Virtual Screening: Hope, Hype, and the Fine Line In Between

Hossam Nada[1,*], Nicholas A. Meanwell[2,3], Moustafa T. Gabr[1,*]

[1]Department of Radiology, Molecular Imaging Innovations Institute (MI3), Weill Cornell Medicine, New York, NY 10065, USA
[2]Baruch S. Blumberg Institute, Doylestown, PA, USA; School of Pharmacy, University of Michigan, Ann Arbor, MI, USA
[3]Ernest Mario School of Pharmacy, Rutgers University, Piscataway, NJ, USA

Corresponding:
[*]Moustafa T. Gabr: Department of Radiology, Molecular Imaging Innovations Institute (MI3), Weill Cornell Medicine, New York, NY 10065, USA. Email: mog4005@med.cornell.edu
[*]Hossam Nada: Department of Radiology, Molecular Imaging Innovations Institute (MI3), Weill Cornell Medicine, New York, NY 10065, USA. Email: hoh4004@med.cornell.edu

Abstract

Introduction: Technological advancements in virtual screening (VS) have rapidly accelerated its application in drug discovery, as reflected by the exponential growth in VS-related publications. However, a significant gap remains between the volume of computational predictions and their experimental validation. This discrepancy has led to a rise in the number of unverified 'claimed' hits which impedes the drug discovery efforts.

Areas covered This perspective examines the current VS landscape, highlighting essential practices and identifying critical challenges, limitations, and common pitfalls. Using case studies and practices, this perspective aims to highlight strategies that can effectively mitigate or overcome these challenges. Furthermore, the perspective explores common approaches for addressing pharmacodynamic and pharmacokinetic issues in optimizing VS hits.

Expert opinion: VS has become a tried-and-true technique of drug discovery due to the rapid advances in computational methods and machine learning (ML) over the past two decades. Although each VS

workflow varies depending on the chosen approach and methodology, integrated strategies that combine biological and *in silico* data have consistently yielded higher success rates. Moreover, the widespread adoption of ML has enhanced the integration of VS into the drug discovery pipeline.

Key words: Virtual screening, structure-based virtual screening, Ligand-based virtual screening, Drug discovery.

Article Highlights

- The increasing adoption of virtual screening (VS) is driven by ongoing technological advances, promising outcomes, and increasing cost-effectiveness.
- Database selection plays a critical role in determining VS outcomes, with factors such as size, diversity, and preparation methods significantly influencing the results.
- Both prospective and retrospective validations are vital for advancing and refining VS methodologies.
- The lack of standardized evaluation criteria remains a significant challenge. Implementing such criteria would enable the objective assessment of the success or failure of VS studies.
- VS and high-throughput screening (HTS) are not mutually exclusive; when used together, they provide a more comprehensive evaluation framework.

1. Introduction

Virtual screening (VS) has become a cornerstone of computer-aided drug design (CADD) due to its contribution to the rapid and efficient identification of potential leads [1]. This importance has been highlighted by the growing number of publications employing VS protocols [2-4]. Figure 1 demonstrates the growing impact and application of VS, illustrating the number of articles containing the keyword “virtual screening” published since 1990. This rise in popularity can be attributed to the promise of VS to offer a high-throughput, cost-effective alternative method to traditional experimental screening methods such as high throughput screening (HTS) [5,6]. Moreover, the increasing availability of large public chemical libraries has further increased the popularity and attractiveness of VS [7,8].

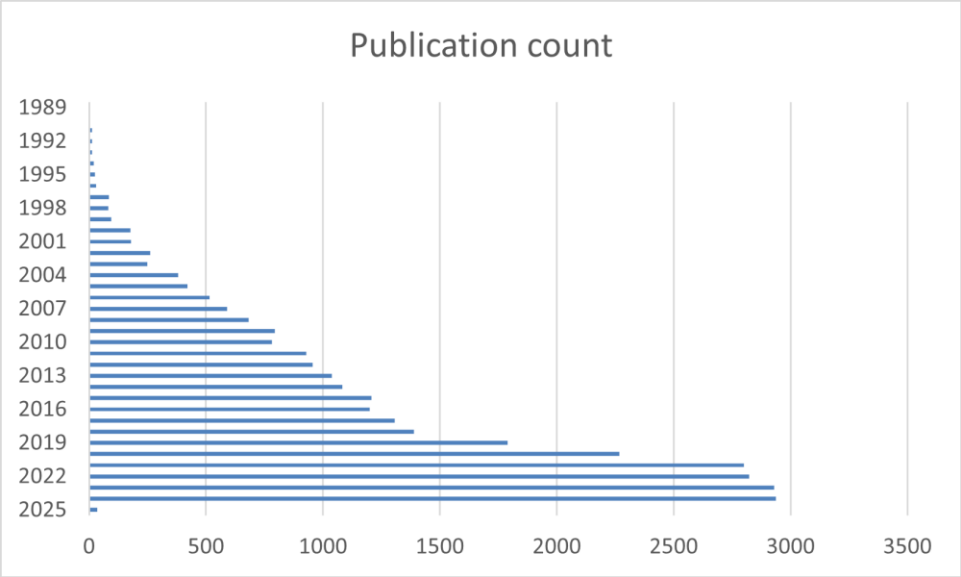

Figure 1. **The Rising Impact of VS:** Annual publication volume in Web of Science indexed for the keyword "virtual screening" from 1990 to January, 2025).

However, the widespread adoption of VS has been accompanied by a growing concern over the accuracy of predictions [9,10]. A significant gap often exists between computationally predicted active compounds and experimentally validated hits, underscoring the critical need for rigorous validation

strategies and the establishment of robust quality control standards in VS. Additionally, both VS and HTS share a common conceptual framework whereby both methods compensate for their limited accuracy by investigating a large number of compounds. Consequently, since both VS and HTS aim to identify active compounds, they are better viewed as complementary techniques. VS excels in predicting potential leads using computational methods, while HTS serves as a high-throughput experimental validation step. Both methods require integration in order to achieve more effective drug discovery outcomes. Active compounds identified through either VS or HTS must undergo experimental validation before progressing to the hit-to-lead stage. Thus, HTS and VS are not competitive methodologies; rather, they are complementary approaches [2,11]. However, in practice the integration of HTS and VS remains less extensive than expected.

The core principle of VS involves predicting the binding affinity to and selectivity of compounds for a target protein. This process encompasses several key components, including accurate molecular representation, the construction of diverse chemical databases and the application of ligand-based or structure-based computational methods [12]. Ligand-based VS methods often rely on similarities between known active compounds and database molecules while structure-based VS methods utilize the three-dimensional structure of the target protein to predict ligand binding [13-15]. Ultimately, VS aims to prioritize compounds with the highest potential for biological activity leading to the acceleration of the drug discovery process and at reduced cost [16,17]. However, the predictive nature of VS requires rigorous experimental validation to confirm the biological activity and therapeutic potential of identified compounds. Experimental validation is essential to verify computational predictions in order to ensure that the VS process ultimately translates into tangible and effective molecules for optimization.

This perspective will examine the fundamental principles of VS by exploring the various computational methods and techniques employed. Additionally, the main challenges and limitations of current VS approaches will be explored with the aim of highlighting the common misconceptions about

VS, share valuable lessons learned, and recommending best practices for its effective implementation in the drug discovery process.

2. Virtual screening: an overview

VS employs structural and chemical information that is applied toward the identification of potential lead molecules from compound libraries (Figure 2). This section provides an overview of ligand-based, structure-based, and ML-based VS methods and their associated validation strategies. A well-curated and validated database constitutes the foundation of successful VS, followed by the application of appropriate screening techniques and subsequent validation of the results.

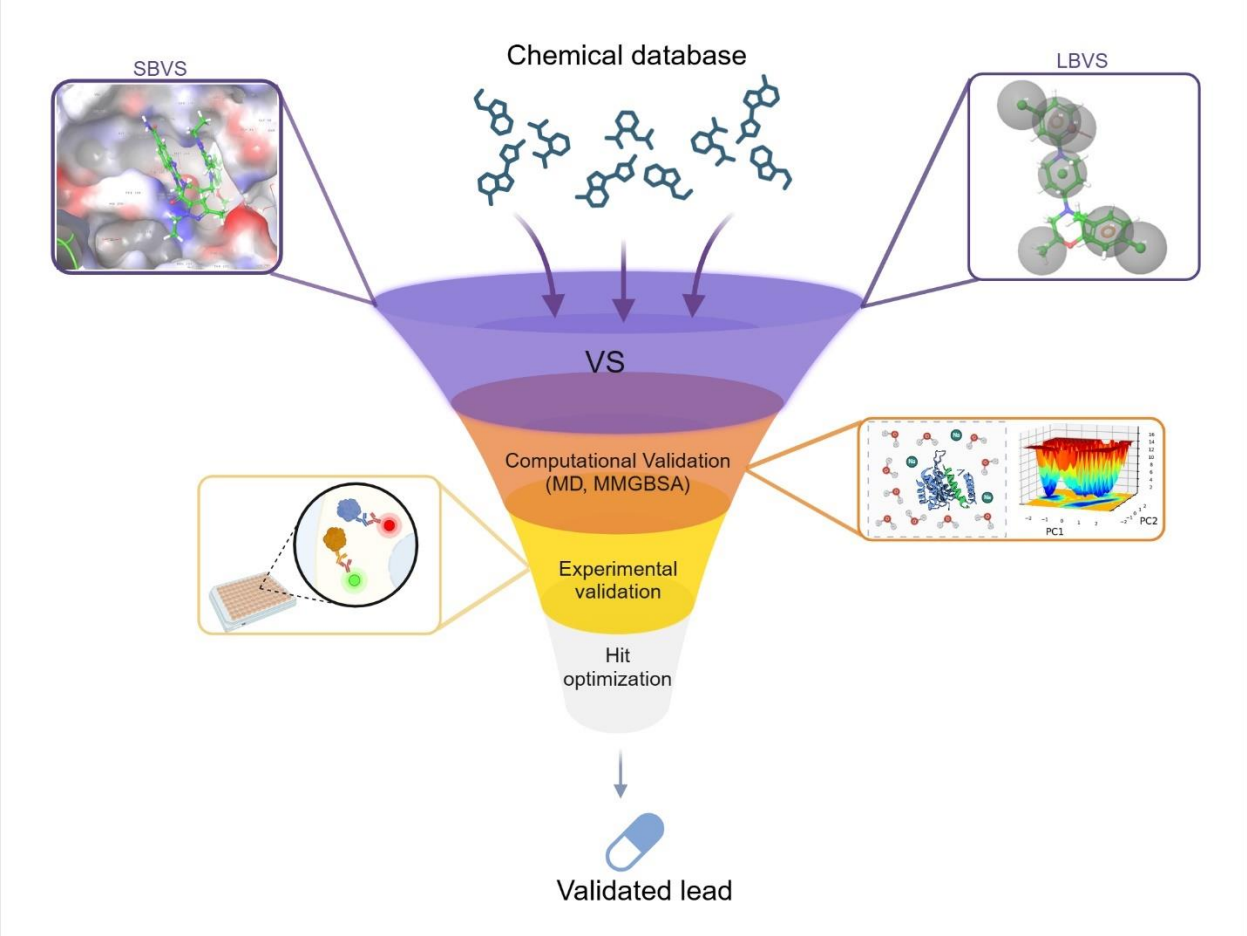

Figure 2. Representative workflow in virtual screening-based studies. An example of how VS methodologies can be integrated with computational and experimental validation techniques to accelerate lead discovery.

2.1. Database: choice and preparation

One of the key factors that influences the eventual success or failure of a VS study is the quality and the proper curation of a selected ligand database [18,19]. Proper database compilation necessitates accurate charge assignment, correct stereochemistry representation, and consideration of other pertinent chemical parameters [20,21]. The chosen scoring function are another factor that plays a key role in the success of any VS study since it must effectively incorporate and appropriately balance all relevant energetic components. Scoring functions can be classified into physics-based, knowledge-based, empirical and more recently ML-based scoring functions. Each type of scoring function has its own advantages and limitations, with the consensus being that a combination of different scoring functions to generate hybrid scoring functions can lead to improved accuracy. There are several literature studies that have both categorized and summarized the different scoring functions in order to facilitate their informed application in VS [22-25]. However, it should be noted that relying solely on scoring functions for selecting compounds for optimization or validation is not advisable. It is crucial to follow up initial scoring function assessments with rigorous validation to confirm the predicted interactions and ensure the reliability of the selected compounds.

