TẠP CHÍ KHOA HỌC & CÔNG NGHỆ ĐẠI HỌC DUY TÂN
DTU Journal of Science and Technology 02(45) (2021) 42-50

TRƯỜNG ĐẠI HỌC DUY TÂN
DUYTAN UNIVERSITY

The future of virtual screening in drug discovery

Tương lai của sàng lọc ảo trong khám phá thuốc chữa bệnh

Pham Thi Ly[a], Le Quoc Chon[a,b*]
Phạm Thị Ly[a], Lê Quốc Chơn[a,b*]

[a]*Faculty of Pharmacy, Duy Tan University, Da Nang, 550000, Vietnam.*
[a]*Khoa Dược, Trường Đại học Duy Tân, Đà Nẵng, Việt Nam*
[b]*Institute of Research and Development, Duy Tan University, Da Nang, 550000, Vietnam*
[b]*Viện Nghiên cứu và Phát triển Công nghệ Cao, Trường Đại học Duy Tân, Đà Nẵng, Việt Nam*

(Ngày nhận bài: 26/01/2021, ngày phản biện xong: 06/02/2021, ngày chấp nhận đăng: 03/03/2021)

Abstract

One of the key elements in the early stages of drug discovery is finding good hits to develop lead compounds. Although HTS has been used as a standardized technology for hit finding, it still bears some challenging drawbacks: expensive and low-quality data. Aiming at the same goal as HTS, virtual screening (VS) has been developed to reduce cost and increase efficiency. Recent studies show that VS can deliver numerous quality hits and a few of them even reach clinical trials. This paper uses HTS as a background to discuss the contributions, limitations and research trends in VS field as these two technologies complement each other.

Keywords: high-throughput screening, virtual screening, hit compounds, lead compounds, structure-based drug design, ligand-based drug design.

Tóm tắt

Bước quan trọng đầu tiên của quá trình khám phá thuốc chữa bệnh là tìm ra các hợp chất chất dẫn. Và phương pháp sàng lọc hiệu suất cao (HTS) được xem là công nghệ chuẩn để thực hiện bước này. Tuy nhiên, sử dụng HTS còn gặp nhiều khó khăn do chi phí cao và hiệu suất thấp. Với cùng mục đích sử dụng, sàng lọc ảo (VS) được phát triển để khắc phục các nhược điểm của HTS, để giảm chi phí và tăng hiệu quả tìm kiếm các chất dẫn. Những nghiên cứu mới đây cho thấy VS có thể cung cấp các chất dẫn có chất lượng cao, một số đã và đang được thử nghiệm lâm sàn. Bài báo này thảo luận các đóng góp, hạn chế và xu hướng phát triển của VS trong tương lai.

Từ khóa: sàng lọc hiệu suất cao, sàng lọc ảo, hợp chất dẫn, khám phá thuốc, thiết kế thuốc dựa trên máy tính.

1. Introduction

Drug discovery and development (DDD) is a risky, and expensive process (figure 1). Screening to find hits is conducted after identifying drug targets and this task is routinely carried out by high-throughput screening (HTS), filtering and selecting the most suitable molecules among a library of

* *Corresponding Author:* Le Quoc Chon; Institute for Research and Training in Medicine, Biology and Pharmacy, Duy Tan University, Da Nang; Department of Pharmacy, Duy Tan University, Da Nang, 550000, Vietnam
Email: lequocchon@duytan.edu.vn

small chemicals or biologics [1]. For over two decades, HTS has become a standard technology in pharma and biotech companies [2], [3].

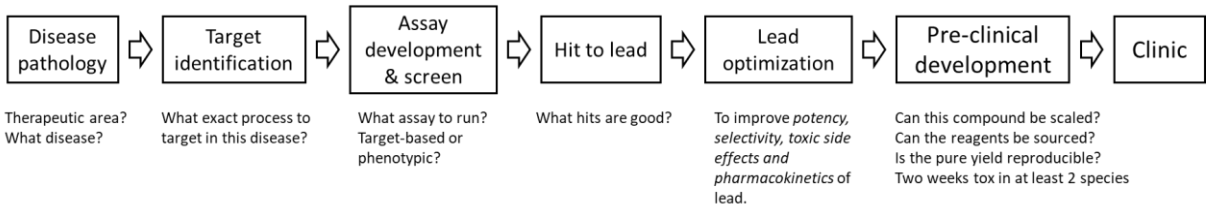

Figure 1: Seven steps in drug discovery

HTS technology has evolved over three generations [4]: the first focused on quantity of compound screened, the second concentrated on efficiency and the third has emphasized on the flexibility and quality of library (figure 2).

