International Journal of Scientific Research in Chemistry (IJSRCH)
© 2020 IJSRCH | Volume 5 | Issue 5 | ISSN : 2456-8457

Pharmacophore Mapping and Virtual Screening

Akshay R.Yadav*, Dr. Shrinivas K. Mohite

Department of Pharmaceutical Chemistry, Rajarambapu College of Pharmacy, Kasegaon, Dist- Sangli, Maharashtra, India

*Corresponding author E-mail:- akshayyadav24197@gmail.com

Article Info
Volume 5, Issue 5
Page Number: 77-82
Publication Issue :
September-October-2020

Article History
Accepted : 18 Oct 2020
Published : 30 Oct 2020

ABSTRACT

3D pharmacophore-based techniques have become one of the most important approaches for the fast and accurate virtual screening of databases with millions of compounds. The success of 3D pharmacophores is largely based on their intuitive interpretation and creation, but the virtual screening with such three-dimensional geometric models still poses a considerable algorithmic and conceptual challenge. Most current implementations favor fast screening speed at the detriment of accuracy. This review describes the general strategies and algorithms employed for 3D pharmacophore searching by some current pharmacophore modeling platforms and will highlight their differences. Developing new medical drugs is expensive. Among the first steps is a screening process, in which molecules in existing chemical libraries are tested for activity against a given target. This requires a lot of resources and manpower. Therefore it has become common to perform a virtual screening, where computers are used for predicting the activity of very large libraries of molecules, to identify the most promising leads for further laboratory experiments. Since computer simulations generally require fewer resources than physical experimentation this can lower the cost of medical and biological research significantly. In this paper we review practically fast algorithms for screening databases of molecules in order to find molecules that are sufficiently similar to a query molecule.

Keywords: 3D pharmacophore, Three-dimensional geometric models, Pharmacophore modeling, Virtual screening.

I. INTRODUCTION

When the 3D structure of the protein target has not been characterized, and/or when a certainnumber of ligands (with or without associated binding affinity) are available, pharmacophore models can be developed and used as search queries for virtual screening of databases[1]. Pharmacophore models may range from sub-structural type pharmacophores to feature- based pharmacophores, in the latter the pharmacophoric points are represented by chemical features like hydrogen - bond acceptors/donors,

Copyright: © the author(s), publisher and licensee Technoscience Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited

hydrophobic points, acidic or basic features etc. Moreover when the necessity occurs to move to the 3D level, virtual screening has to deal with enhanced complexity with regard to functionality and flexibility of molecules, which requires more sophisticated tools for analyzing this type of data[2]. For implementation of this concept into virtual screening, the chemical function based approach is the most generic one. The originality of this type of pharmacophores mostly resides in the fact that their definition is general and represents different types of interactions between organic molecules and proteins. The utility of such models as queries for 3D database search has been recently reviewed. Such pharmacophores can be generated indifferently from ligand sets or from an active site structure. At the end of virtual screening filtering procedure a reliable method for ranking the hits obtained according to their expected bioactivity is required[3].

Methods for Pharmacophore Generation

There are two ways to deduce a pharmacophore: direct- and indirect- methods. The former uses both the ligand and the receptor information, while the latter employs only a collection of ligands that have been experimentally observed to interact with a given receptor. Indirect methods can be used even in the absence of structure of the receptor and hence are more advantageous in the present scenario where the crystal structures of less than 10 % of drug targets are available. However, direct methods are becoming extremely important with the rapidly increasing number of known protein structures, which is the outcome of the Structural Genomics project. Once identified, a pharmacophore model is a versatile tool for the discovery and development of new lead compounds[4].

Steps in Identifying a Pharmacophore

In general, all the algorithms for pharmacophore identification utilize the following six steps:

1) Input
2) Conformational Search
3) Feature extraction
4) Structure Representation
5) Pattern Identification
6) Scoring

Inputs Required for Pharmacophore Identification

Selecting the ligands that will be used in the pharmacophore analysis will have a huge impacton the resulting pharmacophore model. In this context, there are three issues that should beconsidered: the type of the ligand molecules, the size of the data set and its diversity[5].

Ligand Type

A major application of a pharmacophore is in its use as a query (in the preliminary screening layer) for the elimination of in actives, which also implies prioritization of actives. Hence the development of such models often referred to as common feature pharmacophores, requires the input of set of molecules that share the same activity. However in the recent years it is increasingly clear that models that are trained using just the data of actives are incapable of discriminating between actives and inactives. Hence the subsequent attempts to improvise this methodology focused on including the data from inactives as well into the training set. Finally a third type of model, the predictive pharmacophoric model can be developed in cases where a range of activity data exists for the training set molecules[6].