The size of the chosen database and the diversity of the chemical space that it explores are additional factors that can affect the eventual outcome of VS [26-28]. The size of the database influences the likelihood of discovering active molecules, with larger databases affording an increased probability of identifying relevant hits. Meanwhile, a diverse compound collection containing a range of substructures (scaffolds) that are amenable to rapid library expansion is highly advantageous [29-31]. Database diversity can be quantified using distance-based, cell-partitioning, or clustering methods [32-34]. Currently, there are a large number of available chemical databases and compound collections which are readily accessible

from commercial vendors and academic institutions. Notable examples include ZINC, ChemDB, PubChem, the National Cancer Institute (NCI) database and the DrugBank chemical database among many others [35-39]. One strategy to enhance the quality of chemical databases for VS involves including compounds that are known to bind to the desired target in the sample set of screening candidates. This allows assessment of how well these known binders "enrich" during the screening process, thereby improving the overall effectiveness of the VS exercise [40,41].

Despite the capacity of VS to evaluate a vast number of compounds, screening the entire available chemical space is impractical due to its scale. This problem can be overcome by introducing rules to eliminate unlikely candidates leading to a more efficient and streamlined process. Among the various rules for database preparation are drug-likeness criteria, including Lipinski's Rule of Five, and more stringent lead-likeness criteria, like the Rule of Three. Additionally, "Pfizer's Rule of 3/75" targets the identification and elimination of compounds with potential safety and toxicity concerns, while "Jorgensen's Rule-of-Three" emphasizes lead-like properties [42-46]. Depending on the strategy employed in the subsequent VS campaign, multiple conformers can be pre-calculated and stored as separate entries in the database. However, these rules may overlook candidates for unconventional targets such as protein-protein interactions or the generation of peptidomimetics. Therefore, the choice of which rules to apply depends entirely on the specific target of the VS and the targeted lead properties.

2.2. Structure based virtual screening (SBVS)

2.2.1. Target receptor: drugability and structural features

There are several target receptor-related factors that must be considered and resolved prior to starting a VS study. This section will discuss some of those factors and the criteria that should be met before starting the VS survey.

The likelihood of successfully modulating a target with a therapeutic agent is referred to as target drugability. However, not all biological targets of interest have been considered suitable for drug discovery, necessitating a proper understanding of drug properties is essential for the effective evaluation of the drugability of a specific target. Most drugs bind to defined sites on protein targets known as “hot spots” and comprehending these binding sites is fundamental to structure-based drug design and the success of VS. Currently, there are several programs, software and ML algorithms that can be exploited to predict the possible binding sites of the intended target with varying accuracy [47-49]. Experimental techniques like SPR and chemical probes can complement computational modeling to identify key binding residues [50,51].

The flexibility of the target receptor is another challenging aspect of VS due to the number of possible conformations exhibited by a target protein as a function of its innate flexibility.[52,53] The flexible nature of molecular targets can be accounted for by incorporating the dynamic behavior of a protein during screening or the use of an ensemble of rigid protein conformers. Protein conformations suitable for ligand binding can be derived from two primary approaches. The first involves exhaustive conformational sampling using computational methods such as molecular dynamics (MD) simulations to explore the dynamic landscape of the target protein. Alternatively, a more static approach can be employed to assess protein conformations by analyzing multiple protein-ligand complex crystal structures to identify a representative set of conformations. This latter method assumes that experimentally-determined structures correspond to low energy, biologically relevant conformations [54-58].

Lastly, the choice of which crystal structure to employ for VS is a key factor contributing to the potential success of a computational survey. Protein-ligand complexes are typically obtained through soaking or co-crystallization methods. In soaking, a ligand diffuses into preformed protein crystals, whereas in co-crystallization the complex forms in solution prior to crystallization. Different soaking protocols can yield distinct protein conformations which significantly impact docking and scoring outcomes. Ligands may adopt reversed orientations or induce protein conformational changes depending

on the crystallization conditions. Increasing the ligand concentration in soaking buffers can enhance occupancy but may also introduce kinetic traps which leads to unexpected co-crystallization results such as multiple ligands or multiple binding modes. Therefore, before utilizing a crystal structure for VS it is essential to carefully analyze parameters such as ligand occupancy and hydrogen bond networks toward selecting the most representative conformation [40,59,60].

An alternative to X-ray crystallography is Cryo-EM, a biophysical technique employed to establish the structure of biological molecules that enables the structural determination of large or dynamic macromolecular assemblies in solution without requiring crystallization. One of the advantages of Cryo-EM stems from its ability to capture biological structures in their near-native state, including proteins subject to post-translational modifications. Additionally, Cryo-EM can resolve multiple conformational states within a single experiment. There are distinct differences between X-ray crystallography and Cryo-EM which should be considered when choosing a biological structure for VS. Table 1 summarizes some of the key differences between X-ray crystallography and Cryo-EM [61].

Table 1. Comparative analysis of X-ray crystallography and Cryo-EM [61]

Feature	X-ray Crystallography	Cryo-EM
Resolution	High resolution (<3 Å) is readily achievable and reliable.	Effective for intermediate resolutions (4–7.5 Å), with clear visualization of secondary structures; <3 Å resolutions are becoming more common but remains a challenge.
Throughput	Extremely fast; ~500 crystals can be screened in 24 hours, data processed within 1 week.	Significantly slower; similar throughput would require months for data collection and analysis.
Automation	Highly automated processes for data integration, structure solution, and refinement.	Automation is increasing, but manual efforts are still required for model building and refinement.
Sample Preparation	Requires large amounts of protein and successful crystallization, which can be challenging, especially for complex proteins.	Requires smaller quantities of protein and does not rely on crystallization; suitable for large, dynamic, or membrane proteins.
Native State Analysis	Limited in studying proteins in their native, dynamic states or proteins with post-translational modifications.	Ideal for analyzing proteins in their native state, including those that are flexible of incorporate post-translational modifications .
Speed of Analysis	Very fast: data collection at a synchrotron takes <2 minutes; analysis completed in ~1 hour.	Slow: ~8 hours per dataset for collection and extensive computational time for analysis.
Phase Information	Phase information is lost and must be reconstructed, which can be challenging at intermediate resolutions.	Phase information is directly accessible, enabling better visualization of structural features at intermediate resolutions.

Suitability for Drug Discovery	Well-suited due to rapid ligand screening and high throughput.	Increasingly suitable for drug discovery, particularly for targets that are challenging to crystallize.
High-Resolution Challenges	Reliable for <3 Å; challenges arise beyond 3.5 Å due to poor observation-to-parameter ratios.	Achieving <3 Å resolution can be challenging due to conformational variability, beam damage, and optical aberrations.
Intermediate Resolution Quality	Structures at 3.5 Å resolution may lack clarity compared to cryo-EM at similar resolutions.	Better-defined structural features at intermediate resolutions (~3.5 Å) compared to X-ray crystallography.

2.2.2. Structure-based virtual screening (SBVS): techniques, limitations and solutions

Following the careful preparation of the target receptor, SBVS involves docking each compound in a library into the receptor's binding site using software such as AutoDock, Glide, Gold, LigandFit and FLARE [62-66]. This process predicts the three-dimensional orientation of the ligands within the protein binding pocket of the target receptor. Subsequently, a scoring function is applied to estimate the binding affinity of each predicted pose. These scores are derived from various computational methods including force field-based, empirical, knowledge-based approaches, or combinations thereof. The compounds are then ranked based on their predicted binding affinities [67,68].

From a physics perspective, scoring functions can be broadly classified into two categories: force field-based approaches and quantum mechanics-based approaches. Force field-based scoring functions estimate the binding energy of protein-ligand complexes by summing the van der Waals and electrostatic interactions between the individual atom pairs of the protein and ligand. While the force-field approach accounts for the contribution of enthalpy to energy while calculating the docking score it doesn't account for the effect of the solvents and entropy. The limitations of force field-based scoring functions can be addressed by employing quantum mechanics-based scoring functions. Quantum-mechanics-based scoring functions are characterized by improved accuracy and the ability to account for covalent interactions, polarization effects and charge transfers, which are often challenging for force field-based approaches. However, the enhanced precision of quantum mechanics-based scoring functions comes at the cost of significantly higher computational demands and longer calculation times. Scoring functions can also be classified into knowledge-based, empirical, and machine learning-based approaches [22,25,69].

Among the different applications for performing SBVS, 3D structure-based pharmacophore modeling has become a widely employed strategy. Leveraging the 3D structure of a macromolecular target or a macromolecule-ligand complex, these approaches generate pharmacophore models obtained from the analysis of the active site and the identification of the spatial arrangement of key chemical features. The generated features are then compiled to establish a 3D pharmacophore model capable of representing the key interactions which are necessary for ligand binding [70].

Despite its strengths, SBVS is not without limitations. The accuracy of the predictions is heavily reliant on the quality of the target protein structure which often include inaccuracies or missing information, potentially leading to erroneous results. Furthermore, the employed scoring functions often exhibit limitations in accurately predicting binding affinities. This is particularly evident when comparing structurally similar compounds. As a result, the top-ranked compound from a virtual screen frequently does not correspond to the most potent compound in experimental assays [71,72]. The assumption of protein rigidity in many docking programs can also hinder prediction accuracy, since proteins are inherently dynamic molecules [73]. Lastly, the composition and representation of chemical compounds within available databases can introduce biases in the screening process [18]. To bridge the gap between computational predictions and experimental outcomes, robust experimental validation strategies are essential.

Overall, validation methods for VS, both structure-based and ligand-based, can be broadly categorized into two types: prospective and retrospective validation. Retrospective validation is a method commonly used to evaluate the performance of the chosen VS protocol. Retrospective validation employs a dataset with known active and inactive compounds which gauges the effectiveness of the chosen method and identifies areas for improvement [74-76]. Conversely, prospective screening applies validated methodologies to screen databases of novel compounds which is followed by experimental evaluation to confirm the predicted activity.

VS is akin to HTS in drug discovery, as both enable the rapid, broad identification of potential hits. Once identified, these initial hits are then subjected to more sophisticated, resource-intensive methods that refine and validate the results. Methods commonly used to validate VS results include consensus docking, MD simulations, free energy calculations and the incorporation of experimental data to refine the prediction models and improve the accuracy of the screening outcomes. Consensus docking is an advanced computational strategy designed to improve the accuracy of predicting protein-ligand binding poses which can be used to enhance the accuracy of the VS results. Unlike traditional approaches that rely upon a single docking program, consensus docking integrates the outputs from multiple docking tools. This integration leverages the strengths of diverse algorithms and scoring functions which allows consensus docking to often outperform individual docking programs in terms of predictive accuracy. For example, studies have shown that consensus docking can achieve a success rate of 82% in pose prediction, compared to 55% for Autodock, 58% for DOCK, and 64% for Vina. Consensus docking emphasizes the concordance among predicted binding modes rather than the absolute values of predicted binding affinities. This approach helps to reduce the bias and limitations associated with any single docking method which leads to a more reliable and robust predictions [77,78]. Docking results can be further validated by utilizing MD simulations and free energy calculations which will be discussed in a separate section.

Meanwhile, combining VS with HTS can speed up the hit discovery process by utilizing VS to analyze the large datasets generated from HTS experiments and generate a more focused selection of compounds for further investigation [79,80]. Additionally, experimental data can be leveraged to create target-focused libraries for VS with higher likelihood of binding to the target [81]. Collectively, the aforementioned approaches optimize hit prioritization and accelerate the drug discovery process.

2.3. Ligand based virtual screening (LBVS)

LBVS is a method where known active ligands can be utilized to identify new hits from vast chemical databases [12,14]. The fundamental principle of LBVS is molecular similarity, which operates under the assumption that structurally similar compounds are likely to possess similar biological activities [82]. By applying this principle, LBVS employs computational algorithms to search extensive chemical libraries leading to the identification of hits which resemble known active compounds which are then experimentally tested as part of the validation process. A key application of LBVS is the discovery of novel chemotypes that can serve as lead structures for further optimization [83]. This approach not only expands the chemical space explored but also offers the potential to identify innovative leads. Moreover, it has the potential to address issues such as toxicity or pharmacokinetic (PK) problems associated with previously known active compounds, ultimately improving the overall drug discovery process [84,85].

LBVS is carried out using various methodologies that leverage different algorithms and molecular representation techniques toward the identification of promising leads within chemical databases. At its core, LBVS methodologies center around two key elements: an efficient similarity measure or coefficient and a robust scoring method [86]. Similarity coefficient is the parameter which quantitatively defines the required level of similarity between the lead compound and the database compounds [87-89]. A high similarity threshold enforces strict screening and reduces the likelihood of false positives while risking the exclusion of novel active compounds. Conversely, lowering the similarity threshold expands the search space toward potentially identifying more diverse hits but at the cost of increasing the probability of harvesting false positives. Accordingly, the optimal threshold is context-dependent and is influenced by the quality of the lead compound, the chemical diversity of the database and the balance between the desired specificity and novelty in the screening result.

