanimoto coefficient between them is higher than 0.85. However, this value is not always reliable [36]. Keiser et al. [37] used an E-value computed from the 2D similarity with the set of ligands of a target. This *E*-value is derived from the statistics of similarity values with all ligands (above a certain threshold), and it indicates how likely it would be to find a molecule with a given average similarity to the set of ligands of a target. The SwissTargetPrediction server uses

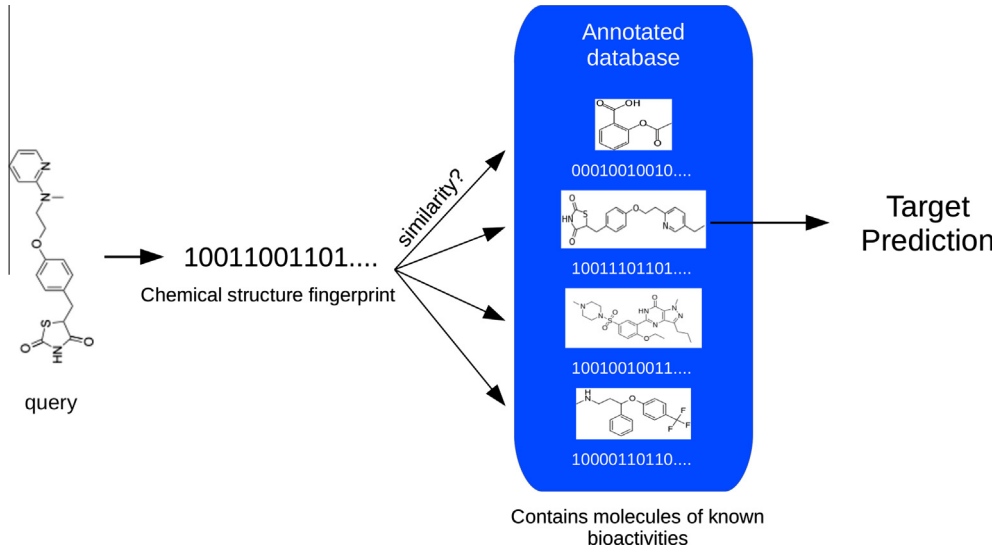

Fig. 1. Chemical similarity through the comparison of fingerprints can be used to predict novel targets or functions of a query molecule.