Data Set Size

Most of the currently available methods are designed to handle small data sets, which are composed of less than 100 ligands. This is usually a reasonable limitation especially at the early stages of the project when a large data set of ligands is unavailable.

Data Set Diversity

In order to get an accurate pharmacophore model, the data set should be as diverse as possible. This will allow identifying features that are most critical for the binding. However, it is important that the outliers will not have a high influence on the obtained model. In addition, one should remember that very different

ligands may bind at different binding sites and this may lead to a wrong pharmacophore model[7].

Conformational Search

The pharmacophore identification problem is complicated substantially by the fact that ligands are very flexible molecules. That is, they possess many internal degrees of freedom. The most common one is the rotation of molecular parts around a connecting single bond. As a result, a ligand may have many possible conformations. Each conformation may bind in the active site of the considered receptor. Thus, all the conformations of each input ligand have to be considered during a search for a pharmacophore.

Feature Extraction

In order to perform pharmacophore analyses relevant features in a molecule need to be identified. This can be achieved through the use of predefined atom types with optionally additional centroid "dummy" atoms [Smellie, A., *et al.,* 1995a; Smellie, A., *et al.,* 1995b] or topological substructural definitions at search time or function based pharmacophoric features. There are three main levels of resolution for defining the features:

Atom-Based: One of the simplest ways to define a feature is by the 3D position of an atom, associated with the atom type.

Topology-Based: In some methods the atoms are grouped into topological features like phenyl ring and carbonyl group.

Function-Based: In other methods the atoms are grouped into chemical functional features that describe the kind of interactions important for ligand-receptor binding. The most common functional groups are:

1. Hydrogen bond acceptor, for example carbonyl, aliphatic ether and hydroxyl.
2. Hydrogen bond donor, such as primary/secondary amide, aniline nitrogens and Hydroxyl Base (positively charged at physiological pH 7), for example sp3 N aliphatic amines, hydrazines, guanidines and 2/4 amino pyridines.
3. Acid (negatively charged at physiological pH 7), such as carboxylic acid, acylsulfonamide, unsubstituted tetrazole and on occasion phenols.
4. Aromatic ring, generally (but not always) in the form of ring centroid.
5. Hydrophobic group, for example certain 5/6 membered aromatic rings, isopropyl,butyl and cyclopentyl

The difference between the topological representations to the functional representation is that the resolution of the functional features is lower. For example, a phenyl ring is only one specific type of aromatic ring. Several topological features may have the same chemical function and thus can be classified as the same functional feature. Note that the functional features are not mutually exclusive. For example, hydroxyl oxygen can be classified as both a hydrogen-bond acceptor and donor. In addition, hydrogen-bond acceptor can also be negatively charged[8-9].

Structure Representation

For each ligand structure the selected features are combined to form a representation of thewhole structure.

Pattern Identification

The various stages involved in identification of common features of pharmacophore are:

a. The constructive stage identifies pharmacophore candidates that are common among the most active set of ligands. This is done in the following way: First the set of maximumeight most active compounds is determined. Then, all pharmacophore candidates consisting of up to five features between the two most active ligands are identified by a pruned exhaustive search on all their conformations. Finally, only pharmacophore candidates which fit a minimum subset of features of the remaining most active compounds are retained. The resulting pharmacophore candidates are influenced by the diversity of the data set. For example, the diversity of the two most active ligands influences the number of enumerated pharmacophore candidates[10].

b. The subtractive stage removes those pharmacophore candidates constructed in the previous stage that are also present in more than half of the least active ligands.

c. The optimization stage attempts to improve the score of the pharmacophore candidates that pass the subtractive stage by simulated annealing.

Scoring

In this stage, the pharmacophore candidates were scored and ranked, which are obtained bythe previous stages. The basic requirement from a scoring scheme is that the higher thescoring, the less likely it is that the ligands satisfy the pharmacophore model by a chancecorrelation. The size of the pharmacophore candidates can sometimes be misleading as ascore. For instance, a charged center is rarer than a hydrophobic one. In HipHop the scoring scheme can account for partial fits. Specifically, pharmacophore candidates are ranked based on the portion of input ligands that fit the proposed pharmacophore model. Furthermore, the scoring function takes into account the infrequentand exceptional of the features. For example, a negative charge center is an infrequent and exceptional feature and, therefore, has the largest weight[11].