There are various coefficients which can be employed with different fingerprint molecular databases to measure the similarity coefficient. Notable examples of commonly employed similarity coefficients in LBVS include the Tanimoto coefficient, Euclidean distance, Soergel distance, Dice index and

Cosine coefficient.[90-92] Among these, the Tanimoto coefficient is one of the most widely used similarity coefficients due to its ability in leveraging molecular fingerprints and 3D shape metrics to assess similarity. While metrics such as Euclidean distance are less common for molecular fingerprints, they are useful in specific property-based comparisons. Overall, the Tanimoto index, Dice index, Cosine coefficient, and Soergel distance are reported to be the best metrics for similarity as they produced the rankings closest to the composite (average) ranking in a comparative study [93].

LBVS typically can be classified into three main approaches: pharmacophores, molecular shapes and molecular fields [94]. Pharmacophore-based approaches rely on calculating the patterns of similarity derived from the calculation of distances between predefined molecular features such as aromatic rings or hydrogen-bond donors/acceptors [95,96]. Molecular shape-based methods aim to maximize the overlap of molecular shapes where the similarity percentage is determined by the degree of shape overlap.[97] Shape-based screening is particularly useful when the target shape is well-defined, such as when a ligand-bound structure is available [98]. Lastly, molecular field-based LBVS involves the comparison of electrostatic or steric fields around molecules to identify similarity patterns [99]. Meanwhile, the scoring method in LBVS is crucial for effectively ranking compounds by their activity. Accordingly, the chosen scoring method must possess the ability to distinguish between active and inactive compounds to allow for efficient identification of a small number of active compounds from a vast chemical library dominated by inactive ones.

Quantitative structure activity relationship (QSAR) modelling is a technique which, when integrated with LBVS, can result in improved accuracy [100]. QSAR models are constructed based on generating molecular descriptors for the different molecular features (e.g., size, shape, electronic properties, hydrophobicity) [101]. These molecular descriptors are then used to establish a quantitative relationship between the molecular structure and its associated biological activity. By correlating molecular descriptors with observed activities in known compounds QSAR models provide a cost-effective

approach that does not require prior knowledge of the target protein structure. Accordingly, QSAR models can be employed to validate the LBVS results or prioritize the predicted hits for further testing. However, QSAR modelling suffers from limitations, including its heavy reliance on the quality and representativeness of the training set data. A biased and/or limited diversity data set will result in a model with unreliable predictions. Another limitation of QSAR models is that they may struggle to accurately predict the activity of compounds with novel scaffolds due to the absence of similar scaffolds in the training set [102,103].

Both QSAR modelling and LBVS have made significant strides with the integration of ML algorithms due to their ability to learn complex patterns from large datasets[104]. These models are trained on known active and inactive compounds, learning to distinguish between them based on molecular descriptors or fingerprints. Once trained, the ML models can predict the activity of novel compounds, enhancing the accuracy and efficiency of VS. However, one major pitfall that computational chemists often encounter during the construction of ML models is the assumption that a high degree of structural similarity guarantees similar biological activity. Structural similarity has been shown to not necessarily imply similar interactions with a target macromolecule. A study that examined the IC_{50} values derived from 115 HTS assays showed that compounds with a Tanimoto similarity score of $\geq$0.85 to an active compound had only a 30% probability of being active themselves [105,106]. Although this 30% hit rate is considerably higher than the typical random screening hit rate of less than 0.01%, it still falls short of an optimal performance. It should be noted that the 30% probability was found to be approximately 10-fold higher than what was achieved using VS methods with Docking and Potential Of Mean Force (PMF) scoring, and 2-fold higher than Dock and Amber forcefield-based scoring [107].

The limitations of structural similarity-based screening arises from the inherent limitations of the chosen fingerprints and similarity calculations (e.g. Daylight fingerprints and Tanimoto similarity calculations) as well as the innate variability in how similar compounds interact with target macromolecules [108-110]. Accordingly, the integration of several methods, fingerprint and similarity

calculation methods is required for increased accuracy. Several reviews have comprehensively summarized and compiled the ML methodologies and applications utilized in VS [111-116]. Furthermore, the application of similarity searching is not confined to hit identification but can also be used for drug repurposing and identifying novel targets. For example, if a compound exhibits favorable cellular activity, such as antitumor activity, but fails to demonstrate activity against its intended target, these tools can be employed to rapidly identify potential alternative targets and mechanisms of action. Tools such as Swiss Similarity, Target hunter and PT-Finder are example of such applications [117-119].

2.4. *MD and free energy calculations*

MD simulations model the time-, pressure- and temperature-dependent behaviors of proteins and their interactions with ligands leading to a more realistic and detailed representation of the binding process [120]. MD can assess structural flexibility and the entropic effects which allows for the prediction of the kinetic and thermodynamic paramtets associated with the binding of drugs to their targets as well as mechanistic studies of biological proteins. MD simulations offer insights into the stability of ligand-protein complexes by analyzing the trajectory of the ligand within the binding site over time. This dynamic information complements the static snapshots derived from X-ray crystallography, leading to improved predictions [121]. Moreover, MD simulations can identify potential issues such as steric clashes or unfavorable interactions that might not be evident in traditional molecular docking or VS studies due to the static nature of their methodologies. Overall, MD simulations are a valuable tool for refining VS results, reducing false positives and prioritizing compounds with promising binding characteristics [122-125].

The use of MD simulations in the drug discovery pipeline has been widely adopted, with the diverse applications extensively documented in several studies [126-128]. This widespread adoption is due to the ability of MD to overcome the rigid limitations associated with traditional docking methodologies, thereby resulting in a more accurate but computationally demanding approach. MD simulations have

played a key role in various stages of docking and virtual screening workflows. These roles include early steps such as the preparation of the protein receptor where MD is employed to refine the receptor by accounting for protein flexibility and optimizing conformations prior to docking. More commonly, MD simulations are used to enhance the accuracy of docked poses by incorporating solvent effects and addressing the phenomenom of induced fit, where ligand binding triggers conformational changes in the receptor.

While MD simulations have become an invaluable tool in drug discovery, their successful application requires careful consideration of both potential pitfalls and the implementation of best practices. One common concern is the potential for the identification of false positives as the result of an analysis of MD simulation trajectories. One approach to reducing the identification of false positives is to conduct multiple independent simulations to assess the reproducibility of the results [129]. Furthermore, the duration of MD simulations significantly impacts the accuracy and reliability of the outcomes. While shorter simulations (50-100 nanoseconds) are often employed and have become the standard, longer simulations, which ideally extend to 1 millisecond or more, have shown the ability to capture more comprehensive dynamics and insights [130-132].

Over-reliance on root-mean-square deviation (RMSD) as the sole metric for assessing protein-ligand interactions is another factor which results in the identification of false positives in MD approaches. One metric that can be a strong indicator of the stability of the hit-protein complex is the number, distance and % of occupancy of the established hydrogen bonds between the hit and its target protein. A ligand forming fewer but stronger, fully occupied hydrogen bonds may be more promising than one forming numerous weaker or transient interactions. Gibbs free energy diagrams, which reveal the number of energy minima, are another tool with the capacity to efficiently identify and eliminate false positives. Ideally, a strong candidate from VS should exhibit a single, stable energy minimum, while multiple minima may suggest weaker binding [124,133].

Free energy calculations are another parameter which can be employed in reducing false positives in VS studies. Free energy calculations provide a more precise assessment of binding affinities by incorporating factors such as solvation and entropy. Free energy calculations can enhance VS by rescoring the top predicted hits, elucidating binding modes, profiling selectivity and quantifying energetic contributions. Such insights can more effectively guide the optimization of lead compounds, contributing to a more efficient drug discovery process. Although free energy calculations are computationally intensive, the rapid advances in computational power coupled with methodological advances have elevated free energy calculations to that of a key tool for validating the VS process.

Another key practice toward reducing false positives is the analysis of the individual components (e.g., van der Waals, electrostatic, solvation) that make up the free energy of interaction rather than simply focusing on the total free energy. This approach often provides deeper insights into the driving forces subtending ligand binding and can guide the optimization of potential hit molecules. Lastly, enhanced sampling techniques, such as umbrella sampling, metadynamics, and steered MD, can provide deeper insights into the binding stability of a hit with its target protein. These methods are particularly effective at reducing the number of false positives in a VS. By leveraging these advanced tools and addressing key considerations, the reliability, accuracy, and interpretability of MD simulations can be significantly enhanced[128,134,135].

3. VS challenges & solutions

Several review articles have highlighted the challenges, pitfalls and misconceptions associated with VS [9,16,136,137]. This section highlights some of the most common challenges encountered in VS and provides potential solutions toward addressing and avoiding these issues (Table 3). For simplicity, the challenges have been categorized to be: data-, SBVS- and LBVS-related challenges.

3.1. *Overcoming database related challenges*

One of the main challenges associated with the database are issues arising from the limited size or diversity of the prepared data set which leads to limited generalizability and biased results.[16,138] As such, the bigger and more diverse a dataset is the better the expected outcome of a VS study. Increasing the data set size and diversity will also solve the hit rate issues and biased results commonly observed with small or diversly biased data sets [9,16,136,137].

During the dataset preparation stage for a VS, it is crucial to identify and remove problematic molecules. For example, pan-assay interfering substances (PAINS) are chemically reactive molecules, promiscuous binders or aggregating molecules that can interfere with biological assays [139]. These compounds are notorious for producing false-positive results in HTS due to their nonspecific interactions. To address this issue, several computational platforms have been developed to identify and filter out such compounds. Beyond PAINS, several additional rules and frameworks have been developed that are designed to enhance and streamline the VS process. Table 2 provides an overview of representative tools and frameworks commonly utilized during the dataset preparation stage for VS. It is important to note that these rules and frameworks are not mutually exclusive and can be used in conjunction to provide a more comprehensive assessment.

Table 2. Common tools, frameworks and databases utilized in dataset preparation for VS.

Tool/Framework	Description	Advantages	Limitations
Lipinski's Rule of Five [140]	Filters compounds based on molecular weight (<500 Da), octanol-water partition coefficient (logP < 5), hydrogen bond donors (<5), and hydrogen bond acceptors (<10).	Simple, fast, widely accepted guideline for initial assessment of drug-likeness for oral administration.	May be overly restrictive, potentially excluding promising drug leads.
Veber's Rules [141]	Predicts oral bioavailability based on the number of rotatable bonds (<10)	Simple, fast, considers both flexibility and polarity.	May not be as accurate as more sophisticated models,

	and topological polar surface area (TPSA <140 Å²).		especially for complex molecules.
Ghose Filter [142]	Empirical guidelines that evaluate drug-likeness based on molecular weight (130-490 Da), number of heavy atoms (20-70), and octanol-water partition coefficient (-0.4 < logP < 5.6).	Simple, fast, provides a broader range of criteria for drug-likeness for oral administration.	May be overly restrictive, potentially excluding promising drug leads.
Egan's Rule [143]	Estimates oral absorption based on topological polar surface area (TPSA <140 Å²) and octanol-water partition coefficient (logP < 5).	Simple, fast, considers both polarity and lipophilicity.	May not be as accurate as more sophisticated models, especially for complex molecules.
Muegge's Rule [144]	Assesses drug-likeness based on molecular weight (<450 Da), octanol-water partition coefficient (-0.5 < logP < 5), hydrogen bond donors (<5), and hydrogen bond acceptors (<5).	Similar to Lipinski's Rule, but with slightly stricter criteria.	May be more restrictive than Lipinski's Rule, potentially excluding more compounds.
PAINS [139]	Identifies pan-assay interference compounds (PAINS) that often give false positives in screening assays due to inherent chemical reactivity or non-specific interactions.	Helps to remove problematic compounds early in the screening process, improving data quality.	May sometimes flag legitimate compounds, requiring careful interpretation.
Aggrescan [145]	Predicts protein aggregation propensity based on amino acid sequences.	Can help identify potential protein aggregation issues.	Accuracy may vary depending on the specific protein and the model used.
PubChem [146]	A database that provides free access to chemical structures and properties, including fluorescence data.	Can be used to identify compounds with known fluorescent properties.	Fluorescence data may not be comprehensive for all compounds.
ChemSpider [147]	An online database that provides free access to chemical information and allows for predictions of molecular properties including solubility and log*P*.	Easy to use, provides access to a large chemical database, useful for initial property estimations.	Prediction accuracy may vary depending on the specific property.

