



IN SILICO APPROACHES FOR DRUG DISCOVERY FOCUS ON VIRTUAL SCREENINGS

Simona Distinto*,
Elias Maccioni

Department of Life and
Environmental Sciences,
University of Cagliari,
Cagliari, Italy

Abstract:

The field of medicinal chemistry has become increasingly dynamic and medicinal chemists face the challenge of rapidly evolving new technologies. In the last decade, medicinal chemistry methodologies have been largely replaced from an individual scheme to an interdisciplinary approach. Furthermore, the shift from traditional to Omics-based applications is needed to develop computational, chemo, and bioinformatic tools that could help medicinal chemists to analyse, link, and compare the research results. Hence, drug research has necessarily oriented drug discovery toward more rational strategies. In silico Virtual Screening (VS) is one of the most promising approaches to accelerate the drug development process. Efficient analysis of key compounds and target properties is crucial for carrying out a virtual screening process. At the same time, it can reduce the attrition rates in drug development. Of course, the main purpose of VS is to identify novel chemical scaffolds as hits for further optimization using medicinal chemistry approaches. An overview of the most employed methods for VS, challenges, and new directions will be discussed.

Keywords:

Drug Discovery, Virtual Screenings, Ligand-based, Structure-based.

INTRODUCTION

Pharmaceutical research is moving towards a more interdisciplinary endeavour. Indeed, drug development requires information on a broad range of topics: not only chemical structures and reactions but also target structure, biological pathways, drug-target interactions. Hence, the collaboration between the researchers working in different disciplines is essential to maximize the potential benefits of drug discovery (DD).

DD involves various steps: a discovery phase, which includes target discovery (only 10% of the human genome is druggable) and the identification of active molecules or hits. This phase ends with the identification of lead molecules. The second phase consists in the lead optimization. This phase focuses on optimizing lead activity and ADMET (absorption, distribution, metabolism, excretion, toxicity) properties.

Correspondence:

Simona Distinto

e-mail:

s.distinto@unica.it



The third phase includes the product development, with the preclinical and clinical trials (I, II, and III), and, finally, the registration phase, which will enable distribution on the market and the clinical use of drugs.[1]

The estimation of the median cost of efficacy trials for new drugs approved by the FDA amounted to \$19 million.[2]

Using computational approaches, many of the steps involved in the drug discovery projects can be more efficient and rewarding. Nowadays, computational tools are useful not only for hit/lead identification and optimization, but also for target identification, prediction of druggable pockets, and accurate prediction of ADME-Tox-related properties and metabolism.[3-5]

Several marketed drugs such as oxymorphone, saquinavir, imatinib, zanamivir, dorzolamide norfloxacin, and several clinical candidates, have been discovered or optimized with the aid of molecular modelling techniques.[6]

One of the most promising techniques to accelerate the DD process is to perform *in silico* virtual screening (VS).[7] VS hit identification rates vary from about 40% to 1%,[8] with most active hit compound activities ranging from μM to nM.

However, hit rates can provide a loose measure of the success of the VS method since the main aim is not necessarily maximizing hit rates and reaching nM activity, but identifying novel active compounds that contain new scaffolds, are synthesizable, and provide a basis for hit-to-lead optimisation.

There is a continuous flow of publications reporting ligand-based (LB) and structure-based (SB) VS applications where new active compounds have been identified employing a variety of VS methods.[9]

The huge computational demand of such VS applications requires developing parallel algorithms and exploiting the computational power of large high-performance computing (HPC) systems to accomplish such screenings, within an affordable time. In fact, DD can be significantly boosted using big data resources.

In this framework, VS approaches, their applications, new trends, and challenges need to be examined and discussed.

2. OVERVIEW OF CURRENT APPROACHES USED IN VS

VS methodologies can be broadly grouped into two main categories: LBVS and SBVS. The former requires a set of known active ligands for structure similarity search, while the latter requires the 3D structure of the target.

When both ligand and structural information exists, it is possible to combine methods to yield improved results. In general, the combination of more approaches, in a hierarchical or parallel way, can lead to an increase in both scaffold diversity of the retrieved hits and hits rate.[7, 10]

VS methods are typically validated by retrospective analysis on benchmark datasets considering the diversity of targets, the diversity of ligands, and the selection of appropriate decoys. This led to high-quality and reliable benchmarking datasets, proving their strength.[11]

Benchmark datasets consist of series of active and inactive molecules, each associated with a specific target. Often the active compounds are experimentally validated, instead, the documentation of experimentally validated inactive molecules is scarce. For this reason, assumed inactive molecules (decoys) are frequently employed.