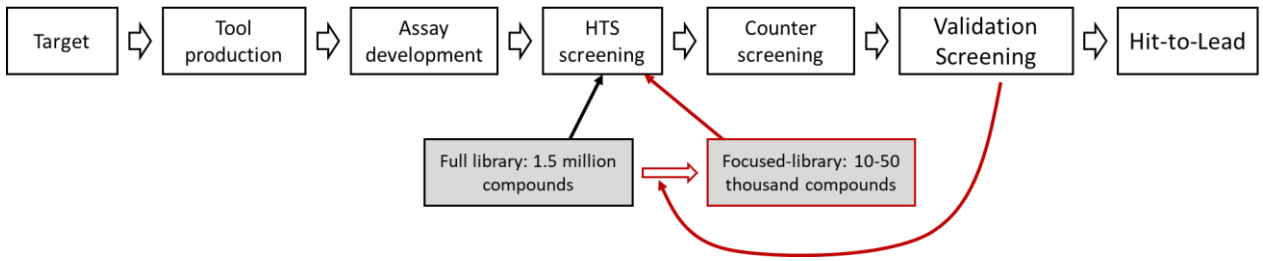

Figure 2: Workflow of contemporary HTS [5].

HTS has been substantially contributing to DDD through providing new chemical entities for lead development [6]–[9]. Leveraging automation and miniaturization technologies, it has accelerated the drug discovery process [10] and generated many FDA-approved drugs [9]. Three critical factors determining the success of HTS campaign are druggable target, compound library and predictive assay [11], [12].

Even though HTS has been widely used, the number of drugs approved by US-FDA is constantly low over two decades [13]. Given that HTS is a routine operation in pharmaceutical [14], it has been partly to blame for the decline in DDD [9], [15].

Indeed, HTS has some inherent drawbacks. First of all, the compound library is too small in comparison to the possible chemical space [16]. Medicinal chemistry assumes that there are sufficient small molecules for all binding sites found in biology [17]. Even though the exact number is unknown, the chemical space may contain about 10^{60} molecules [18]. While a good library must be diverse and lead-like [16], a typical library for HTS holds a few ten thousand to less than two million compounds [2].

The next disadvantage of HTS is high cost. Running HTS is expensive and time-consuming because it is experimental-based. A single HTS screening program costs approximate 75 000 USD [13]. HTS service cost ranges from 0.1 to 1 USD per well [19]. Furthermore, the robotic systems and assay readers in HTS are costly, requiring up to a few millions of USD to set-up and maintain [2], [20]. In addition to resource, conducting HTS campaign is time-consuming, taking several months to a year to finish [21]. Furthermore, HTS program is highly specific, automatizing some expensive steps as assay development and validation is impossible. Moreover, HTS needs real high-quality library, containing many drug-like molecules to screen, and an enormous amount

of time and money is demanded for collecting, synthesizing or/and buying.

In spite of being equipped with advanced technologies, HTS is possibly already reaching its limited capacity. Currently, ultra-HTS can sample 100 000 compounds per day using 384 or 1536-well microplates. In addition, this capacity depends on the availability of real compounds for cell-based or biochemical assays, which is not always the case.

Finally, HTS often reports biased results. It encounters frequent hitters, aggregation problem, and "Pan-Assay Interference" compounds [22], [23] and frequently reports false-positive [24] and false-negative [25], [26]. Moreover, data handling and analysis are challenging in addition to the artifacts from readout technologies [27].

2. The role of virtual screening (VS)

One approach to alleviate the ugliness of HTS is to embrace virtual screening (VS), which complements HTS, reducing the number of compounds to be tested experimentally.