Lastly, context-dependent chemical characteristics such as tautomerism, ionization and chirality are often left unconsidered, which leads to reduced accuracy and can be a source of false positives or negatives. Accordingly, the proper preparation of the chemical database toward considering accurate tautomer selection, proper ionization state assignment and considering all chiral configurations are key factors toward improved VS assay results [148-150].

The employment of negative controls in VS is highly critical to success and these should be carefully selected to ensure the robustness of the screen. The chosen negative controls should ensure representativeness by being structurally or functionally dissimilar to the compounds being tested. Additionally, the negative controls should cover a wide range of chemical classes and properties as well as

being validated through prior experiments or literature to ensure they reliably show no activity. The use of putative inactive compounds such as decoys which may not be truly inactive may lead to false negatives and reduced sensitivity. As such, limiting negatives controls to experimentally confirmed, inactive compounds as decoys would lead to improved accuracy. Lastly, the conditions under which negative controls are tested should be consistent with those applied to the experimental compounds. Any variations in testing conditions could affect the reliability of the controls [9,151-153].

3.2. *SBVS pitfalls and solutions*

There are several challenges that face SBVS, among which is the ability of the chosen docking protocol to differentiate between the correct and incorrect binding poses, which can be an important limiting factor. Potential solutions for this problem are to integrate improved docking algorithms with the docking protocol or to employ enhanced scoring functions such as the previously highlighted ensemble methods. Additionally, there are several reviews and case studies that have compared the different commonly-employed docking systems and programs and which can be used to assess the best system to employ for a VS assay [69,154-158]. Another challenge in SBVS is the potential presence of multiple or allosteric binding pockets on the target protein with the result that SBVS may struggle to identify ligands that bind to these alternative or less conventional sites. To address this issue, it is crucial to conduct a comprehensive study that maps all potential binding sites on a target protein before the implementation of the VS protocol. Such a thorough analysis helps to ensure that all possible binding pockets are considered which improves the likelihood of identifying a broad range of potential modulators, including those that interact with allosteric sites. Conformational sampling and flexibility present another challenge toward the success of SBVS assay where generating representative conformations of molecules is difficult, leading to reduced accuracy and missed active compounds. Improved conformational sampling methods and proper consideration of the target flexibility can help mitigate this issue [159,160].

3.3. *LBVS pitfalls and solutions*

Accurate pharmacophore feature definition is a key to a successful LBVS assay where the pharmacophoric features need to be carefully considered based on using multiple active molecules for model building [14]. Additionally, software selection and associated limitations play key roles in the success of LBVS due to different tools exhibiting different levels of performance. The selection of software or models should be tailored to the specific requirements of the LBVS assay and supported by relevant literature to ensure its suitability for the intended application. Another important consideration in LBVS is the lack of prospective validation in many models. Often, models are not tested on new or independent datasets which can lead to overestimation of their performance and limit their generalizability to different or novel scenarios. Finally, overreliance on drug-likeness in VS often limits the exploration of non-drug-like compounds which may lead to missed opportunities for the discovery of novel compounds [161,162]. Expanding the chemical space and exploring non-drug-like libraries can help mitigate this issue.

Understanding these challenges and implementing the suggested solutions will increase the accuracy and effectiveness of VS. Table 3 summarizes the aforementioned challenges facing VS and their respective potential solutions.

Table 3. Overview of Challenges in VS and Potential Solutions

Category	Challenge	Description	Impact on VS	Potential Solutions
Data-Related Challenges	Data design and content issues	Limited/bias in benchmark data sets	Limited generalizability, biased results	Use larger/diverse data sets, careful negative control selection
	Hit rate issues in benchmark data sets	Low hit rates in standard data sets	Limited generalizability, biased results	Use high hit rate data sets, adjust performance metrics
	Problematic molecules (PAINS)	PAINS interfere with HTS detection	False positives, misleading results	Remove PAINS, use PAINS filters
	Context-dependent chemical characteristics	Tautomerism, ionization, chirality affects VS results	Reduced accuracy, false positives/negatives	Accurate tautomer selection, proper ionization state assignment
	Inactive decoys	Decoys may not be truly inactive	False negatives, reduced sensitivity	Use confirmed inactive decoys, curate decoy sets carefully

SBVS Challenges	Predicting incorrect binding poses	Docking-based VS may predict correct binding for wrong reasons	Low accuracy, false positives	Improved docking algorithms, better scoring functions, ensemble methods
	Variable water-mediated interactions	Difficulty in predicting water-mediated hydrogen bonds	Reduced accuracy, false negatives	Consider water molecules explicitly, improve water modeling
	Single vs. multiple/allosteric binding pockets	VS may miss ligands for alternative binding pockets	Missed novel compounds	Explore multiple binding sites, develop allosteric modulator identification
	Conformational sampling & flexibility	Difficulty in generating representative conformations	Reduced accuracy, missed compounds	Improved sampling methods, consider target flexibility
LBVS Challenges	Feature definition challenges	Incorrectly defined pharmacophore features affect accuracy	Reduced accuracy, false positives/negatives	Careful feature definitions, use multiple active molecules for modeling
	Software selection and limitations	Different tools vary in performance	Inconsistent results, suboptimal performance	Benchmark tools, select based on specific needs
	Lack of prospective validation	Models rarely validated on new data sets	Overconfidence, limited generalizability	Rigorous prospective/external validation
	Overreliance on drug-likeness	Focus on drug-like compounds limits applicability	Missed opportunities in non-drug-like compounds	Explore broader chemical space, non-drug-like libraries

4. Case studies

While numerous investigations and case studies have explored the diverse applications of VS and analyses through big data approaches, a comprehensive review of all such studies would be unwieldly. Therefore, this section highlights representative studies which were carefully selected to highlight the successes and limitations of VS as well as elucidating possible future directions.

4.1. *Integrating VS and experimental validation on a large scale*

A VS study was performed to identify novel bromodomain BRD4 inhibitors from a library of 2 million publicly-available compounds (the Enamine company collection) [163]. ML models, including Support Vector Machine (SVM) and Generative Topographic Mapping (GTM) models, were developed and employed in conjunction with structure-based pharmacophore models to screen the library. Molecular docking was subsequently used to rank the 12000 predicted hits to yield 3000 candidates. Out of the 3000

selected compounds, experimental testing of 2992 revealed that 29 were active which translated to a 1% hit rate.

The BRD4 study highlighted several issues facing VS studies, including the challenge of utilizing public databases due to their heterogeneous nature. The heterogenous nature of public databases stems from the fact that they are often compiled from different sources that use different protocols and measurements which makes it difficult to create a well-defined training set. For example, the weak to moderate correlation displayed by the actual hit detection criterion when compared to the dose-dependent public data affinity scores resulted in significantly impacting the success rate of the VS study.

4.2. *VS on a budget*

The above-mentioned strategy for identifying BRD4 inhibitors employed LBVS in conjunction with molecular docking to identify 3,000 potential hits which were subsequently subjected to experimental validation. Although this approach was effective, as demonstrated by its success, large-scale experimental validation is costly and often unfeasible for smaller laboratories or projects. Therefore, alternative methodologies are required for smaller-scale research endeavors. One such strategy is to prioritize computational validation methods to assess the predicted hits from VS. By identifying a promising hit through computational techniques, the need for extensive experimental validation is minimized. The hit selected using this approach can then undergo optimization using medicinal chemistry approaches, allowing for a more resource-efficient and focused development process.

Our group's attempt toward identifying novel and selective DDR1 inhibitors is a representative of this approach [96,131]. In a two-year effort to identify a selective DDR1 inhibitor, a multi-stage approach was employed. Initially, VS utilized a validated 3D-pharmacophore to assess 20 million drug-like compounds. HTVS then ranked the generated hits, with the top eight subjected to molecular docking and MM-GBSA calculations. This *in silico* analysis yielded three potential DDR1 inhibitors. Subsequent synthesis

and evaluation of these hits demonstrated good selectivity for DDR1, though the inhibitory activity for all three compounds was modest, with only 30% inhibition at 100 µM, considerably lower than that of many reported DDR1 inhibitors which exhibit inhibitory activity in the nanomolar range [131,164,165].

To reconcile the predicted versus observed activity, a second round of optimization in conjunction with a mechanistic study was conducted. This involved millisecond MD simulations and free energy calculations using the original hit compound which identified key structural features essential for further optimization. These insights guided the design and synthesis of more potent compounds which demonstrated improved DDR1 inhibition, with IC_{50} values in the nanomolar range (Figure 3). Despite the success of the DDR1 investigation in utilizing limited resources to accelerate the hit-to-lead pipeline, the process still required two years of effort to yield a single lead compound. Furthermore, the investigation concentrated all resources on optimizing a single scaffold, a strategy that carries significant risk due to the potential for failure stemming from poor pharmacokinetics and pharmacodynamics.

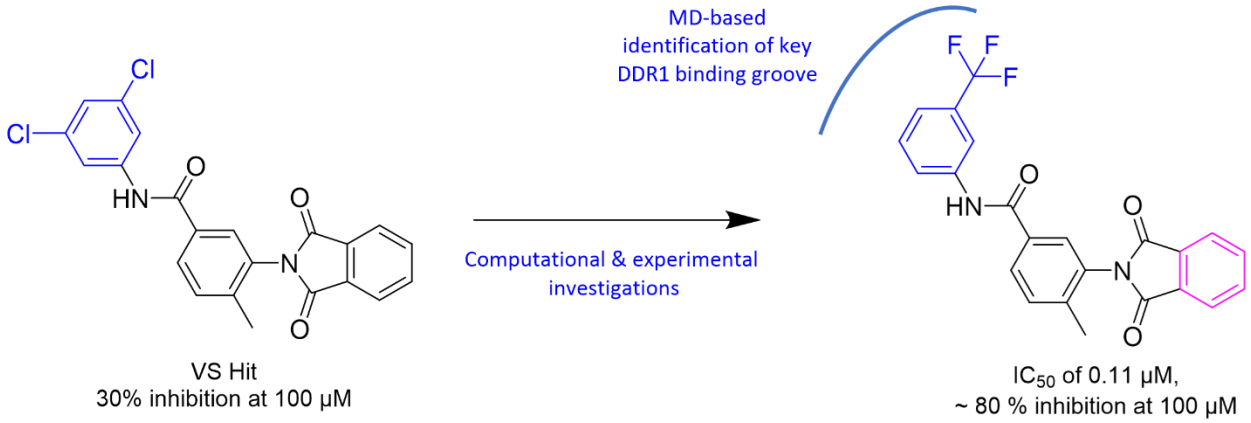

Figure 3. VS on a budget. Successful optimization of a virtual hit a successful lead

4.3. *The CACHE challenge*

A recent collaborative effort involving multiple scientific teams sought to evaluate the effectiveness of various computational methods in identifying drug leads for challenging targets, particularly those with limited or no known ligands. This large-scale initiative, known as "CACHE Challenge

#1," specifically focused on targeting the WD40 repeat (WDR) domain of the LRRK2 protein [166]. The primary goal of CACHE #1 was the identification of compounds capable of binding to the WDR domain of the LRRK2 protein which is believed to be involved in the progression of Parkinson's disease [167-169]. This domain was selected due to its unexplored potential in drug discovery and the absence of reported ligands. In addition to identifying LRRK2 WDR inhibitors, the study had aimed to benchmark different computational hit-finding methods and compare them to emerging ML approaches. This assessment was achieved by evaluating the performance of 23 different computational workflows submitted by participating research teams.

In the initial round of CACHE #1, 73 compounds exhibited measurable binding (K_D <150 µM and >30% binding as determined by surface plasmon resonance (SPR). However, issues related to compound behavior in solution emerged, where dynamic light scattering (DLS) revealed poor solubility for just over half (37) of the compounds. Moreover, orthogonal assays, such as differential scanning fluorimetry (DSF) and isothermal titration calorimetry (ITC), failed to confirm binding, with only two compounds showing binding to LRRK2-WDR in a ^{19}F-nuclear magnetic resonance (NMR) assay. One potential strategy to avoid predicted hits with poor solubility is to use screening methods or to train models specifically designed to identify such compounds. However, the effectiveness of this approach is often limited by the inherent unreliability of solubility predictions. These predictions typically require models to be trained on each chemical scaffold involved in the screening, which becomes impractical and unreliable when dealing with the vast chemical diversity found in millions of compounds. To avoid discarding potentially valuable hits all 73 compounds were advanced to the second round of the CACHE challenge for evaluation. In the second round, 61 compounds out of 714 analogues tested exhibited measurable K_D values (8.5% hit rate) and seven chemical series were confirmed as strong hits for LRRK2-WDR binding through multiple assays. Among these seven chemical series, the CACHE_1183 series yielded the most potent derivative (Figure 4).