When new targets are studied consensus approach and studies of target similarities can help to maximize the VS success.

Similarity LB methods are based on the calculation of molecular descriptors, which consider molecular properties of different complexity 1D-, 2D- or 3D.[12, 13] In particular, these methods estimate similarity metrics considering coefficients such as Tanimoto index, Dice coefficient (Hodgkin index), Cosine coefficient or distances, Soergel distance, Euclidean distance, Hamming (Manhattan or city-block) distance.[14] Furthermore, also quantitative structure-activity relationship (QSAR) methods are largely applied for LBVS. QSAR model development and application include preparation of data, analysis of data, model development, model validation, and VS of chemical databases.[15]

SBDD and SBVS have contributed to the introduction of ~50 new compounds into clinical trials and numerous drug approvals.[16] To improve the predictive power of docking experiments, it is necessary to have well-established protocols and robust metrics to measure it. Docking approach benchmarking for VS application includes two properties:



- a. RMSD computed for the predicted binding pose against the experimental pose;
- b. Binding free energies/docking energies which are proportional to experimental inhibition/dissociation constants.

If the binding mode prediction is recognised to be often accurate ($\text{RMSD} < 2 \text{ \AA}$), the second is not equally as precisely determined. In fact, docking ends often with several false-positive hits because of a lack of exactness in predicting binding affinity.

Many factors can influence the performance of docking experiments. Some limitations can be overcome considering alternative protonation states for critical residues, the flexible side-chain minimization, consensus scoring, rescoring of the docking complexes with different scoring functions, or the inclusion of solvation effects through specific protocols.[17]

No single docking software is superior to the others in all respects.[18] The validation process can help to choose the most appropriate program and protocol for a specific target. However, in a new target VS context, consensus docking can improve the reliability of docking by using different docking programs.[19] Pre-processing and curation of data are mandatory to correctly assess the quality of information and avoid any potential bias in VS methods validation.[6]

Aiming at finding new scaffolds, fragment-based VS has emerged to be a powerful approach by exploiting molecular fragments with molecular weight $<150\text{--}250$. Ideally, the different identified small fragments can be subsequently connected by opportune linkers to increase the hits potency.[20]

Pharmacophore modelling (LB or SB) is an important and useful method for drug discovery.[21] Among the 3D methods, no doubt that the pharmacophore approach for VS is the most appreciated by medicinal chemists because of a common language.[13]

SBDD is also important to derive structure-activity relationships of a chemical series, especially in the lead optimization phase, when very accurate modifications are needed to adjust an ADME/tox profile while maintaining binding affinity.

All VS methods have their strength and pitfall, over the time many improvements have been obtained. In the last decades, Artificial Intelligence (AI) methods appeared in the panel of new strategies.

3. AI METHODS IN VS

Computer science advances and speeds find AI broadly benefitting several fields. AI refers to an algorithm capable of mimicking cognitive functions without supervision or user input.

In pharmaceutical research, because of the complex nature of big data, some relations may not be apparent at first glance and might lead to wrong results or hypotheses.

Among AI methods, machine learning (ML), and in particular deep learning (DL) using deep neural networks (DNN), has been responsible for recent progress. [22, 23] DL involves scaling machine learning using multi-layered neural networks to attempt to model the abstraction of big data.

However, because of the poor interpretability of AI methods, and the limited accuracy of prediction results for molecules with large structural differences from the training compounds, HTVS rather than DL methods remain the dominant choice in drug design.[19]

Of course, their application is growing and many studies showed interesting results,[24] but particular attention should be given to data collection and cure.[25]

Great efforts are being made to produce better algorithms. Nevertheless, software is not (yet) meant to replace chemical intuition or deep knowledge of the biological target, which is essential for the identification of hits.

The combination of ligand- and structure-based methods has become a common approach in virtual screening since it has been hypothesized that their integration can enhance the strengths and reduce the drawbacks of each method.

