MMSL Military Medical Science Letters

Mil. Med. Sci. Lett. (Voj. Zdrav. Listy) 2016, vol. 85(2), p. 75-79
ISSN 0372-7025
DOI: 10.31482/mmsl.2016.014

REVIEW ARTICLE

VIRTUAL SCREENING IN DRUG DESIGN – OVERVIEW OF MOST FREQUENT TECHNIQUES

Tomas Kucera

Department of Toxicology and Military Pharmacy, Faculty of Military Health Sciences, University of Defence, Trebesska 1575, 500 01 Hradec Kralove, Czech Republic

Received 22nd May 2016.
Revised 5th June 2016.
Published 14th June 2016.

Summary

New and modern techniques of drug design are extensively used in parallel or instead of the classic ones. Applicability of virtual screening (VS) is growing with the computational performance. This article includes list and short description of most frequent used methods in VS. These methods are divided into two groups – ligand-based VS and structure-based VS. Ligand based methods include chemical similarity, pharmacophore and quantitative structure-activity relationship. Molecular docking and scoring are methods of the structure-based VS.

Key words: virtual screening, drug design, docking, scoring, pharmacophore, fingerprint, QSAR

INTRODUCTION

Drug design is a costly and time-consuming process. Some new methods are carried out in parallel or instead of the traditional ones in the process of searching for new drugs. This new performance techniques are called screening methods. We can enumerate e.g. high throughput screening, biophysics supported methods, crystallography, virtual screening and others. The virtual screening (VS) is part of the computer-aided drug design [1].

The VS became popular in 1980s and their popularity has increased with the development of computational technologies and growing computational performance [2]. The VS has three main advantages – performance, price and safety – compared with traditional methods or with *in vitro* high through-put screening. It enables to test large databases of compounds without the necessity to hold them. A lot of compounds predicted to be inactive can be eliminated before *in vitro* testing and the compounds with predicted activity can be prioritized [3]. This pre-selection is saving time and money.

There are two main strategies in VS – ligand-based methods and structure-based methods. In between, there are so called hybrid methods.

✉ University of Defence, Faculty of Military Health Sciences, Department of Toxicology and Military Pharmacy, Trebesská 1575, 500 01 Hradec Kralove, Czech Republic
tomas.kucera2@unob.cz
☎ + 420 720 395 913

STRUCTURE-BASED VIRTUAL SCREENING

Knowledge of target 3D-structure is a big advantage in VS. It is possible to use methods of molecular docking and scoring with this knowledge for VS.

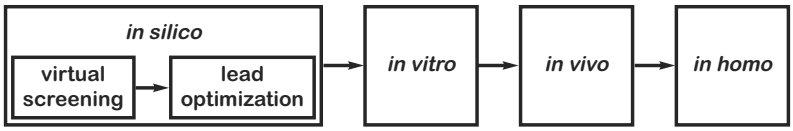

Figure 1. Position of virtual screening in current drug design

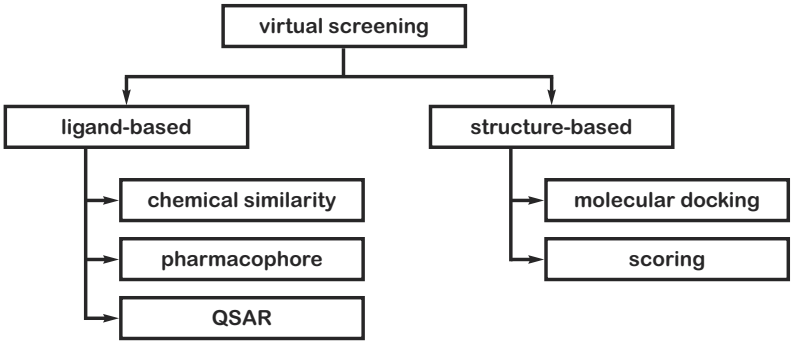

Figure 2. Overview of most frequent virtual screening methods

Molecular docking

Molecular docking helps to predict ligand-receptor interaction. Within virtual screening use, the inter-action of small molecule with a protein is most frequent. This method studies the binding mode and the position of the ligand in a protein's binding site [4].

There are two theories important to understand this technique. The key-lock theory compares the receptor to a lock and the ligand to a key. The ligand has to have accurate size and shape to be able to fit the receptor. This theory is important mainly for docking with inflexible receptor.