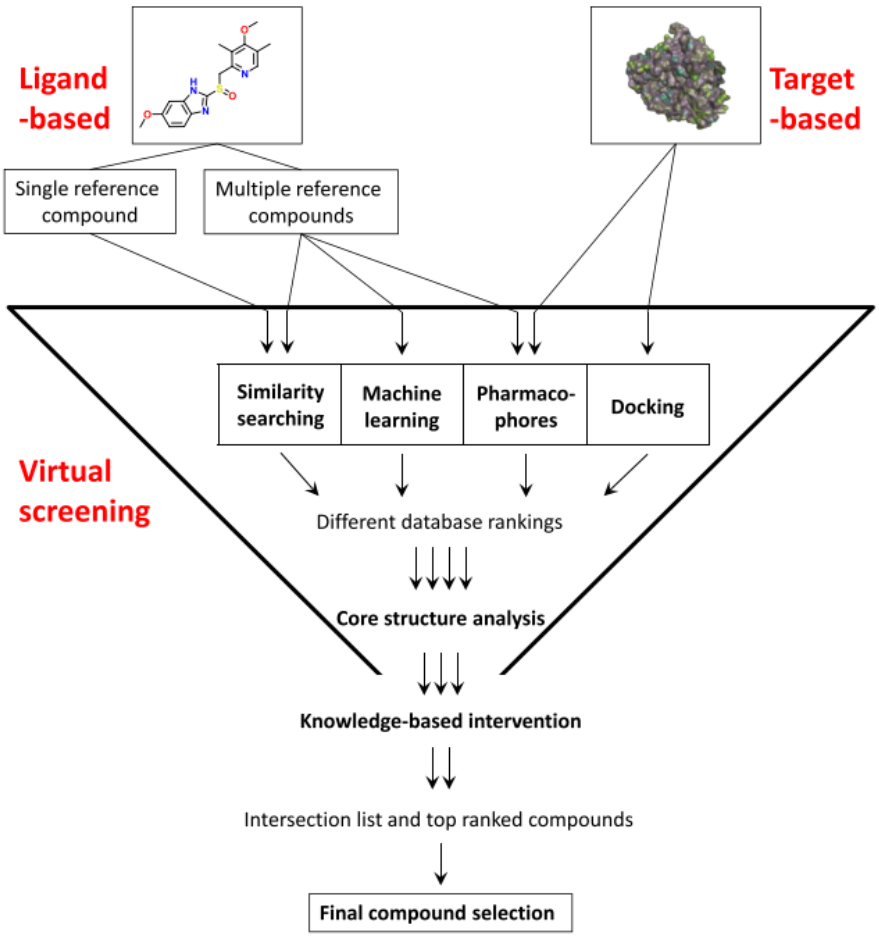

Figure 3. Overview of two categories in virtual screening [28].

Two major VS categories (figure 3) have been used over two decades are ligand-based virtual screening (LBVS) and target-based virtual screening (TBVS) [28], [29]. TBVS is conducted when structures of target molecules are known based on data from X-ray diffraction and NMR [29] or relied on homology modeling [29]–[31]. LBVS, on the other hand, is used when active ligands are known or the structure of targets is established [32].

Many scientists have tried to figure out the contributions of VS [33]–[37], but it has not been easy because it takes many years to see if the hits found by VS turn up to be approved drugs. In addition, this field is still evolving and not all VS results are in public domain, particularly those from pharmaceutical industry. To circumvent this difficulty, for instance, Van Vlijmen et al. had assessed the contribution of computational chemist in drug discovery based on the number of parents that these chemists hold [37], while others have mentioned highly potent and diverse chemotypes as successful proofs [33], [34], [36], [38]–[41].

To offer a global view on how VS contributes to DDD, Slater suggested that VS performance can be evaluated by focusing on retrospective studies to extract statistical data [42].

At smaller and specific case, VS performance can be assessed based on the ability to find hit compounds from a vast chemical space. In the best scenario hit compounds become approved drugs. Two questions guide us to assess different aspects of VS: (i) are VS methods reliable? (ii) is VS more efficient than other approaches in searching for hits?

The first question asks for the validation of computational methods used in VS. The most common way based on a reference library, containing known active compounds and decoys [43], [44]. If VS reproduces the experimental result, the method is accepted. In addition, the binding modes of the hits found must be experimentally verified [40], [45], [46] and likewise, the biological activities must be reassessed through many bioassay formats in laboratory setting to ensure the reproducibility of results [47], [48].

A similar evaluation method is comparing VS results with that of HTS [49]. Doman et al. for instance found that hit rate of VS was 1700-fold higher than HTS [38]. Similarly, Paiva et al. reported that hit rates from VS was 30-fold higher than HTS [50]. Most recently, Damm-Ganamet et al. reported that hits found by HTS overlapped with more than 70% found by VS [40]. However, these groups used different libraries for screening, the comparison is therefore imperfect.

The second question concerns the quantity and quality of hits found. Despite the great effort made, few hits discovered by VS have become drug candidates and approved for patients [34]. So far, structure-based discovery has helped to bring approximate 20 drugs in clinical uses [51].

Typically, about 0.1 – 2.5% top-ranked molecules in VS result are selected as hit compounds, containing new chemotypes, on which potency can be improved for further exploration [18], [52], [53]. From 2014 to 2018, PubMed indexes about more than 100 publications using VS, altogether filtered over 93 million compounds and found more than 12500 hits and bioassays confirmed that approximately 10% of these hits were actually bioactive [42].

The success of VS in drug discovery can be demonstrated occasionally through some triumphant stories. For example, compound PRX-08066 (**1** in figure 4), a potent and selective antagonist at serotonin 5-HT_{2B} receptor against pulmonary arterial hypertension [54], was discovered and designed with the aid of computational approaches at EPIX Pharmaceuticals [34]. This compound has 5-HT_{2B} binding affinity (Ki) of 3.4 nM and has been also studied for cancer treatment [55]. Similarly, Damm-ganamet et al. has found a

highly potent hit series which became leads, among them one lead - a quinoline tertiary alcohol (**2** in figure 4), has been developed to be a New Chemical Entity [40]. Another example comes from study of Al-Sha et al., they screened more than 240000 molecules against Hsp90α protein and found many highly active hits which were confirmed by subsequent *in vitro* assays. The most potent hits are compound **3**, **4** and **5** (in figure 4) with experimental IC_{50} of 3, 5, 6 nM, respectively.