Notably, some problematic compounds from Round 1 yielded promising analogues in Round 2 which stresses the importance of a two-round approach.

Analysis of the CACHE #1 study results revealed that both traditional methods and ML algorithms demonstrated comparable success rates in identifying novel LRRK2-WDR ligands. While some workflows incorporating ML were successful, their performance was not significantly superior to conventional methods which indicates that advanced computational techniques do not always guarantee improved outcomes. Moreover, the study highlighted the significant challenges in computational screening for targets with limited prior knowledge, such as the LRRK2-WDR domain. Another area of limitation was the ability of current VS workflows in predicting the solubility of small molecules which was highlighted by the high proportion of poorly soluble compounds among initial hits. It is also important to stress that the study's findings may not directly translate to other drug targets with different structural characteristics. Despite these limitations and challenges, CACHE #1 successfully identified seven chemical series as direct LRRK2-WDR binders, demonstrating the effectiveness of VS in identifying novel modulators.

Lastly, a key lesson that can be learned from the CACHE challenge is its iterative approach consisting of two rounds. The two rounds consist of a preliminary round which is used to identify potential hits which is followed by a more focused validation and optimization phase to refine the identified hits into promising candidates. This iterative process allows for a balance between exploring a wide range of chemical space and ensuring the quality and potential of identified compounds.

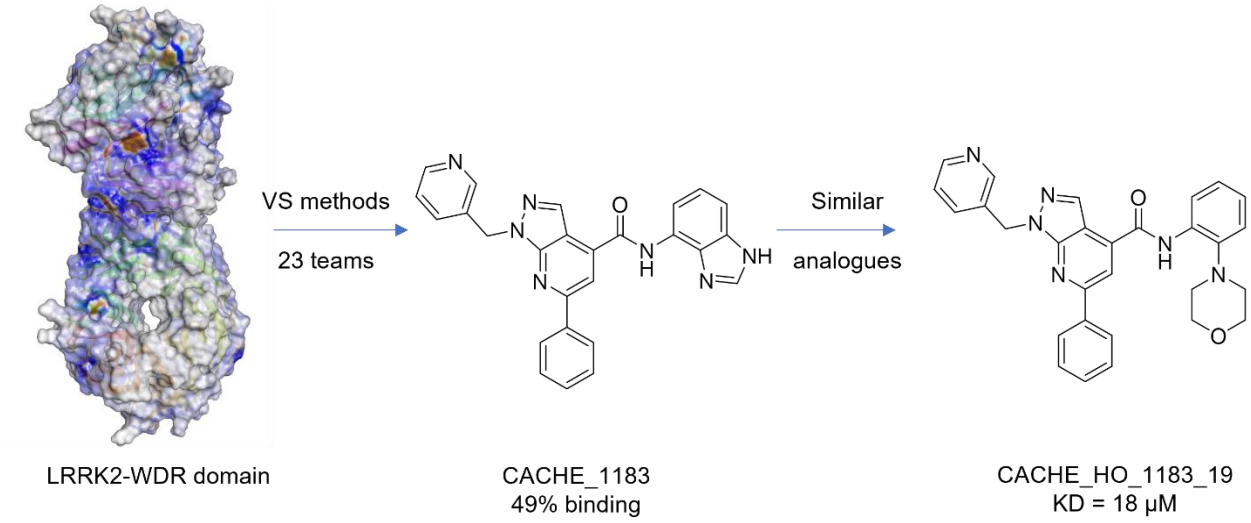

Figure 4. Overview of the CACHE challenge. VS of the LRRK2 protein (PDB: 6DLO [170]) was conducted by 23 teams from 10 different countries. The parent compound, CACHE_1183_25, demonstrated 49% binding activity at 200 µM. Subsequent testing of its analogues led to the identification of CACHE_HO_1183_19, with a dissociation constant (K_D) of 18 µM.

4.4. *RosettaVS*

RosettaVS is an SBVS platform that was developed recently using artificial intelligence (AI) algorithms [171]. Among the various AI-based platforms for VS, RosettaVS was chosen to be highlighted in this section as an example of how AI is changing the drug discovery field. Moreover, the open-source nature of RosettaVS combined with its reported promising results marks it as an ideal candidate to be highlighted. The RosettaVS platform is capable of efficiently screening multi-billion compound libraries which significantly accelerates virtual screening. What differentiates RosettaVS from other platforms is its open-source nature, which makes it a widely accessible tool for scientists worldwide.

To demonstrate the effectiveness of RosettaVS it was applied to two case studies: ubiquitin ligase KLHDC2 and the human voltage-gated sodium channel $Na_V1.7$. In the case of KLHDC2, the VS study yielded seven hit compounds, representing a 14% hit rate. The screening resulted in the discovery of compound C2.8 (Figure 5) which exhibited a potency of 1.1 µM. Meanwhile, VS for $Na_V1.7$ yielded four hits with a 44% hit rate. All hit compounds displayed binding affinities in the low micromolar range which underscores the

ability of RosettaVS to identify highly potent ligands [171]. Significantly, the entire screening process was completed in under seven days for both targets which highlights the platform's efficiency in high-throughput applications. An X-ray cocrystal structure generated for one of the KLHDC2-ligand complexes confirmed the platform's predicted binding pose. This validation showed a high degree of agreement between the predicted binding pose and the experimental structure and reinforces the reliability of RosettaVS predictions.

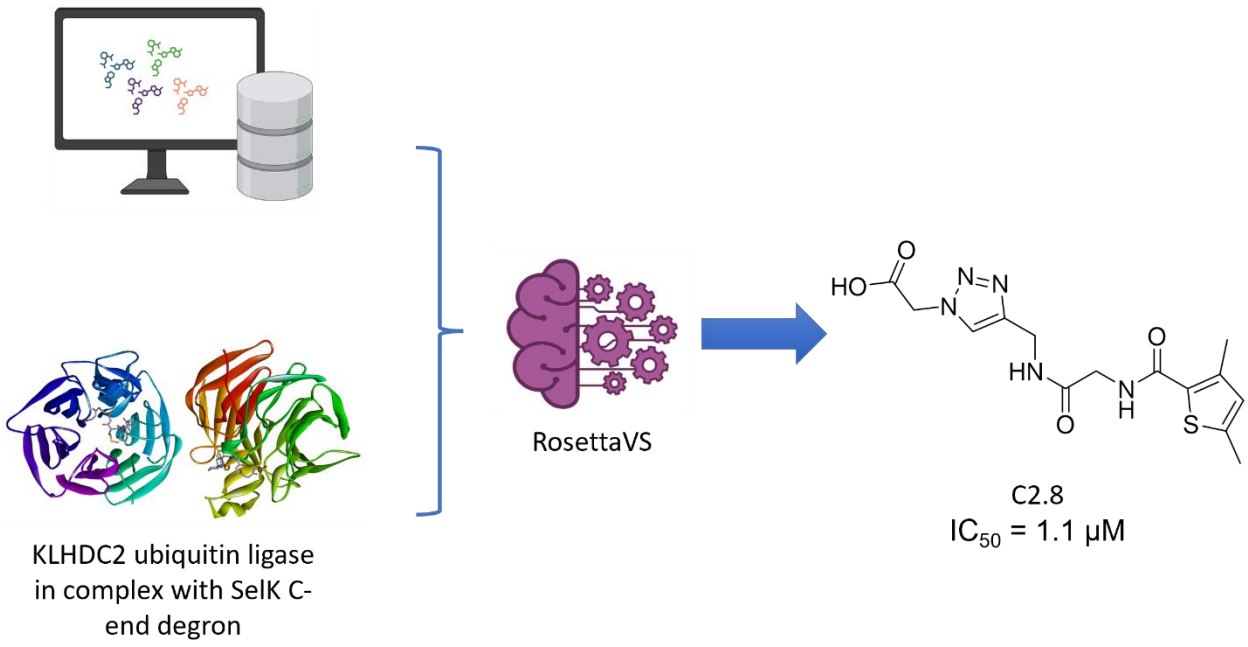

Figure 5. RosettaVS validation. VS of small molecule libraries over KLHDC2 (PDB: 6DO3 [172]) using RosettaVS yielded the best hit compound, compound C2.8 with IC_{50} of 1.1 µM.

Another key feature of RosettaVS is its ability to accommodate the flexibility of protein side chains and, to a lesser extent, the backbone. This is crucial, as proteins and ligands are inherently dynamic, existing in multiple conformations. Rigid docking, which assumes static structures, cannot capture this dynamic nature, potentially leading to inaccurate predictions. While flexible docking provides a more realistic representation of molecular interactions, it is computationally demanding making it less widely adopted.

As such, the ability of RosettaVS to account for protein flexibility when coupled with its open-source nature makes it a valuable tool for drug discovery [173,174].

It is important to note that RosettaVS is one of many platforms and applications developed for AI-based VS [175-178]. Given the mounting interest and the extensive body of work reported in this field, a comprehensive review of all AI-based VS methodologies would require an entire book. RosettaVS was highlighted in this work to exemplify the growing maturity and significance of AI in VS. Each AI-based platform and application possesses its own unique strengths and limitations. Consequently, the choice of platform largely depends on the specific requirements and objectives of the screening process which includes the size and nature of the chemical library, the desired level of accuracy, and the available computational resources. Another point to keep in mind is that while AI-based platforms like RosettaVS have the potential to be more powerful and faster than traditional LBVS tools, they also present their own set of challenges. For example, AI-based applications are often associated with higher computational costs and the complexity of installation due to the specific systems requirements for each ML model or package. Additionally, challenges such as overfitting of training data and the increased difficulty in interpreting the results and understanding the underlying rationale of AI models remain significant hurdles.

5. Conclusion

VS has become a tried-and-true technique of drug discovery due to the rapid advances in the computational and ML space over the last two decades. Although each VS workflow varies depending on the chosen approach and techniques. VS studies can be classified into retrospective and prospective studies based on the approach employed to evaluate the performance and outcome of a VS study. Retrospective VS studies are studies which assess the ability of the chosen methods to correctly identify known active compounds within a preselected class. Retrospective VS studies struggle to accurately judge

the performance of novel VS methods and often fail to allow for reproducible results. This failure is due to the lack the context to effectively compare findings across different studies.

Meanwhile, Prospective VS studies are studies involving computational screening of compounds to identify potential hits which is followed by experimental validation. However, simply identifying active compounds does not unequivocally validate a method. Demonstrating that simpler approaches were ineffective in identifying these hits is crucial for establishing the true value of a new method. Furthermore, the criteria for a compound to be considered a "hit" is not standardized across the different studies, leading to an unbalanced and biased evaluation. In conclusion, VS has matured into a widely adopted technique due to promising results and the significant potential for future growth.

6. Expert opinion

Despite the proven success and widespread use, VS still suffers from several limitations, misconceptions, and pitfalls. One of the hurdles facing the development of VS is a lack of standardized evaluation criteria on which to decide if a VS study is a success. Establishing a community-wide set of evaluation standards for published VS studies is essential in order to minimize false positives, promote reproducibility, and eliminate suboptimal practices. Another key challenge in the development of VS is that scientific publications in general are biased toward publishing positive results with a reluctance to publish negative results. The dearth of negative or inconclusive results in the literature creates a distorted view of the true success rates and limitations of the various VS approaches. This problem is further exacerbated by the fact that many of the industrial tools and applications of VS remain confidential. However, there is a promising shift in these practices, with many of the new ML-based applications and tools being shared as open-source thereby offering potential for the future advancement of the field.

While there are no guaranteed paths to success in VS, this section highlights key practices and concepts with the aim of providing a framework for understanding and overcoming some of the significant

challenges facing VS. One such a practice is that the success of a VS study hinges on careful and thorough preparation. Before carrying out a VS study, careful consideration must be given to database selection, preparation, and a detailed understanding of the intended target. Additionally, the optimal VS methodology should be chosen based on factors such as desired outcome, available computational resources, and planned validation strategies. Another key concept in the success of a VS study is that the size of screening libraries is a critical factor in VS. Expanding screening libraries to billions of compounds can significantly increase the chances of identifying potent, selective, and novel ligands.