The induced-fit theory rejects the conception of rigid receptor. It supposes that the final protein-ligand structure is a result of their mutual interaction. This theory led to development of docking methods based on receptor flexibility [5].

Molecular docking has two main phases. The first is docking in the narrow sense or it could be called "sampling". It looks for the best pose of ligand in the receptor binding site. The second one is scoring and it evaluates the ligand pose through prediction and quantification of the binding affinity.

Here is short overview of sampling algorithms is based on review published by Sousa et al. (2013) [6].

Geometry-based methods divide the binding cavity and also the ligand to set of spheres. Then the software combines position of ligand-spheres and cavity-spheres [5,7].

Hash functions transform information about ligand and receptor to a hash key. In the second, recognition step, hash keys are matched and best combinations are evaluated [8,9].

Incremental construction methods are first algorithms that respect ligand flexibility. The software divides the ligand along each rotatable bond. Then the first fragment is placed and its best pose is the base for addition of next parts and "grow" of the ligand [5].

Genetic method uses principles of evolutionary biology to locking for the best solution. It starts with zero generation of randomly created solutions. The next generations are created as combination of best-evaluated individuals from parent generation with some impact of random mutations (analogous to evolution of biological species) [5].

Simulated annealing is a part of molecular dynamic methods. The system is cooled during the simulation run, the energy is decreased and system is approaching the local minimum of energy. Using *Monte Carlo* algorithm particularly solves the dependence of the result on a starting position. It enables to overcome energetic barriers and to find global energetic minimum [8,10].

Scoring

Scoring is sometimes presented as a separate method of structure-based drug design but it is irreplaceable part of molecular docking. In the docking approaches, it is the second phase and it follows the sampling.

There are three basic scoring methodologies based on different theories.

The *force field-based scoring* uses classical molecular-mechanistic calculations. All these scoring functions have similar formulas based on sum of partial energies (bond interactions, van der Waals interactions, electrostatics interactions, angle bending, out-of-plane bending, torsion interactions and others) [11,12].

The basis for calculation is often Lennard-Jones potential:

$$E = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left(\frac{A_{ij}}{r^{12}} - \frac{B_{ij}}{r^{6}} + 332 \frac{q_i q_j}{D r_{ij}} \right),$$

where A_{ij} and B_{ij} are van der Waals repulsion and attraction parameters, r_{ij} is the distance between atoms i and j, q_i and q_j are the point charges on atoms i and j, D is the dielectric function.

This type of scoring function has many advantages of e.g. speed of computation, strong physical basis and good theoretical description. The main disadvantage of this method is exclusion of the entropy parameters (such as desolvatation energy and restriction of ligand flexibility) [5].

Empirical scoring functions calculate the binding energy as a sum of particular increments (e.g. ionic interactions, hydrogen bonds, π-interactions, desolvatation, lipophilic interactions). Proportional weight of these increments is defined by their coefficients. Values of the coefficients are determined on training set of protein-ligand complexes using multiple linear regression [13].

As an example, London dG scoring function:

$$\Delta G = c + E_{flex} + \sum_{h-bonds} c_{hb} f_{hb} + \sum_{metal-lig} c_m f_m + \sum_{atomsi} \Delta D_i,$$

where E_{flex} is a topological estimate of ligand entropy, f_{hb} and f_m are measures of geometric imperfections of protein-ligand and metal-ligand interactions, ΔDi is the desolvatation energy and c, c_{hb} and c_m are empirical determined coefficients.

The main advantage of this method is inclusion of enthalpy and entropy contributions. A disadvantage is dependence on used training set.

The third possibility is using of *knowledge-based scoring functions*. They are based on statistical investigation of protein-ligand complexes. The theory assumes that frequency of distance of two specific atom types is proportional to its favour. Partial increment of an atom-type pair can be calculated as:

$$A(r) = -k_B T \ln g_{ij}(r),$$

where k_B is the Boltzmann constant, T is the thermodynamic temperature, g_{ij} is the distribution function for atom-type pair ij and r is the distance between atoms i and j [14].

In advantage, this method is relatively fast. A disadvantage is poor physical theoretical-physical basis.