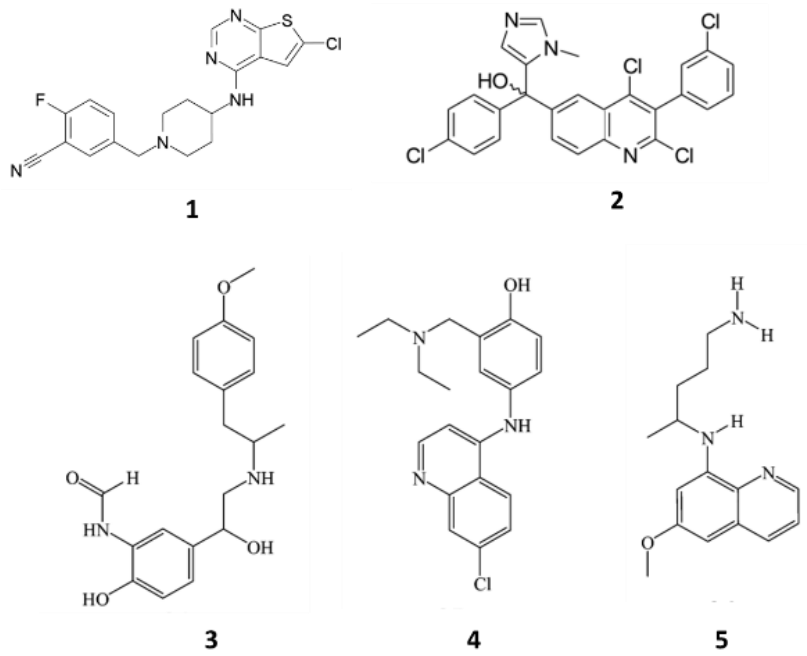

Figure 4. Some most potent hits found by VS.

Most recently, Lyu et al. screened 130 and 170 million compounds against AmpC and D4 dopamine receptor, finding a phenolate inhibitor of AmpC and after being optimized, it attained binding affinity of 77 nM, placing it at the top most potent non-covalent inhibitors known [18]. Likewise, Gahlawat et al. screened libraries of natural products, FDA-approved drugs and known inhibitors; they found many potential lead compounds [56].

Even though we are unable to assess holistically VS's performance because of biased publish data [57] or lack of information from pharmaceutical industry [42], these few successful stories clearly demonstrated that VS has highly potential in DDD providing that it accesses to quality libraries and equipped with reliable computational methods. Unfortunately, these two conditions uncover the inherent limitations of VS.

3. Limitation of VS

Although VS is a promising technique to find hits and new chemotypes for lead compounds, it needs to overcome two limitations related to algorithm and database to achieve its full potential.

The largest drawback of VS is its suboptimal algorithms. Hits reported from VS frequently contain false-positive compounds [34], [57], [58]. Because VS screens huge chemical libraries, the number of false-positive can be large, the cost of synthesis and *in vitro* tests to confirm results can be prohibitive. One reason is attributed to the current algorithms, which do not take into account the flexibility of target

molecules, knowing as a critical factor determining the accurate docking pose and binding affinity prediction, thus the ranking result in VS [42], [59]. Sometimes the algorithm is too specific; it works effectively for some systems but not the others [60], preventing it from screening across ligand-target systems.

The other limitation links with the quantity, quality and accessibility of ligand and target databases [61], [62]. Presently, chemical databases for DDD are fragmentary, some are in public domain, others are industrial. VS currently is unable synchronize all these sources of databases to access potential compounds more efficiently. Furthermore, the ligand databases are still too tiny in comparison to the possible chemical space [62]. This limitation asks for more effort to build standard and libraries with easily access [63]. If size of and accessibility to databases are important, so is quality. Since VS cannot find good hits if the stock does not hold that molecules as experiences demonstrated [64].

4. Outlook

VS has made grand contributions to DDD regardless of its many limitations. Fortunately, this field is still evolving, especially to cope with big data in Chemistry [65]. Building larger target databases to accelerate drug discovery process is especially useful for VS [66]. Recently, sequencing of human genome and pathogen [2], [67], [68] and the advancement of high-throughput crystallography and NMR [42] have offered more targets for VS.

Similarly, constructing huge and focused ligand libraries for VS is significant [62], [69]–[71]. Compounds are sourced from natural products [72], [73], combinatorial chemistry, DNA-encode library [74], [75]. These databases can be stored on cloud-based system to ease access [76]. Aiming at target screening, focused chemical databases have been built using lead-like and drug-like properties as well as target relevant characteristics as filtration criteria before screening. Today, many libraries (table 2) have been made available for public usage [77]–[79].