4. PITFALL AND CHALLENGES

The druglike chemical space is estimated to be around 10^{63} molecules.[26] However, a drug candidate needs to possess the right combination of properties to provide efficacy and safety, and formulation. Only a small fraction of chemical compounds possesses these properties. Furthermore, accessible synthetic pathways of compounds and purchasability are not guaranteed. This would make the use of virtual libraries more appealing.[27]



The sizes of screenings databases range from a few thousand (e.g. DrugBank) to almost one hundred million (e.g. Pubchem). But, are bigger screenings better?[28]

Of course, a large database gives more chance of finding new scaffolds, but the analysis of published works highlighted the interesting conclusion that even less complex studies can lead to success stories.[29] This is encouraging for small labs where, albeit the access to thousands of CPUs/GPUs also via the cloud is feasible, the cost of thousands of licenses for virtual screening software may be unaffordable.

CADD still faces many challenges, which include increasing the efficiency of virtual screening; further developing the computational chemogenomic field; predictive animal models, more attention for earlier toxicology evaluation, data curation, and quality.

Moreover, there is a large room for improvement on studies that involve multiple molecular targets, for both synergy and side effect prediction.[30] In fact, many studies suggest that the partial inhibition of a small number of targets, involved in a specific disease, can be more efficient than the complete inhibition of a single target. In this regard, VS that consider the polypharmacology aspect might solve the problem of fighting challenging diseases by retrieving more efficient weapons.

5. CONCLUSIONS

Researchers are currently attacking diseases of great complexity such as virus infections, cancer, and neurodegenerative disorders. Furthermore, drug targets being tackled include a growing number of less druggable targets than those pursued previously. In addition, the entry bar for new drugs is becoming higher because of the enhanced standard of care. Indeed, the VS approaches can help the DD process. However, the awareness of how a method can fail is as useful as knowing how it works. Hence, it is mandatory to maintain a critical attitude when dealing with results. The VS success rate may vary considerably, depending on the target and the expertise of the user. But, if a "first-in-class hit" or a novel scaffold is identified, it should be considered as a success, regardless of its potency. Therefore, the recommendation is to pursue VS for DD, even on a relatively small scale.

6. REFERENCES

- [1] J. P. Hughes, S. Rees, S. B. Kalindjian, and K. L. Philpott, "Principles of early drug discovery," *British journal of pharmacology*, vol. 162, no. 6, pp. 1239-1249, 2011.
- [2] T. J. Moore, H. Zhang, G. Anderson, and G. C. Alexander, "Estimated Costs of Pivotal Trials for Novel Therapeutic Agents Approved by the US Food and Drug Administration, 2015-2016," *JAMA Internal Medicine*, vol. 178, no. 11, pp. 1451-1457, 2018.
- [3] J. Kirchmair *et al.*, "Predicting drug metabolism: experiment and/or computation?," *Nature Reviews Drug Discovery*, vol. 14, no. 6, pp. 387-404, 2015.
- [4] T. Katsila, G. A. Spyroulias, G. P. Patrinos, and M.-T. Matsoukas, "Computational approaches in target identification and drug discovery," *Computational and structural biotechnology journal*, vol. 14, pp. 177-184, 2016.
- [5] S. Alqahtani, "In silico ADME-Tox modeling: progress and prospects," *Expert Opinion on Drug Metabolism & Toxicology*, vol. 13, no. 11, pp. 1147-1158, 2017.
- [6] F. D. Prieto-Martínez, E. López-López, K. Eurídice Juárez-Mercado, and J. L. Medina-Franco, "Chapter 2 - Computational Drug Design Methods—Current and Future Perspectives," in *In Silico Drug Design*, K. Roy, Ed.: Academic Press, 2019, pp. 19-44.
- [7] S. Distinto *et al.*, "Identification of HIV-1 reverse transcriptase dual inhibitors by a combined shape-, 2D-fingerprint- and pharmacophore-based virtual screening approach," *European Journal of Medicinal Chemistry*, vol. 50, pp. 216-29, 2012.
- [8] T. Zhu *et al.*, "Hit identification and optimization in virtual screening: practical recommendations based on a critical literature analysis," *Journal of medicinal chemistry*, vol. 56, no. 17, pp. 6560-6572, 2013.
- [9] D. Stumpfe and J. Bajorath, "Current Trends, Overlooked Issues, and Unmet Challenges in Virtual Screening," *Journal of Chemical Information and Modeling*, vol. 60, no. 9, pp. 4112-4115, 2020.
- [10] S. Alcaro *et al.*, "Identification and characterization of new DNA G-quadruplex binders selected by a combination of ligand and structure-based virtual screening approaches," *Journal of Medicinal Chemistry*, vol. 56, no. 3, pp. 843-55, 2013.
- [11] N. Lagarde, J.-F. Zagury, and M. Montes, "Benchmarking Data Sets for the Evaluation of Virtual Ligand Screening Methods: Review and Perspectives," *Journal of Chemical Information and Modeling*, vol. 55, no. 7, pp. 1297-1307, 2015.