Each scoring function has some limitations and this is the reason for many false high-scored results. A possible solution of this problem is *consensual scoring*. It lies in re-scoring the result with usually 2–3 scoring functions from different previously described groups. Selection of more-times high-scored molecules decreases the false positive rate [15,16,17].

LIGAND-BASED VIRTUAL SCREENING

Knowledge of target 3D-structure is an advantage in the drug design process. Across fast development in this area, there are many proteins without described spatial information. In these cases, it is possible to use target-independent approaches.

Chemical similarity

Usage of methods of chemical similarity is based on the theory of "neighborhood behavior". It assumes then similar molecules have similar biological effect [18]. Chemical fingerprint is a way to compare resemblance of the chemical structures. Fingerprint uses descriptors that convert chemical information into a binary string. Thus it is possible to compare two bit strings using mathematical methods [19,20].

Main advantage of this method is a little input data demand. It is necessary to know only one or several active ligands. Speed and very low computational demands are further arguments supporting use of this method. Molecular similarity is possible to use favorably as a pre-filtration of ligand database before application of a computationally more expensive method [21,22].

Pharmacophore

The idea of "neighborhood behavior" is adopted and improved by so called pharmacophore method. Chemical fingerprint states similarity (or dissimilarity) of two molecules without their detailed investigation. Pharmacophore is presented as a description of specific features that causes a biological activity (or that are necessary for the activity). Description of these features does not include real functional groups, but it describes abstract electrostatic and steric characteristics [23].

There are two main ways to define a pharmacophore. The first is based on knowledge of ligand-receptor complex (usually from X-ray crystallography) and it is structure-based pharmacophore. Second possi-bility is definition of features based on structural alignment of the known ligands – so called ligand-based pharmacophore.

Defined pharmacophore is possible to use for screening of large databases of compounds. Software compares compounds with the suitable features and selects molecules that match the pharmacophore.

This method is relatively computationally undemanding but it place demand to researcher skill to correctly define the pharmacophore [24].

Quantitative structure-activity relationship

An advanced method of computer-aided drug design is QSAR (quantitative structure-activity relationship). QSAR method adds the quantification of the activity to the calculation. It tries to find correlation between molecule structure and activity [25,26].

It is relatively easy to quantify biological activity with parameters of e.g. K_i, IC_{50}, EC_{50}. Quantification of the chemical structure is the real challenge. In the beginning, there were described relations between activity and physical-chemical properties (e.g. molecular weight, lipophilicity, acidity). It was so called 1D-QSAR. In the course of time, many descriptors were added and the predictions became more accurate. The 2D-QSAR takes into account structural patterns and connectivity.

The gold standard in this development became the 3D-QSAR that describes electrostatic and steric properties of molecules in the three-dimensional space. The 4D-QSAR computes with more conformations of the ligand, the 5D-QSAR adds induced-fit models to the 4D-QSAR and the 6D-QSAR adds solation models. [25,27].

Besides the quantification of chemical structures, there is further challenge in statistical processing of large data sets. For such processing, many methods were developed. They are mostly based on multilinear regression that is usually used after data set simplification by the transformation algorithms [28].

CONCLUSIONS

Herein, the most common methods of virtual screening were summarized and described. The choice of a method is highly case dependent and there is no simple manual describing which method is the right one. Each method has its advantages or disadvantages. Combination of two or more computational methods is usually a favorable tool for drug design purposes.

It seems that it is possible to use VS methods in design of new drugs for military purposes including e.g. acetylcholinesterase or butyrylcholinesterase reactivators for antidotal treatment. In this case, the use of VS could help to develop molecules tailor-made for exact enzyme structure. As such, a designed highly active reactivator could be applied for pseudo-catalytic scavenging or treat-ment of intoxications caused by organophosphorus nerve agents.

ACKNOWLEDGEMENT

The work was supported by the grant SV/FVZ201601.