Other practices require expertise and careful application. For instance, it is necessary to recognize that the success of any VS study is highly dependent on the knowledge and expertise of the practitioner. This is highlighted by the fact that the majority of VS studies include visual inspection (a highly subjective and personal skill) as part of their protocol. As such the same VS screening protocol could be carried out on the same library and yield different results. Additionally, a two-round approach in VS, as demonstrated in the CACHE #1 challenge, is advised. VS screening studies with two rounded approach involve a preliminary stage of broad screening to identify potential hits which is followed by a more focused validation and optimization phase to refine promising candidates. This iterative process allows for a balance between exploring a wide range of chemical space and ensuring the quality and potential of identified compounds.

An integrated approach which combines in silico and in vitro methods with access to vast on-demand chemical libraries is essential for efficient drug discovery. This stems from that fact that while HTS can effectively evaluate the biological activity of large libraries of physical compounds, its accuracy can be impacted by experimental variability. Conversely, VS screens a significantly larger portion of virtual chemical space but is constrained by the predictive accuracy of its scoring functions and models. Integrating VS with HTS can mitigate these limitations. Trained models on HTS outcomes can be utilized to enrich the chemical diversity of tested compounds. By prioritizing virtual compounds with a high probability of exhibiting biological activity, VS can guide HTS experiments and enable a more focused

evaluation of a diverse and targeted subset of physical compounds. This synergistic approach may expand the applicability of HTS while maximizing the strengths of VS. Ultimately, VS and HTS should be viewed as complementary tools in drug discovery rather than as mutually exclusive alternatives.

Lastly, there are numerous case studies and applications of VS which can serve as a guideline for an intended VS study. Therefore, the chosen approach should be carefully tailored to align with the desired outcomes, available expertise and to draw upon lessons from both the successes and failures of case studies with similar applications. In 2018 Dean Brown noted that there are no reported examples of clinical candidates emerging from a virtual screening [179]. However, in 2022 the optimization of a virtual screening hit resulted in the discovery of the clinical candidate RP-6306 which is an orally bioavailable and selective PKMYT1 inhibitor.[180] RP-6303 represents the value of VS and reflect on the value of utilizing previous studies and results to "stand on the shoulders of giants,"

Funding

This research was supported by the National Institute on Aging of the National Institutes of Health.

Declaration of Interest:

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Reviewer Disclosures:

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

References

[1] Walters WP, Wang R. New Trends in Virtual Screening. Journal of Chemical Information and Modeling. 2020;60(9):4109-4111.

[2] Tanrikulu Y, Krüger B, Proschak E. The holistic integration of virtual screening in drug discovery. Drug Discovery Today. 2013;18(7):358-364.

[3] Lavecchia A, Di Giovanni C. Virtual screening strategies in drug discovery: a critical review. Current medicinal chemistry. 2013;20(23):2839-60.

[4] Slater O, Kontoyianni M. The compromise of virtual screening and its impact on drug discovery. Expert opinion on drug discovery. 2019;14(7):619-637.

[5] Zhang SY, Zhao J, Ni JJ, et al. Application and prospects of high-throughput screening for in vitro neurogenesis. World journal of stem cells. 2022;14(6):393-419.

[6] Gorgulla C. Recent Developments in Ultralarge and Structure-Based Virtual Screening Approaches. 2023;6(Volume 6, 2023):229-258.

[7] Warr WA, Nicklaus MC, Nicolaou CA, et al. Exploration of Ultralarge Compound Collections for Drug Discovery. Journal of Chemical Information and Modeling. 2022;62(9):2021-2034.

[8] Dunn TB, Seabra GM, Kim TD, et al. Diversity and Chemical Library Networks of Large Data Sets. Journal of Chemical Information and Modeling. 2022;62(9):2186-2201.

[9] Scior T, Bender A, Tresadern G, et al. Recognizing Pitfalls in Virtual Screening: A Critical Review. Journal of Chemical Information and Modeling. 2012;52(4):867-881. ** A comprehensive review of pitfalls when carrying out VS

[10] Lyne PD. Structure-based virtual screening: an overview. Drug Discovery Today. 2002;7(20):1047-1055.

[11] Bajorath J. Integration of virtual and high-throughput screening. Nature Reviews Drug Discovery. 2002;1(11):882-894. * Stresses the importance of integrating HTS and VS and the different ways for their integration.

[12] Banegas-Luna A-J, Cerón-Carrasco JP, Pérez-Sánchez H. A Review of Ligand-Based Virtual Screening Web Tools and Screening Algorithms in Large Molecular Databases in the Age of Big Data. Future Medicinal Chemistry. 2018;10(22):2641-2658.

[13] Yang S-Q, Ye Q, Ding J-J, et al. Current advances in ligand-based target prediction. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2021;11(3):e1504.

[14] Bhunia SS, Saxena M, Saxena AK. Ligand- and Structure-Based Virtual Screening in Drug Discovery. In: Saxena AK, editor. Biophysical and Computational Tools in Drug Discovery. Cham: Springer International Publishing; 2021. p. 281-339.

[15] Wilson GL, Lill MA. Integrating Structure-Based and Ligand-Based Approaches for Computational Drug Design. Future Medicinal Chemistry. 2011;3(6):735-750.

[16] Sadybekov AV, Katritch V. Computational approaches streamlining drug discovery. Nature. 2023;616(7958):673-685.

[17] Yan XC, Sanders JM, Gao Y-D, et al. Augmenting Hit Identification by Virtual Screening Techniques in Small Molecule Drug Discovery. Journal of Chemical Information and Modeling. 2020;60(9):4144-4152.

[18] Kitchen DB, Decornez H, Furr JR, et al. Docking and scoring in virtual screening for drug discovery: methods and applications. Nature Reviews Drug Discovery. 2004;3(11):935-949.

[19] Kontoyianni M. Docking and Virtual Screening in Drug Discovery. In: Lazar IM, Kontoyianni M, Lazar AC, editors. Proteomics for Drug Discovery: Methods and Protocols. New York, NY: Springer New York; 2017. p. 255-266.

[20] Knox AJS, Meegan MJ, Carta G, et al. Considerations in Compound Database Preparation“Hidden” Impact on Virtual Screening Results. Journal of Chemical Information and Modeling. 2005;45(6):1908-1919.

[21] Vuorinen A, Schuster D. Methods for generating and applying pharmacophore models as virtual screening filters and for bioactivity profiling. Methods. 2015;71:113-134.

[22] Wang R, Lu Y, Wang S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. Journal of Medicinal Chemistry. 2003;46(12):2287-2303.

[23] Kastritis PL, Bonvin AMJJ. Are Scoring Functions in Protein−Protein Docking Ready To Predict Interactomes? Clues from a Novel Binding Affinity Benchmark. Journal of Proteome Research. 2010;9(5):2216-2225.

[24] Halperin I, Ma B, Wolfson H, et al. Principles of docking: An overview of search algorithms and a guide to scoring functions. Proteins: Structure, Function, and Bioinformatics. 2002;47(4):409-443.

[25] Li J, Fu A, Zhang L. An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking. Interdisciplinary Sciences: Computational Life Sciences. 2019;11(2):320-328.

[26] Li Q. Chapter 4 - Virtual screening of small-molecule libraries. In: Trabocchi A, Lenci E, editors. Small Molecule Drug Discovery: Elsevier; 2020. p. 103-125.

[27] Irwin JJ. How good is your screening library? Current Opinion in Chemical Biology. 2006;10(4):352-356.

[28] Bender A, Glen RC. Molecular similarity: a key technique in molecular informatics [10.1039/B409813G]. Organic & Biomolecular Chemistry. 2004;2(22):3204-3218.

[29] Lumley JA. Compound Selection and Filtering in Library Design. 2005;24(9):1066-1075.

[30] Garcia-Castro M, Zimmermann S, Sankar MG, et al. Scaffold Diversity Synthesis and Its Application in Probe and Drug Discovery. Angewandte Chemie International Edition. 2016;55(27):7586-7605.

[31] Karthikeyan M, Vyas R. Role of Open Source Tools and Resources in Virtual Screening for Drug Discovery. Combinatorial Chemistry & High Throughput Screening. 2015;18(6):528-543.

[32] Subramaniam S, Mehrotra M, Gupta D. Virtual high throughput screening (vHTS)--a perspective. Bioinformation. 2008;3(1):14-7.

[33] Xie X-Q, Chen J-Z. Data Mining a Small Molecule Drug Screening Representative Subset from NIH PubChem. Journal of Chemical Information and Modeling. 2008;48(3):465-475.

[34] Waldman M, Li H, Hassan M. Novel algorithms for the optimization of molecular diversity of combinatorial libraries. Journal of Molecular Graphics and Modelling. 2000;18(4):412-426.

[35] Tingle BI, Tang KG, Castanon M, et al. ZINC-22—A Free Multi-Billion-Scale Database of Tangible Compounds for Ligand Discovery. Journal of Chemical Information and Modeling. 2023;63(4):1166-1176.

[36] Chen J, Swamidass SJ, Dou Y, et al. ChemDB: a public database of small molecules and related chemoinformatics resources. Bioinformatics. 2005;21(22):4133-4139.

[37] Kim S, Chen J, Cheng T, et al. PubChem 2019 update: improved access to chemical data. Nucleic Acids Research. 2018;47(D1):D1102-D1109.

[38] Bilimoria KY, Stewart AK, Winchester DP, et al. The National Cancer Data Base: A Powerful Initiative to Improve Cancer Care in the United States. Annals of Surgical Oncology. 2008;15(3):683-690.

[39] Knox C, Wilson M, Klinger Christen M, et al. DrugBank 6.0: the DrugBank Knowledgebase for 2024. Nucleic Acids Research. 2023;52(D1):D1265-D1275.

[40] Klebe G. Virtual ligand screening: strategies, perspectives and limitations. Drug Discovery Today. 2006;11(13):580-594.

[41] Verdonk ML, Berdini V, Hartshorn MJ, et al. Virtual Screening Using Protein−Ligand Docking: Avoiding Artificial Enrichment. Journal of Chemical Information and Computer Sciences. 2004;44(3):793-806.

[42] Congreve M, Carr R, Murray C, et al. A 'Rule of Three' for fragment-based lead discovery? Drug Discovery Today. 2003;8(19):876-877.

[43] Pollastri MP. Overview on the Rule of Five. Current Protocols in Pharmacology. 2010;49(1):9.12.1-9.12.8.

[44] Lee J, Myeong IS, Kim Y. The Drug-Like Molecule Pre-Training Strategy for Drug Discovery. IEEE Access. 2023;11:61680-61687.

[45] Hughes JD, Blagg J, Price DA, et al. Physiochemical drug properties associated with in vivo toxicological outcomes. Bioorganic & Medicinal Chemistry Letters. 2008;18(17):4872-4875.

[46] Jorgensen WL. Efficient Drug Lead Discovery and Optimization. Accounts of Chemical Research. 2009;42(6):724-733.

[47] Lionta E, Spyrou G, K. Vassilatis D, et al. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. Current Topics in Medicinal Chemistry. 2014;14(16):1923-1938.

[48] Wass MN, Kelley LA, Sternberg MJE. 3DLigandSite: predicting ligand-binding sites using similar structures. Nucleic Acids Research. 2010;38(suppl_2):W469-W473.

[49] Krivák R, Hoksza D. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. Journal of Cheminformatics. 2018;10(1):39.

[50] Kairys V, Baranauskiene L, Kazlauskiene M, et al. Binding affinity in drug design: experimental and computational techniques. Expert opinion on drug discovery. 2019;14(8):755-768. ** Comprehensive review of binding affinity determination

[51] Khoshbin Z, Housaindokht MR, Izadyar M, et al. Recent advances in computational methods for biosensor design. Biotechnology and Bioengineering. 2021;118(2):555-578.

[52] Yadav MR, Murumkar PR, Yadav R, et al. Chapter 3 - Structure-based virtual screening in drug discovery. In: Roy K, editor. Cheminformatics, QSAR and Machine Learning Applications for Novel Drug Development: Academic Press; 2023. p. 69-88.

[53] Spyrakis F, Cavasotto CN. Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration and active site water molecules description. Archives of Biochemistry and Biophysics. 2015;583:105-119. * Highlights critical issues in SBVS

[54] Wei G, Xi W, Nussinov R, et al. Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell. Chemical Reviews. 2016;116(11):6516-6551.

[55] Antunes DA, Devaurs D, Kavraki LE. Understanding the challenges of protein flexibility in drug design. Expert opinion on drug discovery. 2015;10(12):1301-1313. * Summarize the challenges associated with protein flexibility

[56] Atilgan C. Chapter Two - Computational Methods for Efficient Sampling of Protein Landscapes and Disclosing Allosteric Regions. In: Karabencheva-Christova TG, Christov CZ, editors. Advances in Protein Chemistry and Structural Biology. Vol. 113: Academic Press; 2018. p. 33-63.