Table 2: Some big and popular libraries for virtual screening

Library	Source and purpose	Number of compounds	URL
ChEMBL	bioactive compounds with drug-like properties from medicinal chemistry literature	15 million	https://www.ebi.ac.uk/chembl/
PubChem	compounds from academic screening centers	109 million	https://pubchem.ncbi.nlm.nih.gov/
ChemSpider	compound structures from multiple sources	101 million	http://www.chemspider.com/
ZINC	commercially-available compounds for virtual screening	980 million	https://zinc.docking.org/
SureChEMBL	compounds from chemical patents	17 million	https://www.surechembl.org

Developing better computational methods to account for the flexibility of drug target is also an obvious trend [42], [80]. To simplify the complexity of the system, most of the algorithms in the past studies ignore the flexibility of target molecules and this is one of the reasons the screening result is not satisfactory. The improved methods are expected to deliver accurate VS results.

Nowadays, the most obvious trend in drug discovery is the application of AI, specially machine learning and deep learning have already show many promising potentials [81], [82] dealing successfully with huge chemical databases.

5. Conclusion

Virtual screening has made important contributions to DDD, providing many hit and lead compounds with diverse chemical scaffold for further exploitation. Hitherto, to achieve its full potential, VS must overcome two challenges. First, it requires more reliable computational methods to account for the flexibility of drug targets and better scoring functions. Second, VS demands big and high-quality chemical libraries to be used as reference database to validate computational methods, simultaneously as a pool to search for good hits. The screening efficiency would be higher if the chemical libraries are synergized and focused, and can be accessed easily. The future of VS relies on its capability to handle reliably big chemical databases using AI.

References

[1] J. Inglese *et al.*, "High-throughput screening assays for the identification of chemical probes," *Nat. Chem. Biol.*, vol. 3, no. 8, pp. 466–479, 2007.

[2] L. M. Mayr and P. Fuerst, "The future of high-throughput screening," *J. Biomol. Screen.*, vol. 13, no. 6, pp. 443–448, 2008.

[3] J. Noah, "New developments and emerging trends in high-throughput screening methods for lead compound identification," *Int. J. High Throughput Screen.*, p. 141, 2010.

[4] D. A. Pereira and J. A. Williams, "Origin and evolution of high throughput screening," *Br. J. Pharmacol.*, vol. 152, no. 1, pp. 53–61, 2007.

[5] L. M. Mayr and D. Bojanic, "Novel trends in high-throughput screening," *Curr. Opin. Pharmacol.*, vol. 9, no. 5, pp. 580–588, 2009.

[6] A. Golebiowski, S. R. Klopfenstein, and D. E. Portlock, "Lead compounds discovered from libraries: Part 2," *Curr. Opin. Chem. Biol.*, vol. 7, no. 3, pp. 308–325, 2003.

[7] S. Fox, H. Wang, L. Sopchak, and R. Khoury, "High-Throughput Screening: Early Successes Indicate a Promising Future," *J. Biomol.*, vol. 6, no. 3, pp. 137–140, 2001.

[8] A. Davis, D. Keeling, J. Steele, N. Tomkinson, and A. Tinker, "Components of Successful Lead Generation," *Curr. Top. Med. Chem.*, vol. 5, no. 4, pp. 421–439, 2005.

[9] R. MacArron *et al.*, "Impact of high-throughput screening in biomedical research," *Nat. Rev. Drug Discov.*, vol. 10, no. 3, pp. 188–195, 2011, [Online].

[10] R. Macarron, "Critical review of the role of HTS in drug discovery," *Drug Discov. Today*, vol. 11, no. 7–8, pp. 277–279, 2006.

[11] F. Prinz, T. Schlange, and K. Asadullah, "Believe it or not: How much can we rely on published data on potential drug targets?," *Nat. Rev. Drug Discov.*, vol. 10, no. 9, pp. 712–713, 2011, [Online]. Available: http://dx.doi.org/10.1038/nrd3439-c1.

[12] Editorial, "Stay on target," *Nat. Chem. Biol.*, vol. 9, no. 4, p. 193, 2013.

[13] J. Bajorath, "Integration of virtual and high-throughput screening," *Nat. Rev. Drug Discov.*, vol. 1, no. 11, pp. 882–894, 2002.

[14] E. J. Gordon, "Small-Molecule Screening: It takes a Village," *ACS Chem. Biol.*, vol. 2, no. 1, pp. 9–16, 2007.

[15] R. Lahana, "How many leads from HTS?," *Drug Discov. Today*, vol. 4, no. 10, 1999.