Methods for Similarity-based Virtual Screening

One way to combat a disease is to find a ligand that will dock with a protein important for that disease, and disrupt its normal function. In general one will have a chemical library of molecules that are available for manufacturing. Using computers for predicting the activity of very large libraries of molecules to identify the most promising leads for further laboratory experiments is called virtual screening . Simulating the docking between the protein and each ligand on a computer in order search for promising ligands in a library of available molecules requires a lot of computing time and available protein structures. Instead one may rely on the idea that similar structure leads to similar properties, and predict the properties of a molecule by studying the properties of similar molecules. Hence, if one has identified a ligand that binds to a given target, for example from another medical drug, or observed in nature, one may find other candidate ligands by looking for ligands in a chemical library or database that are similar to the known binder. This similarity- and ligand-based approach to virtual screening works well for the right formalizations of how to represent molecules and quantify their similarity. Due to the size of chemical databases such as PubChem and ChemDB, the similarity-baed approach to virtual screening also needs efficient methods for screening a database of molecular representations for molecules that are sufficiently similar to a query molecule. In this paper we review such screening methods for molecules represented as fingerprints or SMILES strings[12].

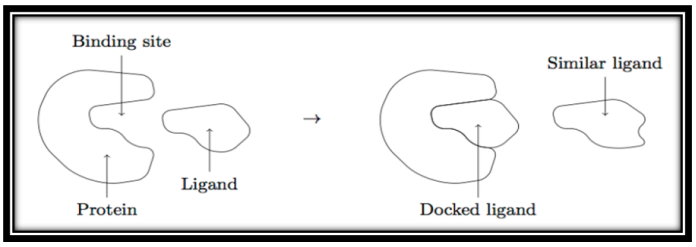

Figure 1. A ligand docking to a protein. Another ligand may dock with the same protein, if it is sufficiently similar.

Similarity between molecules

There are of course several ways to quantify the similarity between two sets (or multi-sets) of features, but the Tanimoto coefficient has proven very useful. If A and B are sets, or multi-sets, of features, then the Tanimoto coefficient, ST (A, B), is:

$$S_T(A,B) = \frac{|A \cap B|}{|A \cup B|}.$$

If A and B are given as two bit-strings, then the Tanimoto coefficient becomes:

$$S_T(A,B) = \frac{|A \wedge B|}{|A \vee B|},$$

where and are bitwise logical 'and' and logical 'or' respectively, and |A| is the number of bits set to one in the bit-string A. See figure 4 for an example.

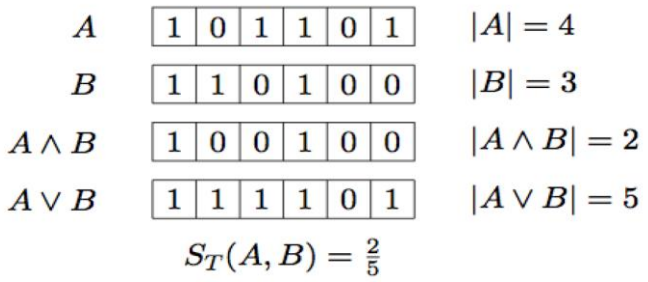

Figure 2. The notation used for bit-strings.

The Tanimoto coefficient as defined above quantifies the similarity between two bit-strings as a number in the interval [0;1], where 0 says that the two bit-strings have no one-bits in common, and 1 says that the two bit-strings are equal[13]. The coefficient is only defined if there is at least one bit set to one in the two bit-strings (i.e. one feature is shared), which is a very reasonable assumption for molecular fingerprints[14-15]. Recall, that the LINGO profile of a molecule is the multi-set of LINGOs in its SMILES string. The similarity between two ligands can thus be measured as the Tanimoto coefficient between their LINGO profiles. This measure is called the LINGOsim between the ligands[16]. One of the major motivations for quantifying molecular similarity is to identify molecules for medical drugs. The problem can be formalized as: We are given a database of representations (for example fingerprints or SMILES) of synthesizable molecules, a query molecule A, and a minimal similarity SMIN. The task is then to find all molecules B in the database where ST (A, B) ≥ SMIN. This query can of course be performed by a naive screening of the database, where we examine every fingerprint A in the database to compute ST (A, B)[17-18]. However, due to the typical size of the database, this is not a desirable approach. In the following sections, we review how to perform such queries more efficiently in practice. We first consider the problem for molecules represented as bit-strings (fingerprints), and secondly, for molecules represented as SMILES[19-20].