[57] Peach ML, Cachau RE, Nicklaus MC. Conformational energy range of ligands in protein crystal structures: The difficult quest for accurate understanding. Journal of Molecular Recognition. 2017;30(8):e2618.

[58] Deller MC, Rupp B. Models of protein–ligand crystal structures: trust, but verify. Journal of Computer-Aided Molecular Design. 2015;29(9):817-836.

[59] Dekker C, Goepfert A. An Industrial Perspective on Protein–Ligand Complex Crystallization. Structural Biology in Drug Discovery2020. p. 165-185.

[60] Ehrmann FR, Stojko J, Metz A, et al. Soaking suggests “alternative facts”: Only co-crystallization discloses major ligand-induced interface rearrangements of a homodimeric tRNA-binding protein indicating a novel mode-of-inhibition. PLOS ONE. 2017;12(4):e0175723.

[61] Renaud J-P, Chari A, Ciferri C, et al. Cryo-EM in drug discovery: achievements, limitations and prospects. Nature Reviews Drug Discovery. 2018;17(7):471-492. ** Comprehensively discusses the differences between Cryo-EM and X-ray crystallography and their applications

[62] Goodsell DS, Sanner MF, Olson AJ, et al. The AutoDock suite at 30. 2021;30(1):31-43.

[63] Friesner RA, Banks JL, Murphy RB, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. Journal of Medicinal Chemistry. 2004;47(7):1739-1749.

[64] Verdonk ML, Cole JC, Hartshorn MJ, et al. Improved protein–ligand docking using GOLD. Proteins: Structure, Function, and Bioinformatics. 2003;52(4):609-623.

[65] Evans ME, Graham LD. A Flexible Lunar Architecture for Exploration (FLARE) supporting NASA's Artemis Program. Acta Astronautica. 2020;177:351-372.

[66] Venkatachalam CM, Jiang X, Oldfield T, et al. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. Journal of Molecular Graphics and Modelling. 2003;21(4):289-307.

[67] Ghosh S, Nie A, An J, et al. Structure-based virtual screening of chemical libraries for drug discovery. Current Opinion in Chemical Biology. 2006;10(3):194-202.

[68] Stanzione F, Giangreco I, Cole JC. Chapter Four - Use of molecular docking computational tools in drug discovery. In: Witty DR, Cox B, editors. Progress in Medicinal Chemistry. Vol. 60: Elsevier; 2021. p. 273-343.

[69] Zheng L, Meng J, Jiang K, et al. Improving protein–ligand docking and screening accuracies by incorporating a scoring function correction term. Briefings in Bioinformatics. 2022;23(3).

[70] Gaurav A, Gautam V. Structure-based three-dimensional pharmacophores as an alternative to traditional methodologies. Journal of Receptor, Ligand and Channel Research. 2014;7(null):27-38.

[71] Chen Y-C. Beware of docking! Trends in Pharmacological Sciences. 2015;36(2):78-95. * Limitations of docking and key practices to overcome these limitations

[72] Ashtawy HM, Mahapatra NR. A Comparative Assessment of Ranking Accuracies of Conventional and Machine-Learning-Based Scoring Functions for Protein-Ligand Binding Affinity Prediction. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2012;9(5):1301-1313.

[73] Saikia S, Bordoloi M. Molecular Docking: Challenges, Advances and its Use in Drug Discovery Perspective. Current Drug Targets. 2019;20(5):501-521.

[74] Wójcikowski M, Ballester PJ, Siedlecki P. Performance of machine-learning scoring functions in structure-based virtual screening. Scientific Reports. 2017;7(1):46710.

[75] Sieg J, Flachsenberg F, Rarey M. In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening. Journal of Chemical Information and Modeling. 2019;59(3):947-961.

[76] Irwin JJ. Community benchmarks for virtual screening. Journal of Computer-Aided Molecular Design. 2008;22(3):193-199.

[77] Morrone JA, Weber JK, Huynh T, et al. Combining Docking Pose Rank and Structure with Deep Learning Improves Protein–Ligand Binding Mode Prediction over a Baseline Docking Approach. Journal of Chemical Information and Modeling. 2020;60(9):4170-4179.

[78] Houston DR, Walkinshaw MD. Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context. Journal of Chemical Information and Modeling. 2013;53(2):384-390. * Importance of Consensus docking

[79] Bajusz D, Keserű GM. Maximizing the integration of virtual and experimental screening in hit discovery. Expert opinion on drug discovery. 2022;17(6):629-640.

[80] Lopez-Vallejo F, Caulfield T, Martinez-Mayorga K, et al. Integrating Virtual Screening and Combinatorial Chemistry for Accelerated Drug Discovery. Combinatorial Chemistry & High Throughput Screening. 2011;14(6):475-487.

[81] Roy K, Mitra I. On Various Metrics Used for Validation of Predictive QSAR Models with Applications in Virtual Screening and Focused Library Design. Combinatorial Chemistry & High Throughput Screening. 2011;14(6):450-474.

[82] Federico LB, Barcelos MP, Silva GM, et al. Key Aspects for Achieving Hits by Virtual Screening Studies. In: La Porta FA, Taft CA, editors. Functional Properties of Advanced Engineering Materials and Biomolecules. Cham: Springer International Publishing; 2021. p. 455-487.

[83] Hurtle BT, Jana S, Cai L, et al. Ligand-Based Virtual Screening as a Path to New Chemotypes for Candidate PET Radioligands for Imaging Tauopathies. Journal of Medicinal Chemistry. 2024;67(16):14095-14109.

[84] Sarkar B, Chakraborty S, Rakshit G, et al. Chapter 12 - Fundamental approaches of drug discovery. In: Rudrapal M, Egbuna C, Cho WC, editors. Biochemical and Molecular Pharmacology in Drug Discovery: Elsevier; 2024. p. 251-282.

[85] Oprea TI, Matter H. Integrating virtual screening in lead discovery. Current Opinion in Chemical Biology. 2004;8(4):349-358.

[86] Hamza A, Wei N-N, Zhan C-G. Ligand-Based Virtual Screening Approach Using a New Scoring Function. Journal of Chemical Information and Modeling. 2012;52(4):963-974.

[87] Durai P, Ko Y-J, Pan C-H, et al. Evolutionary chemical binding similarity approach integrated with 3D-QSAR method for effective virtual screening. BMC Bioinformatics. 2020;21(1):309.

[88] Krüger DM, Evers A. Comparison of Structure- and Ligand-Based Virtual Screening Protocols Considering Hit List Complementarity and Enrichment Factors. ChemMedChem. 2010;5(1):148-158.

[89] Stahura FL, Bajorath J. New Methodologies for Ligand-Based Virtual Screening. Current Pharmaceutical Design. 2005;11(9):1189-1202.

[90] Himmat M, Salim N, Al-Dabbagh MM, et al. Adapting Document Similarity Measures for Ligand-Based Virtual Screening. Molecules. 2016;21(4):476.

[91] Grimm M, Liu Y, Yang X, et al. LigMate: A Multifeature Integration Algorithm for Ligand-Similarity-Based Virtual Screening. Journal of Chemical Information and Modeling. 2020;60(12):6044-6053.

[92] Similarity Methods. An Introduction To Chemoinformatics. Dordrecht: Springer Netherlands; 2007. p. 99-117.

[93] Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? Journal of Cheminformatics. 2015;7(1):20.

[94] Maggiora GM. Introduction to Molecular Similarity and Chemical Space. In: Martinez-Mayorga K, Medina-Franco JL, editors. Foodinformatics: Applications of Chemical Information to Food Chemistry. Cham: Springer International Publishing; 2014. p. 1-81.

[95] von Korff M, Freyss J, Sander T. Flexophore, a New Versatile 3D Pharmacophore Descriptor That Considers Molecular Flexibility. Journal of Chemical Information and Modeling. 2008;48(4):797-810.

[96] Nada H, Lee K, Gotina L, et al. Identification of novel discoidin domain receptor 1 (DDR1) inhibitors using E-pharmacophore modeling, structure-based virtual screening, molecular dynamics simulation and MM-GBSA approaches. Computers in Biology and Medicine. 2022;142:105217.

[97] Nicholls A, McGaughey GB, Sheridan RP, et al. Molecular Shape and Medicinal Chemistry: A Perspective. Journal of Medicinal Chemistry. 2010;53(10):3862-3886.

[98] Lee HS, Lee CS, Kim JS, et al. Improving Virtual Screening Performance against Conformational Variations of Receptors by Shape Matching with Ligand Binding Pocket. Journal of Chemical Information and Modeling. 2009;49(11):2419-2428.

[99] Cheeseright TJ, Mackey MD, Melville JL, et al. FieldScreen: Virtual Screening Using Molecular Fields. Application to the DUD Data Set. Journal of Chemical Information and Modeling. 2008;48(11):2108-2117.

[100] Berrhail F, Belhadef H, Haddad M. Deep Convolutional Neural Network to improve the performances of screening process in LBVS. Expert Systems with Applications. 2022;203:117287.

[101] Emonts J, Buyel JF. An overview of descriptors to capture protein properties – Tools and perspectives in the context of QSAR modeling. Computational and Structural Biotechnology Journal. 2023;21:3234-3247.

[102] Patel HM, Noolvi MN, Sharma P, et al. Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery. Medicinal Chemistry Research. 2014;23(12):4991-5007. * Comprehensive review of QSAR

[103] Cherkasov A, Muratov EN, Fourches D, et al. QSAR Modeling: Where Have You Been? Where Are You Going To? Journal of Medicinal Chemistry. 2014;57(12):4977-5010.

[104] Tropsha A, Isayev O, Varnek A, et al. Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR. Nature Reviews Drug Discovery. 2024;23(2):141-155.

[105] Martin YC, Kofron JL, Traphagen LM. Do Structurally Similar Molecules Have Similar Biological Activity? Journal of Medicinal Chemistry. 2002;45(19):4350-4358.

[106] Willett P. Similarity-based virtual screening using 2D fingerprints. Drug Discovery Today. 2006;11(23):1046-1053.

[107] Muegge I, Martin YC, Hajduk PJ, et al. Evaluation of PMF Scoring in Docking Weak Ligands to the FK506 Binding Protein. Journal of Medicinal Chemistry. 1999;42(14):2498-2503.

[108] Willett P. The Calculation of Molecular Structural Similarity: Principles and Practice. Molecular informatics. 2014;33(6-7):403-413.

[109] Geppert H, Bajorath J. Advances in 2D fingerprint similarity searching. Expert opinion on drug discovery. 2010;5(6):529-542.

[110] Willett P. Similarity Searching Using 2D Structural Fingerprints. In: Bajorath J, editor. Chemoinformatics and Computational Chemical Biology. Totowa, NJ: Humana Press; 2011. p. 133-158.

[111] Lavecchia A. Machine-learning approaches in drug discovery: methods and applications. Drug Discovery Today. 2015;20(3):318-331.

[112] Rahmani AM, Yousefpoor E, Yousefpoor MS, et al. Machine Learning (ML) in Medicine: Review, Applications, and Challenges. Mathematics. 2021;9(22):2970.

[113] Davila Delgado JM, Oyedele L. Deep learning with small datasets: using autoencoders to address limited datasets in construction management. Applied Soft Computing. 2021;112:107836.

[114] Carpenter KA, Huang X. Machine Learning-based Virtual Screening and Its Applications to Alzheimer's Drug Discovery: A Review. Current Pharmaceutical Design. 2018;24(28):3347-3358.

[115] Hatcher WG, Yu W. A Survey of Deep Learning: Platforms, Applications and Emerging Research Trends. IEEE Access. 2018;6:24411-24432.

[116] Tran-Nguyen V-K, Junaid M, Simeon S, et al. A practical guide to machine-learning scoring for structure-based virtual screening. Nature Protocols. 2023;18(11):3460-3511. ** Comprehensive discussion on ML integration with VS

[117] Zoete V, Daina A, Bovigny C, et al. SwissSimilarity: A Web Tool for Low to Ultra High Throughput Ligand-Based Virtual Screening. Journal of Chemical Information and Modeling. 2016;56(8):1399-1404.

[118] Wang L, Ma C, Wipf P, et al. TargetHunter: An In Silico Target Identification Tool for Predicting Therapeutic Potential of Small Organic Molecules Based on Chemogenomic Database. The AAPS Journal. 2013;15(2):395-406.

[119] Nada H, Kim S, Lee K. PT-Finder: A multi-modal neural network approach to target identification. Computers in Biology and Medicine. 2024;174:108444.