[16] C. A. Lipinski, "Lead- and drug-like compounds: The rule-of-five revolution," *Drug Discov. Today Technol.*, vol. 1, no. 4, pp. 337–341, 2004.

[17] J. L. Reymond and M. Awale, "Exploring chemical space for drug discovery using the chemical universe database," *ACS Chem. Neurosci.*, vol. 3, no. 9, pp. 649–657, 2012.

[18] J. Lyu *et al.*, "Ultra-large library docking for discovering new chemotypes," *Nature*, vol. 566, no. 7743, pp. 224–229, 2019, [Online].

[19] UWCCC Research, "Services, Equipment & Pricing," *University of Wisconsin-Madison*. https://cancer.wisc.edu/research/resources/ddc/smsf/equipment-services/ (Jan. 2021).

[20] M. Zaragoza-Sundqvist, H. Eriksson, M. Rohman, and P. J. Greasley, "High-quality cost-effective compound management support for HTS," *J. Biomol. Screen.*, vol. 14, 2009.

[21] U. of Helsinki, "High through-put chem-ical screen-ing." https://www.helsinki.fi/en/infrastructures/drug-discovery-chemical-biology-and-screening/infrastructures/high-throughput-biomedicine/(accessed Jan. 21, 2021).

[22] N. Thorne, D. S. Auld, and J. Inglese, "Apparent activity in high-throughput screening: origins of compound-dependent assay interference," *Curr. Opin. Chem. Biol.*, vol. 14, no. 3, pp. 315–324, 2010.

[23] J. L. Dahlin *et al.*, "Post-HTS case report and structural alert: Promiscuous 4-aroyl-1,5-disubstituted-3-hydroxy-2H-pyrrol-2-one actives verified by ALARM NMR," *Bioorganic Med. Chem. Lett.*, vol. 25, no. 21, pp. 4740–4752, 2015, [Online].

[24] A. Jadhav *et al.*, "Quantitative analyses of aggregation, autofluorescence, and reactivity artifacts in a screen for inhibitors of a thiol protease," *J. Med. Chem.*, vol. 53, 2010.

[25] S. C. Owen, A. K. Doak, P. Wassam, M. S. Shoichet, and B. K. Shoichet, "Colloidal

Aggregation A ff ects the E ffi cacy of Anticancer Drugs in Cell Culture," 2012.

[26] W. P. Walters and M. Namchuk, "Designing screens: How to make your hits a hit," *Nat. Rev. Drug Discov.*, vol. 2, no. 4, pp. 259–266, 2003.

[27] I. Caraus, A. A. Alsuwailem, R. Nadon, and V. Makarenkov, "Detecting and overcoming systematic bias in highthroughput screening technologies: A comprehensive review of practical issues and methodological solutions," *Brief. Bioinform.*, vol. 16, no. 6, 2015.

[28] D. Stumpfe and J. Bajorath, "Current trends, overlooked issues, and unmet challenges in virtual screening," *J. Chem. Inf. Model.*, vol. 60, no. 9, pp. 4112–4115, 2020.

[29] Z. Chen *et al.*, "Pharmacophore-based virtual screening versus docking-based virtual screening: A benchmark comparison against eight targets," *Acta Pharmacol. Sin.*, vol. 30, no. 12, pp. 1694–1708, 2009.

[30] A. Evers and G. Klebe, "Successful virtual screening for a submicromolar antagonist of the neurokinin-1 receptor based on a ligand-supported homology model," *J. Med. Chem.*, vol. 47, no. 22, pp. 5381–5392, 2004.

[31] R. Rodriguez, G. Chinea, N. Lopez, T. Pons, and G. Vriend, "Homology modeling, model and software evaluation: Three related resources," *Bioinformatics*, vol. 14, no. 6, 1998.

[32] J. Y. Lee, J. M. Krieger, H. Li, and I. Bahar, "Pharmmaker: Pharmacophore modeling and hit identification based on druggability simulations," *Protein Sci.*, vol. 29, no. 1, 2020.

[33] D. E. Clark, "What has computer-aided molecular design ever done for drug discovery?," *Expert Opin. Drug Discov.*, vol. 1, no. 2, pp. 103–110, 2006.

[34] D. E. Clark, "What has virtual screening ever done for drug discovery?," *Expert Opin. Drug Discov.*, vol. 3, no. 8, pp. 841–851, 2008.

[35] G. Schneider, "Virtual screening: an endless staircase?," *Nat. Rev. Drug Discov.*, vol. 9, pp. 273–276, 2010.

[36] S. Kar and K. Roy, "How far can virtual screening take us in drug discovery?," *Expert Opin. Drug Discov.*, vol. 8, no. 3, pp. 245–261, 2013.

[37] H. van Vlijmen, R. L. Desjarlais, and T. Mirzadegan, "Computational chemistry at Janssen," *J. Comput. Aided. Mol. Des.*, vol. 31, no. 3, pp. 267–273, 2017.

