Methods 71 (2015) 113–134

Contents lists available at ScienceDirect

Methods

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

Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling

Anna Vuorinen, Daniela Schuster *

Institute of Pharmacy/Pharmaceutical Chemistry and Center for Molecular Biosciences Innsbruck – CMBI, University of Innsbruck, Innrain 80/82, 6020 Innsbruck, Austria

ARTICLE INFO

Article history:
Received 1 June 2014
Received in revised form 29 September 2014
Accepted 14 October 2014
Available online 22 October 2014

Keywords:
Pharmacophore modeling
Virtual screening
3D-structures
Conformational models
Bioactivity predictions

ABSTRACT

Biological effects of small molecules in an organism result from favorable interactions between the molecules and their target proteins. These interactions depend on chemical functionalities, bonds, and their 3D-orientations towards each other. These 3D-arrangements of chemical functionalities that make a small molecule active towards its target can be described by pharmacophore models. In these models, chemical functionalities are represented as so-called features. Commonly, they are obtained either from a set of active compounds or directly from the observed protein–ligand interactions as present in X-ray crystal structures, NMR structures, or docking poses. In this review, we explain the basics of pharmacophore modeling including dataset generation, 3D-representations and conformational analysis of small molecules, pharmacophore model construction, model validation, and its benefits to virtual screening and other applications.

1. Introduction

When a small molecule enters an organism, such as the human body, it has thousands of proteins to potentially interact with. Which of these proteins (typically enzymes, receptors, and transporters) it chooses as targets, is defined by chemical interactions: the small molecule binds to the proteins, where the ligand–protein interactions are energetically favorable. Already in 1890, Emil Fischer described the lock-key hypothesis on enzyme-activity: only the key with the right size and shape opens a specific lock, and similarly an enzyme chooses its substrate [1]. The same principle is also suitable for receptors. In the chemistry world, this translates to the theory that only the small molecules with the right size and correct complementary chemical functionalities can bind to the target protein and cause a biological effect (Fig. 1).

The chemical functionalities of the amino acid residues in the binding site, the binding site size, and its shape determine which small molecules it tolerates. Therefore, all molecules binding to the same binding site share similar chemical functionalities, size, and shape restrictions. The chemical functionalities that are needed for a small molecule to block or activate its target protein can be represented as pharmacophore models [2]. The concept of pharmacophore has evolved since the early 1900s: First, it was describing actual chemical functionalities called haptophore and toxophore by Paul Ehrlich. Since the 1960s, the modern term pharmacophore describing the chemical functionalities as abstract features was introduced by Schuler [3–5]. The basic theory behind pharmacophore modeling is that common chemical functionalities in similar 3D arrangements lead to a biological activity on the same target. Pharmacophore models consist of a defined 3D arrangement of so-called features that represent the chemical functionalities of active small molecules: hydrogen bond acceptors (HBAs), hydrogen bond donors (HBDs), hydrophobic areas (H), positively and negatively ionizable groups (PI/NI), and metal coordinating areas (M). Additional size restrictions in the form of a shape or exclusion volumes (XVOL) – forbidden areas – can be added to represent the size and the shape of the binding pocket. Since the models themselves do not focus on actual atoms, but chemical functionalities, they are good tools in recognizing similarities between molecules. For example, a simple hydrogen bond donor could be an NH_2-group or an OH-group.

In relation to Fischer's theory, pharmacophore models work like soap prints of the keys that fit to a lock: before trying all the available keys to the lock, they are first fitted to the soap print, and the ones that definitely do not fit can be excluded from the trial already. Thus, the aim of pharmacophore modeling and pharmacophore-based virtual screening is to predict activities by sorting the compounds into actives (compounds that match the model) and inactives (compounds not fitting to the model). The output of such a screen is a list of compounds (hit list) that are proposed to be active. Therefore, the advantage of pharmacophore-based virtual

* Corresponding author. Fax: +43 512 507 58299.
E-mail address: Daniela.Schuster@uibk.ac.at (D. Schuster).

http://dx.doi.org/10.1016/j.ymeth.2014.10.013
1046-2023/

screening in the drug discovery process is that most of the compounds with low probability to be active can be excluded from further studies already in a very early stage of a project. Thereby, a lot of resources in the further drug discovery process, especially in *in vitro* experiments, can be saved: To find 50 new active compounds for a specific target, thousands of molecules need to be tested if one uses *in vitro* high-throughput screening methods, but with the help of pharmacophore modeling, one only has to evaluate a few hundred compounds experimentally [6,7].