[120] Du X, Li Y, Xia Y-L, et al. Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. International Journal of Molecular Sciences. 2016;17(2):144.

[121] Mortier J, Rakers C, Bermudez M, et al. The impact of molecular dynamics on drug design: applications for the characterization of ligand–macromolecule complexes. Drug Discovery Today. 2015;20(6):686-702.

[122] Okimoto N, Futatsugi N, Fuji H, et al. High-Performance Drug Discovery: Computational Screening by Combining Docking and Molecular Dynamics Simulations. PLOS Computational Biology. 2009;5(10):e1000528.

[123] Liu X, Shi D, Zhou S, et al. Molecular dynamics simulations and novel drug discovery. Expert opinion on drug discovery. 2018;13(1):23-37.

[124] Nada H, Elkamhawy A, Lee K. Identification of 1H-purine-2,6-dione derivative as a potential SARS-CoV-2 main protease inhibitor: molecular docking, dynamic simulations, and energy calculations. PeerJ. 2022;10:e14120.

[125] Kumari I, Sandhu P, Ahmed M, et al. Molecular Dynamics Simulations, Challenges and Opportunities: A Biologist's Prospective. Current Protein and Peptide Science. 2017;18(11):1163-1179.

[126] Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. Medicinal Research Reviews. 2006;26(5):531-568. * Detailed discussion of MD integration with docking

[127] Katiyar RS, Jha PK. Molecular simulations in drug delivery: Opportunities and challenges. WIREs Computational Molecular Science. 2018;8(4):e1358.

[128] Abel R, Wang L, Harder ED, et al. Advancing Drug Discovery through Enhanced Free Energy Calculations. Accounts of Chemical Research. 2017;50(7):1625-1632.

[129] Yan S, Peck JM, Ilgu M, et al. Sampling Performance of Multiple Independent Molecular Dynamics Simulations of an RNA Aptamer. ACS Omega. 2020;5(32):20187-20201.

[130] Guo X, Yan T, Rao J, et al. Novel Feleucin-K3-Derived Peptides Modified with Sulfono-γ-AA Building Blocks Targeting Pseudomonas aeruginosa and Methicillin-Resistant Staphylococcus aureus Infections. Journal of Medicinal Chemistry. 2023;66(2):1254-1272.

[131] Nada H, Kim S, Jaemin C, et al. From pixels to druggable leads: A CADD strategy for the design and synthesis of potent DDR1 inhibitors. Computer Methods and Programs in Biomedicine. 2024;254:108318. * Example of VS application with a budget and its limitations

[132] Mahdizadeh SJ, Stier M, Carlesso A, et al. Multiscale In Silico Study of the Mechanism of Activation of the RtcB Ligase by the PTP1B Phosphatase. Journal of Chemical Information and Modeling. 2024;64(3):905-917.

[133] Reddy MR, Reddy CR, Rathore RS, et al. Free energy calculations to estimate ligand-binding affinities in structure-based drug design. Current Pharmaceutical Design. 2014;20(20):3323-37.

[134] Malmstrom RD, Watowich SJ. Using Free Energy of Binding Calculations To Improve the Accuracy of Virtual Screening Predictions. Journal of Chemical Information and Modeling. 2011;51(7):1648-1655.

[135] Wang L, Wu Y, Deng Y, et al. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. Journal of the American Chemical Society. 2015;137(7):2695-2703.

[136] Kar S, Roy K. How far can virtual screening take us in drug discovery? Expert opinion on drug discovery. 2013;8(3):245-261.

[137] Stumpfe D, Bajorath J. Current Trends, Overlooked Issues, and Unmet Challenges in Virtual Screening. Journal of Chemical Information and Modeling. 2020;60(9):4112-4115.

[138] Chen L, Cruz A, Ramsey S, et al. Hidden bias in the DUD-E dataset leads to misleading performance of deep learning in structure-based virtual screening. PLOS ONE. 2019;14(8):e0220113.

[139] Baell JB, Holloway GA. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. Journal of Medicinal Chemistry. 2010;53(7):2719-2740.

[140] Lipinski CA, Lombardo F, Dominy BW, et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews. 2001;46(1-3):3-26.

[141] Veber DF, Johnson SR, Cheng HY, et al. Molecular properties that influence the oral bioavailability of drug candidates. Journal of Medicinal Chemistry. 2002;45(12):2615-23.

[142] Ghose AK, Viswanadhan VN, Wendoloski JJ. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. J Comb Chem. 1999;1(1):55-68.

[143] Egan WJ, Merz KM, Baldwin JJ. Prediction of Drug Absorption Using Multivariate Statistics. Journal of Medicinal Chemistry. 2000;43(21):3867-3877.

[144] Muegge I. Selection criteria for drug-like compounds. Medicinal Research Reviews. 2003;23(3):302-321.

[145] Conchillo-Solé O, de Groot NS, Avilés FX, et al. AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC Bioinformatics. 2007;8(1):65.

[146] Kim S, Thiessen PA, Bolton EE, et al. PubChem Substance and Compound databases. Nucleic Acids Research. 2015;44(D1):D1202-D1213. * Key databases for drug discovery

[147] Pence HE, Williams A. ChemSpider: An Online Chemical Information Resource. Journal of Chemical Education. 2010;87(11):1123-1124. *

[148] Brooks WH, Guida WC, Daniel KG. The significance of chirality in drug design and development. Current Topics in Medicinal Chemistry. 2011;11(7):760-70.

[149] Brooks WH, Daniel KG, Sung S-S, et al. Computational Validation of the Importance of Absolute Stereochemistry in Virtual Screening. Journal of Chemical Information and Modeling. 2008;48(3):639-645.

[150] Guasch L, Yapamudiyansel W, Peach ML, et al. Experimental and Chemoinformatics Study of Tautomerism in a Database of Commercially Available Screening Samples. Journal of Chemical Information and Modeling. 2016;56(11):2149-2161.

[151] Varela-Rial A, Majewski M, De Fabritiis G. Structure based virtual screening: Fast and slow. 2022;12(2):e1544.

[152] van der Westhuizen CJ, Stander A, Riley DL, et al. Discovery of Novel Acetylcholinesterase Inhibitors by Virtual Screening, In Vitro Screening, and Molecular Dynamics Simulations. Journal of Chemical Information and Modeling. 2022;62(6):1550-1572.

[153] Zhang L, Fourches D, Sedykh A, et al. Discovery of Novel Antimalarial Compounds Enabled by QSAR-Based Virtual Screening. Journal of Chemical Information and Modeling. 2013;53(2):475-492.

[154] Chen H, Lyne PD, Giordanetto F, et al. On Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors. Journal of Chemical Information and Modeling. 2006;46(1):401-415.

[155] Warren GL, Andrews CW, Capelli A-M, et al. A Critical Assessment of Docking Programs and Scoring Functions. Journal of Medicinal Chemistry. 2006;49(20):5912-5931.

[156] Boittier ED, Tang YY, Buckley ME, et al. Assessing Molecular Docking Tools to Guide Targeted Drug Discovery of CD38 Inhibitors. International Journal of Molecular Sciences. 2020;21(15):5183.

[157] Shen C, Wang Z, Yao X, et al. Comprehensive assessment of nine docking programs on type II kinase inhibitors: prediction accuracy of sampling power, scoring power and screening power. Briefings in Bioinformatics. 2018;21(1):282-297.

[158] Masters L, Eagon S, Heying M. Evaluation of consensus scoring methods for AutoDock Vina, smina and idock. Journal of Molecular Graphics and Modelling. 2020;96:107532.

[159] Olubiyi OO, Samantray S, Illig A-M. Chapter 24 - Advances in structure-based virtual screening for drug discovery. In: Tripathi T, Dubey VK, editors. Advances in Protein Molecular and Structural Biology Methods: Academic Press; 2022. p. 387-404.

[160] Cavasotto CN, W. Orry AJ. Ligand Docking and Structure-based Virtual Screening in Drug Discovery. Current Topics in Medicinal Chemistry. 2007;7(10):1006-1014.

[161] Swann SL, Brown SP, Muchmore SW, et al. A Unified, Probabilistic Framework for Structure- and Ligand-Based Virtual Screening. Journal of Medicinal Chemistry. 2011;54(5):1223-1232.

[162] Tuerkova A, Bongers BJ, Norinder U, et al. Identifying Novel Inhibitors for Hepatic Organic Anion Transporting Polypeptides by Machine Learning-Based Virtual Screening. Journal of Chemical Information and Modeling. 2022;62(24):6323-6335.

[163] Casciuc I, Horvath D, Gryniukova A, et al. Pros and cons of virtual screening based on public “Big Data”: In silico mining for new bromodomain inhibitors. European Journal of Medicinal Chemistry. 2019;165:258-272. * Example of the integrating VS and experimental validation on a large scale

[164] Yoshimori A, Asawa Y, Kawasaki E, et al. Design and Synthesis of DDR1 Inhibitors with a Desired Pharmacophore Using Deep Generative Models. ChemMedChem. 2021;16(6):955-958.

[165] Tan X, Li C, Yang R, et al. Discovery of Pyrazolo[3,4-d]pyridazinone Derivatives as Selective DDR1 Inhibitors via Deep Learning Based Design, Synthesis, and Biological Evaluation. Journal of Medicinal Chemistry. 2022;65(1):103-119.

[166] Li F, Ackloo S, Arrowsmith CH, et al. CACHE Challenge #1: Targeting the WDR Domain of LRRK2, A Parkinson's Disease Associated Protein. Journal of Chemical Information and Modeling. 2024;64(22):8521-8536. ** To the best of our knowledge, this is the first competition of this scale to enable a comprehensive and prospective analysis of VS.

[167] Lin X, Parisiadou L, Gu X-L, et al. Leucine-rich repeat kinase 2 regulates the progression of neuropathology induced by Parkinson's-disease-related mutant alpha-synuclein. Neuron. 2009;64(6):807-827.

[168] Azeggagh S, Berwick DC. The development of inhibitors of leucine-rich repeat kinase 2 (LRRK2) as a therapeutic strategy for Parkinson's disease: the current state of play. 2022;179(8):1478-1495.

[169] Bae JR, Lee BD. Function and dysfunction of leucine-rich repeat kinase 2 (LRRK2): Parkinson's disease and beyond. BMB reports. 2015;48(5):243-8.

[170] Zhang P, Fan Y, Ru H, et al. Crystal structure of the WD40 domain dimer of LRRK2. Proceedings of the National Academy of Sciences. 2019;116(5):1579-1584.

[171] Zhou G, Rusnac D-V, Park H, et al. An artificial intelligence accelerated virtual screening platform for drug discovery. Nature Communications. 2024;15(1):7761. * Key example of how the field is moving toward opensource and a valuable tool for VS

[172] Rusnac D-V, Lin H-C, Canzani D, et al. Recognition of the Diglycine C-End Degron by CRL2KLHDC2 Ubiquitin Ligase. Molecular Cell. 2018;72(5):813-822.e4.

[173] Therrien E, Weill N, Tomberg A, et al. Docking Ligands into Flexible and Solvated Macromolecules. 7. Impact of Protein Flexibility and Water Molecules on Docking-Based Virtual Screening Accuracy. Journal of Chemical Information and Modeling. 2014;54(11):3198-3210.

[174] Jakhar R, Dangi M, Khichi A, et al. Relevance of Molecular Docking Studies in Drug Designing. Current Bioinformatics. 2020;15(4):270-278.

[175] Turon G, Hlozek J, Woodland JG, et al. First fully-automated AI/ML virtual screening cascade implemented at a drug discovery centre in Africa. Nature Communications. 2023;14(1):5736.

[176] Deng J, Yang Z, Ojima I, et al. Artificial intelligence in drug discovery: applications and techniques. Briefings in Bioinformatics. 2021;23(1).

[177] Parvatikar PP, Patil S, Khaparkhuntikar K, et al. Artificial intelligence: Machine learning approach for screening large database and drug discovery. Antiviral Research. 2023;220:105740.

[178] Oliveira TAd, Silva MPd, Maia EHB, et al. Virtual Screening Algorithms in Drug Discovery: A Review Focused on Machine and Deep Learning Methods. Drugs and Drug Candidates. 2023;2(2):311-334.

[179] Brown DG, Boström J. Where Do Recent Small Molecule Clinical Development Candidates Come From? Journal of Medicinal Chemistry. 2018;61(21):9442-9468.

[180] Zhao H. The Science and Art of Structure-Based Virtual Screening. ACS Medicinal Chemistry Letters. 2024;15(4):436-440.