- [12] J. Kirchmair *et al.*, "How To Optimize Shape-Based Virtual Screening: Choosing the Right Query and Including Chemical Information," *Journal of Chemical Information and Modeling*, vol. 49, no. 3, pp. 678-692, 2009.
- [13] A. Cereto-Massagué, M. J. Ojeda, C. Valls, M. Mule-ro, S. Garcia-Vallvé, and G. Pujadas, "Molecular fingerprint similarity search in virtual screening," *Methods*, vol. 71, pp. 58-63, 2015.
- [14] R. A. Miranda-Quintana, D. Bajusz, A. Rácz, and K. Héberger, "Differential Consistency Analysis: Which Similarity Measures can be Applied in Drug Discovery?," *Molecular Informatics*, vol. 40, no. 7, p. 2060017, 2021.
- [15] T. Wang, M.-B. Wu, J.-P. Lin, and L.-R. Yang, "Quantitative structure-activity relationship: promising advances in drug discovery platforms," *Expert Opinion on Drug Discovery*, vol. 10, no. 12, pp. 1283-1300, 2015.
- [16] W. L. Jorgensen, "The Many Roles of Computation in Drug Discovery," *Science*, vol. 303, no. 5665, pp. 1813-1818, 2004.
- [17] M. A. Llanos *et al.*, "Strengths and Weaknesses of Docking Simulations in the SARS-CoV-2 Era: the Main Protease (Mpro) Case Study," *Journal of Chemical Information and Modeling*, vol. 61, no. 8, pp. 3758-3770, 2021.
- [18] Z. Wang *et al.*, "Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power," *Physical Chemistry Chemical Physics*, vol. 18, no. 18, pp. 12964-12975, 2016.
- [19] D. R. Houston and M. D. Walkinshaw, "Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context," *Journal of Chemical Information and Modeling*, vol. 53, no. 2, pp. 384-390, 2013.
- [20] L. R. de Souza Neto *et al.*, "In silico Strategies to Support Fragment-to-Lead Optimization in Drug Discovery," *Frontiers in Chemistry*, Review vol. 8, 2020.
- [21] T. Seidel, S. D. Bryant, G. Ibis, G. Poli, and T. Langer, "3D Pharmacophore Modeling Techniques in Computer Aided Molecular Design Using LigandScout," 2017, pp. 281-312: Wiley.
- [22] J. Bajorath, "Artificial intelligence in interdisciplinary life science and drug discovery research," *Future Science OA*, p. FSO792, 2022.
- [23] T. B. Kimber, Y. Chen, and A. Volkamer, "Deep Learning in Virtual Screening: Recent Applications and Developments," *International Journal of Molecular Sciences*, vol. 22, no. 9, p. 4435, 2021.
- [24] M. Wójcikowski, P. J. Ballester, and P. Siedlecki, "Performance of machine-learning scoring functions in structure-based virtual screening," *Scientific reports*, vol. 7, pp. 46710-46710, 2017.
- [25] J. Bajorath, "State-of-the-art of artificial intelligence in medicinal chemistry," *Future Science OA*, vol. 7, no. 6, p. FSO702, 2021.
- [26] R. S. Bohacek, C. McMartin, and W. C. Guida, "The art and practice of structure-based drug design: A molecular modeling perspective," *Medicinal Research Reviews*, vol. 16, no. 1, pp. 3-50, 1996.
- [27] G. Opassi, A. Gesù, and A. Massarotti, "The hitchhiker's guide to the chemical-biological galaxy," *Drug Discovery Today*, vol. 23, no. 3, pp. 565-574, 2018.
- [28] D. E. Gloriam, "Bigger is better in virtual drug screens," *Nature*, vol. 566, no. 7743, pp. 193-194, 2019.
- [29] D. E. Clark, "Virtual Screening: Is Bigger Always Better? Or Can Small Be Beautiful?," *Journal of Chemical Information and Modeling*, vol. 60, no. 9, pp. 4120-4123, 2020.
- [30] J. L. Medina-Franco, "Grand Challenges of Computer-Aided Drug Design: The Road Ahead," *Frontiers in Drug Discovery, Specialty Grand Challenge* vol. 1, 2021.