Pharmacophore models can be constructed *via* two ways: ligand-based and structure-based. The ligand-based method is used when no 3D structure of the target protein is available, but there is at least information on active molecules for this target. The structure-based method is applicable in cases where the 3D structure of the protein is known, e.g. as an X-ray crystal structure (ideally in complex with an active small molecule ligand), NMR structure, or homology model. In addition, Klabunde et al. and Sanders et al. [8,9] introduced a way to generate pharmacophore models from a target protein sequence without any information of the crystal structure or ligands. In their method, they derived structure-based pharmacophore models from G-protein coupled receptor crystal structures and their homology models. Each of these pharmacophore models was analyzed and the pharmacophore feature-interacting residues were identified. In case the interaction was present in multiple cases, it was marked and stored as a residue-feature-pair. These kinds of pairs can be then applied to any G-protein coupled receptor sequence: in case a specific residue is found, the corresponding feature with its coordinates will be added to a pharmacophore model. This method enables therefore pharmacophore model generation based on a protein sequence only. However, it requires pre- calculated data on a protein family with high sequence- and structural identity.

The development of a high quality pharmacophore model is a multi-step procedure. Independent from the generation method, a pharmacophore model should be first theoretically validated before applying it to prospective virtual screening. After virtual screening, compounds from the hit list can be experimentally validated. Depending on the results, the pharmacophore model can be improved using the newly generated activity data. Later, additional pharmacophore model refinement should be done if the model did not perform well in experimental validations or if there is new information on active compounds that do not support the old model hypothesis [10].

In this publication, we comprehensively review the principles of pharmacophore modeling: model construction, its theoretical validation, use as virtual screening filter, application to structure–activity relationship-predictions, and as a bioactivity profiling tool. We guide the reader through the model generation and validation process. Finally, we also outline the limits and future challenges for the pharmacophore modeling field.

2. Pharmacophore modeling

Pharmacophore modeling, as every virtual screening study, begins with a thorough literature survey: What is already known about the target? How is the binding site composed? Are there already known small molecules that bind to the target? Is there already a 3D structure (crystal structure or homology model) of the target? For ligand-based pharmacophore generation, at least two active molecules or one rigid, highly active compound are needed. In case of structure-based modeling, a 3D-protein structure with a bound active ligand is for a valid starting point for pharmacophore model generation.

2.1. Dataset generation

For model generation and theoretical validation, different kinds of datasets are needed and information has to be collected from various sources. Because the models represent the data they are based on, the final quality of the model highly depends on the input data: If the starting data is incorrect or unreliable, the model will also be low quality.

One important aspect in collecting the compounds for model generation and validation is, that the structures need to be correct and curated [11], the activity data should origin from comparable *in vitro* methods, identical species, and directly measure the investigated interaction [12]. Ideally, activity data derive from a target-based, cell-free method, where any other effects like cell uptake, efflux, or metabolism can be excluded. If data reported from cell-based assays are to be used, it needs to be considered that such activities may be based on several mechanisms including direct

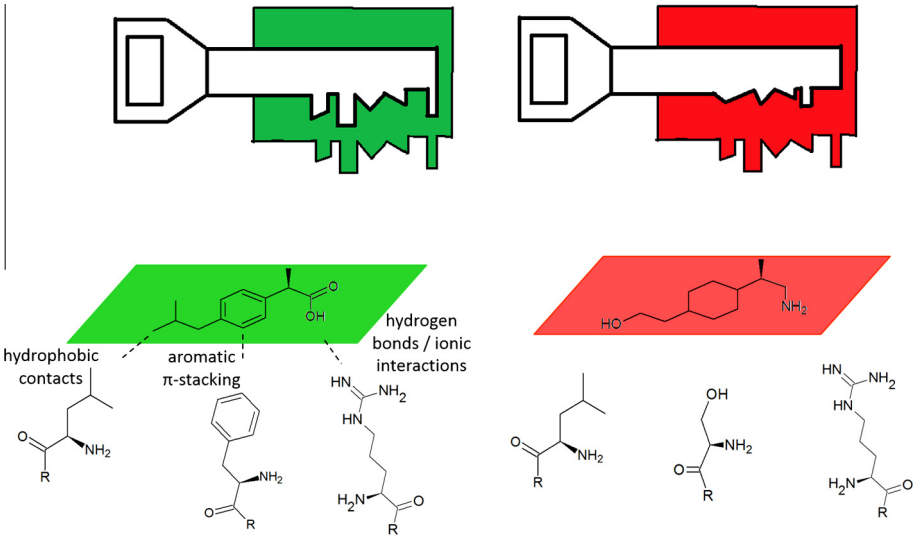

Fig. 1. Illustration of Emil Fischers lock- and key theory and the same principle translated to ligand–protein binding. For enabling interactions, the interaction partners' complementary chemical functionalities need to be positioned in the right distances and angles towards each other: for example, a hydrogen bond acceptor needs a donor on the other side, hydrophobic areas require hydrophobic surroundings, and positive charges can interact with negative charges (green key). In comparison, a hydrogen bond donor can't interact with a hydrophobic moiety and two positively ionizable groups also don't form energetically favorable contacts (red key).