[38] T. N. Doman *et al.*, "Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B," *J. Med. Chem.*, vol. 45, no. 11, 2002.

[39] M. Seifert and M. Lang, "Essential Factors for Successful Virtual Screening," *Mini-Reviews Med. Chem.*, vol. 8, no. 1, pp. 63–72, 2008.

[40] K. L. Damm-Ganamet *et al.*, "Accelerating Lead Identification by High Throughput Virtual Screening: Prospective Case Studies from the Pharmaceutical Industry," *J. Chem. Inf. Model.*, vol. 59, no. 5, pp. 2046–2062, 2019.

[41] J. A. Bikker and L. S. Narasimhan, "Lead Discovery Using Virtual Screening," *Top. Med. Chem.*, vol. 5, pp. 85–124, 2009.

[42] O. Slater and M. Kontoyianni, "The compromise of virtual screening and its impact on drug discovery," *Expert Opin. Drug Discov.*, vol. 14, no. 7, pp. 619–637, 2019, [Online].

[43] K. E. Hevener *et al.*, "Validation of Molecular Docking Programs for Virtual Screening against Dihydropteroate Synthase," *J. Chem. Inf. Model. Model.*, vol. 49, no. 2, 2010.

[44] L. Chaput, J. Martinez-Sanz, N. Saettel, and L. Mouawad, "Benchmark of four popular virtual screening programs: construction of the active/decoy dataset remains a major determinant of measured performance," *J. Cheminform.*, vol. 8, no. 1, pp. 1–17, 2016.

[45] G. L. Warren *et al.*, "A critical assessment of docking programs and scoring functions," *J. Med. Chem.*, vol. 49, no. 20, pp. 5912–5931, 2006.

[46] C. K. Chang *et al.*, "Structure-based virtual screening and experimental validation of the discovery of inhibitors targeted towards the human coronavirus nucleocapsid protein," *Mol. Biosyst.*, vol. 12, no. 1, pp. 59–66, 2016.

[47] W. L. Ye *et al.*, "Virtual screening and experimental validation of eEF2K inhibitors by combining homology modeling, QSAR and molecular docking from FDA approved drugs," *New J. Chem.*, vol. 43, no. 48, pp. 19097–19106, 2019.

[48] D. Lowe, "How to Know When a New Target is Really a New Target," *Science Translational Medicine*, 2017. https://blogs.sciencemag.org/pipeline/archives/2017/06/30/how-to-know-when-a-new-target-is-really-a-new-target (accessed Dec. 11, 2020).

[49] K. Babaoglu *et al.*, "Comprehensive Mechanistic Analysis of Hits from High-Throughput and Docking Screens against β-Lactamase," *J. Med. Chem.*, vol. 51, no. 8, 2008.

[50] A. M. Paiva, D. E. Vanderwall, J. S. Blanchard, and J. W. Kozarich, "Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of Mycobacterium tuberculosis," *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.*, vol. 1545, no. 1–2, pp. 67–77, 2001.

[51] J. J. Irwin and B. K. Shoichet, "Docking Screens for Novel Ligands Conferring New Biology," *J. Med. Chem.*, vol. 59, no. 9, pp. 4103–4120, 2016.

[52] R. D. Cramer, "Leadhopping – and beyond," *Expert Opin. Drug Discov.*, vol. 1, 2006.

[53] V. Katritch *et al.*, "Structure-based discovery of novel chemotypes for adenosine A2A receptor antagonists," *J. Med. Chem.*, vol. 53, no. 4, pp. 1799–1809, 2010.

[54] S. L. Porvasnik *et al.*, "PRX-08066, a novel 5-hydroxytryptamine receptor 2B antagonist, reduces monocrotaline-induced pulmonary arterial hypertension and right ventricular hypertrophy in rats," *J. Pharmacol. Exp. Ther.*, vol. 334, no. 2, pp. 364–372, 2010.

[55] B. Svejda *et al.*, "The 5-HT2B receptor plays a key regulatory role in both neuroendocrine tumor cell proliferation and the modulation of the fibroblast component of the neoplastic microenvironment," *Cancer*, vol. 116, no. 12, pp. 2902–2912, 2010.

[56] A. Gahlawat *et al.*, "Structure-Based Virtual Screening to Discover Potential Lead Molecules for the SARS-CoV-2 Main Protease," *J. Chem. Inf. Model.*, 2020.

[57] D. Stumpfe, P. Ripphausen, and J. Bajorath, "Virtual compound screening in drug discovery," *Future Med. Chem.*, vol. 4, no. 5, pp. 593–602, 2012.

[58] Y. O. Adeshina, E. J. Deeds, and J. Karanicolas, "Machine learning classification can reduce false positives in structure-based virtual screening," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 31, pp. 18477–18488, 2020.