II. CONCLUSION

In rational drug design process, it is common that the biological activity data of a set of compounds acting on a particular protein is known while information of the three dimensional structure of the protein active site is absent. A three-dimensional pharmacophore hypothesis that is consistent with existing molecules should be useful and predictive in evaluating new compounds and directing further synthesis. The pharmacophore modeling of the synthesized molecules shows how a set of active molecules can uncover the molecular characteristics or features essential for activity. By reviewing reviewed computationally efficient methods for solving the problem of identifying all molecules stored in database that have a certain similarity to a query molecule. We have considered to problem when molecules were represented by bit-strings, and when molecules were represented by SMILES string. In both cases, the similarity measure used has been the Tanimoto coefficient. The growing size of chemical databases implies a growing need for solutions to this problem that are efficient in practice. An area for improvement that we have not considered in details is memory usage. Our data structures consume a lot of memory. To store very large molecule databases it might be relevant to create an I/O efficient implementation that stores the data structures on disk in way that can be processed efficiently without reading the entire structure into memory.

III. REFERENCES

[1]. Brown, R.D. and Martin, Y.C. (1997) The information content of 2D an 3D structural descriptors relevant to ligand receptor binding. J. Chem. Inf. Comput. Sci. 37, 1–9.

[2]. Ewing, T. et al. (2006) Novel 2D fingerprints for ligand-based virtual screening. J. Chem. Inf. Model. 46, 2423–2431.

[3]. Hert, J. et al. (2004) Comparison of fingerprint-based methods for virtual screening using multiple bioactive reference structures. J. Chem. Inf. Comput. Sci. 44, 1177–1185

[4]. Willet, P. (2005) Searching techniques for databases of two- and threedimensional chemical structures. J. Med. Chem. 48, 4183–4199

[5]. Humblet, C. and Dunbar, J.B., Jr (1993) Chapter VI. Topics in drug design and discovery. In Annual Reports in Medicinal Chemistry, (Vol. 28) (Venuti, M.C., ed.), pp. 275–284.

[6]. Evers, A. et al. (2005) Virtual screening of biogenic amine-binding Gproteincoupled receptors: comparative evaluation of protein- and ligandbased virtual screening protocols. J. Med. Chem. 48, 5448–5465.

[7]. 38 Hurst, T. (1994) Flexible 3D searching: the directed tweak technique. J. Chem. Inf. Comput. Sci. 34, 190–196.

[8]. Wolber, G. et al. (2006) Efficient overlay of small organic molecules using 3D pharmacophores. J. Comput.- Aided Mol. Des. 20 (12), 773–788.

[9]. Sheridan, R.P. and Kearsley, S.K. (2002) Whydo we need so many chemical similarity search methods? Drug Discov. Today 7, 903–911.

[10]. Johnson, M.A. and Maggiora, G.M. (1990) Concepts and Applications of Molecular Similarity. John Wiley & Sons, Inc. Leach, A. (2001) Molecular Modelling: Principles and Applications (2ndedition), Prentice Hall 43 Zhu, F. and Agrafiotis, D.K. (2007) Recursive distance partitioningalgorithm for common pharmacophore identification. J. Chem. Inf. Model. 47, 1619–1625.

[11]. Brint, A.T. and Willet, P. (1987) Algorithms for the identification of threedimensional maximal common substructures. J. Chem. Inf. Comput. Sci. 27, 152–158.

[12]. Mason, J.S. et al. (2001) 3-D pharmacophores in drug discovery. Curr Pharm. Des. 7, 567–597

[13]. Bohm, H-J. et al. (1996) Wirkstoffdesign. Spektrum Akademischer Verlag Wermuth, C.G. et al. (1998) Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). Pure Appl. Chem. 70, 1129–1143.

[14]. Burger, A. (1991) Isosterism and bioisosterism in drug design. Fortschr. Arzneimittelforsch. 37, 287–371.

[15]. Hansch, C. (1974) Bioisosterism. Intra-Science Chem. Rept. 8, 17–25

[16]. Lipinski, C.A. (1986) Bioisosterism in drug design. Ann. Rep. Med. Chem.21, 283–291.

[17]. Thornber, C.W. (1979) Isosterism and molecular modification in drug design. Chem. Soc. Rev. 8, 563–580

[18]. Catalyst, 4.11, Accelrys Inc., http://accelrys.com

[19]. Greene, J. et al. (1994) Chemical function queries for 3D database search. J. Chem. Inf. Comput. Sci. 34, 1297–1308.

[20]. Kurogi, Y. and Guner, O.F. (2001) Pharmacophore modeling and three dimensional database searching for drug design using catalyst. Curr. Med. Chem. 8, 1035–1055.

Cite this article as :

Akshay R.Yadav, Dr. Shrinivas K. Mohite, "Pharmacophore Mapping and Virtual Screening", International Journal of Scientific Research in Chemistry (IJSRCH), ISSN : 2456-8457, Volume 5 Issue 5, pp. 77-82, September-October 2020. URL : http://ijsrch.com/IJSRCH205617