REFERENCES

1. Klebe, G. Drug design. Springer, New York 2013. 850 p.

2. Stahura, F.; Bajorath, J. New Methodologies for Ligand-Based Virtual Screening. *Curr Pharm Des*. **2005**, 11, 1189–1202.
3. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Computational Methods in Drug Discovery. *Pharmacol Rev*. **2013**, 66, 334–395.
4. Kuchař, M. Výzkum a vývoj léčiv: studijní program: syntéza a výroba léčiv. VŠCHT, Praha **2008**. 166 p.
5. Höltje H. D. Molecular modeling: basic principles and applications. 2nd ed. Wiley-VCH, Weinheim **2003**. 228 p.
6. Sousa, S. F.; Ribeiro, A. J. M.; Coimbra, J. T. S.; Neves, R. P. P.; Martins, S. A.; Moorthy, N. S. H. N. et al. Protein-Ligand Docking in the New Millennium – A Retrospective of 10 Years in the Field. *Curr Med Chem*. **2013**, 20, 2296–2314.
7. Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N. et al. Development and validation of a modular, extensible docking program: DOCK 5. *J Comput Aided Mol Des*. **2006**, 20, 601–619.
8. Lipkowitz, K. B.; Boyd, D. B. Reviews in computational chemistry. [online]. Wiley-vch, New York, 2001 [cit. 2016-05-12]. Available from: http://public.eblib.com/choice/publicfullrecord.aspx?p=468828
9. Fischer, D. A geometry-based suite of molecular docking processes. *J Mol Biol*. **1995**, 248, 459–477.
10. Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins Struct Funct Genet*. **1990**, 8, 195–202.
11. Grinter, S.; Zou, X. Challenges, Applications, and Recent Advances of Protein-Ligand Docking in Structure-Based Drug Design. *Molecules*. **2014**, 19, 10150–10176.
12. Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem*. **1996**, 17, 490–519.
13. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions. 1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des*. **1997**, 11, 425–445.
14. Muegge, I.; Martin, Y. C. A General and Fast Scoring Function for Protein−Ligand Interactions: A Simplified Potential Approach. *J Med Chem*. **1999**, 42, 791–804.
15. Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *J Med Chem*. **1999**, 42, 5100–5109.
16. Cerqueira, N. M. F. S. A.; Gesto, D.; Oliveira, E. F.; Santos-Martins, D.; Brás, N. F.; Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Receptor-based virtual screening protocol for drug discovery. *Archives of Biochemistry and Biophysics*. **2015**, 582, 56–67.
17. Ferreira, L. G.; dos Santos, R. N.; Oliva, G.; Andricopulo A. D. Molecular Docking and Structure-Based Drug Design Strategies. *Molecules*. **2015**, 20, 13384–13421.
18. Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. Neighborhood Behavior: A Useful Concept for Validation of "Molecular Diversity" Descriptors. *J Med Chem*. **1996**, 39, 3049–3059.
19. Sheridan, R. Why do we need so many chemical similarity search methods? *Drug Discov Today*. **2002**, 7, 903–911.
20. Xue, L.; Godden, J. W.; Bajorath, J. Database Searching for Compounds with Similar Biological Activity Using Short Binary Bit String Representations of Molecules. *J Chem Inf Comput Sci*. **1999**, 39, 881–886.
21. Willett, P. Similarity-based virtual screening using 2D fingerprints. *Drug Discov Today*. **2006**, 11, 1046–1053.
22. Vogt, M.; Bajorath, J. Chemoinformatics: A view of the field and current trends in method development. *Bioorg Med Chem*. **2012**, 20, 5317–5323.
23. Wermuth, C. G.; Ganellin, C. R.; Lindberg, P.; Mitscher, L. A. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). Pure Appl Chem [online]. 1998, 70. [cit 2016-05-13]. Available from: http://www.degruyter.com/view/j/pac.1998.70.issue-5/pac199870051129/pac199870051129.xml
24. Hung, C. L., Chen, C. C. Computational Approaches for Drug Discovery: Computational Approaches for Drug Discovery. *Drug Dev Res*. **2014**, 75, 412–418.
25. Verma, J.; Khedkar, V.; Coutinho, E. 3D-QSAR in Drug Design - A Review. *Curr Top Med Chem*. **2010**, 10, 95–115.
26. Kubinyi, H. QSAR and 3D QSAR in drug design Part 1: methodology. *Drug Discov Today*. **1997**, 2, 457–467.
27. Damale, M.; Harke, S.; Kalam, K. F.; Shinde, D.; Sangshetti, J. Recent Advances in Multidimensional QSAR (4D-6D): A Critical Review. *Mini-Rev Med Chem*. **2014**, 14, 35–55.
28. Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol Inform*. **2010**, 29, 476–488.