[59] S. Saikia and M. Bordoloi, "Molecular Docking: Challenges, Advances and its Use in Drug Discovery Perspective," *Curr. Drug Targets*, no. 5, 2019.

[60] E. Lionta, G. Spyrou, D. Vassilatis, and Z. Cournia, "Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances," *Curr. Top. Med. Chem.*, vol. 14, no. 16, pp. 1923–1938, 2014.

[61] X. Zhang, S. Betzi, Z. Morelli, and P. Roche, "Focused chemical libraries – design and enrichment: an example of protein–protein interaction chemical space," *Future Med. Chem.*, vol. 6, no. 11, pp. 1291–1307, 2014.

[62] W. P. Walters, "Virtual Chemical Libraries," *J. Med. Chem.*, vol. 62, no. 3, 2019.

[63] F. I. Saldívar-González, C. S. Huerta-García, and J. L. Medina-Franco, "Chemoinformatics-based enumeration of chemical libraries: a tutorial," *J. Cheminform.*, vol. 12, no. 1, pp. 1–25, 2020, [Online].

[64] Editorial, "Screening we can believe in," *Nat. Chem. Biol.*, vol. 5, no. 3, p. 2009, 2009.

[65] I. V. Tetko, O. Engkvist, U. Koch, J. L. Reymond, and H. Chen, "BIGCHEM: Challenges and Opportunities for Big Data Analysis in Chemistry," *Mol. Inform.*, vol. 35, 2016.

[66] Z. R. Tanoli *et al.*, "Drug Target Commons 2.0: A community platform for systematic analysis of drug-target interaction profiles," *Database*, vol. 2018, no. 2018, 2018.

[67] L. Folkersen *et al.*, "Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals," *Nat. Metab.*, vol. 2, no. 10, pp. 1135–1148, 2020, [Online].

[68] H. Fang, L. Chen, and J. C. Knight, "From genome-wide association studies to rational drug target prioritisation in inflammatory arthritis," *Lancet Rheumatol.*, vol. 2, no. 1, pp. e50–e62, 2020, [Online]. Available: http://dx.doi.org/10.1016/S2665-9913(19)30134-1.

[69] L. Ruddigkeit, R. Van Deursen, L. C. Blum, and J. Reymond, "Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17," *J. Chem. Inf. Model.*, vol. 52, pp. 2864–2875, 2012.

[70] D. E. Gloriam, "Bigger is better in virtual drug screens," *Nature*, vol. 566, 2019.

[71] D. E. Clark, "Virtual screening: Is bigger always better? Or can small be beautiful?," *J. Chem. Inf. Model.*, vol. 60, no. 9, pp. 4120–4123, 2020.

[72] J. Shen *et al.*, "Virtual Screening on Natural Products for Discovering Active Compounds and Target Information," *Curr. Med. Chem.*, vol. 10, no. 21, pp. 2327–2342, 2005.

[73] J. M. Rollinger, H. Stuppner, and T. Langer, "Virtual screening for the discovery of bioactive natural products," *Prog. Drug Res.*, vol. 65, pp. 212–249, 2008.

[74] R. A. Goodnow, C. E. Dumelin, and A. D. Keefe, "DNA-encoded chemistry: Enabling the deeper sampling of chemical space," *Nat. Rev. Drug Discov.*, vol. 16, no. 2, 2017.

[75] A. L. Satz, "What Do You Get from DNA-Encoded Libraries?," *ACS Med. Chem. Lett.*, vol. 9, no. 5, pp. 408–410, 2018.

[76] C. Grebner, E. Malmerberg, A. Shewmaker, J. Batista, A. Nicholls, and J. Sadowski, "Virtual screening in the cloud: How big is big enough?," *J. Chem. Inf. Model.*, vol. 60, no. 9, pp. 4274–4282, 2020.

[77] C. Gorgulla *et al.*, "An open-source drug discovery platform enables ultra-large virtual screens," *Nature*, vol. 580, no. 7805, pp. 663–668, 2020.

[78] A. Mullard and A. Mullard, "Large Pharmas Set Up Screening," vol. 18, 2019.

[79] J. J. Irwin and B. K. Shoichet, "ZINC – A Free Database of Commercially Available Compounds for Virtual Screening," *J. Chem. Inf. Model.*, vol. 45, no. 1, 2005.

[80] F. Spyrakis and C. N. Cavasotto, "Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration and active site water molecules description," *Arch. Biochem. Biophys.*, vol. 583, pp. 105–119, 2015, [Online].

[81] X. Yang, Y. Wang, R. Byrne, G. Schneider, and S. Yang, "Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery," *Chem. Rev.*, vol. 119, 2019.

[82] N. Fleming, "How artificial intelligence is changing drug discovery," *Nature*, vol. 557, no. 7707, pp. S55–S57, 2018.