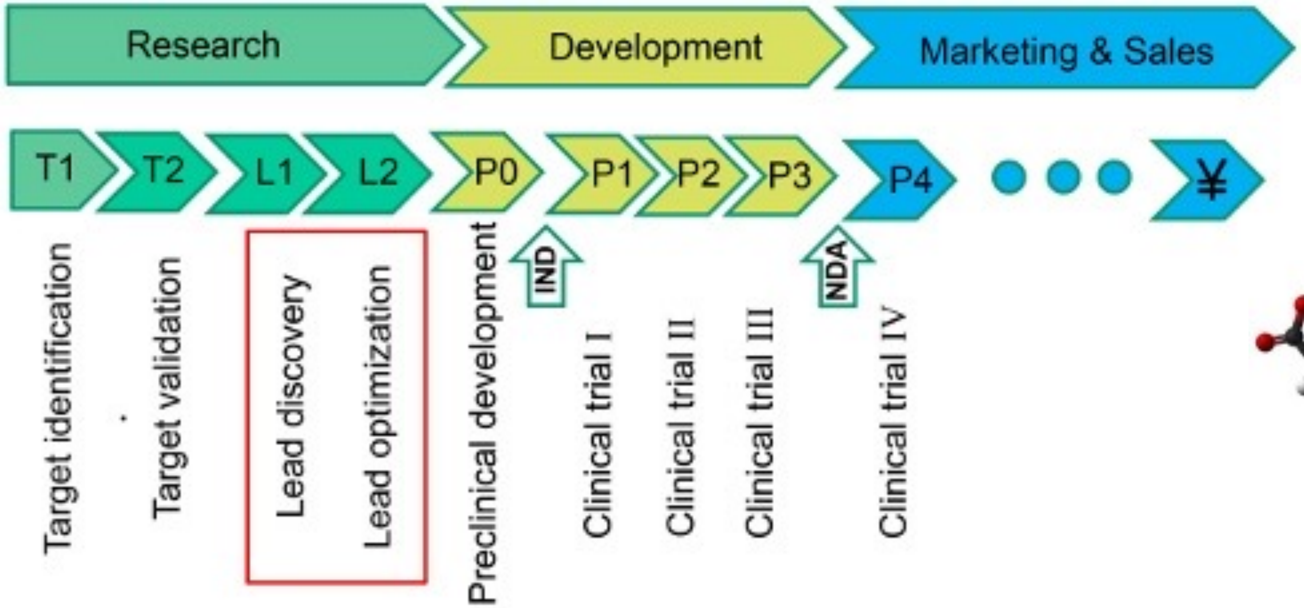
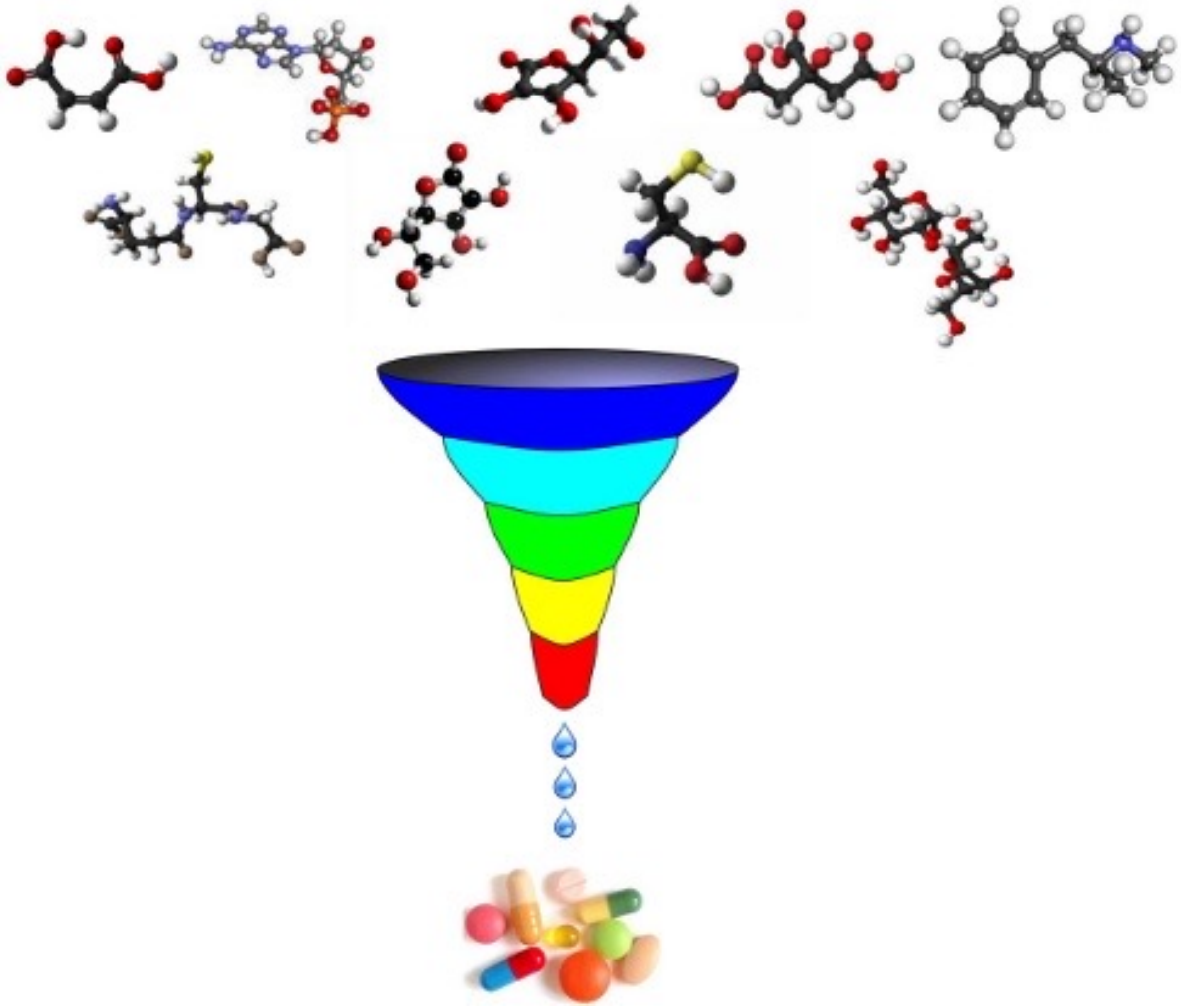
Review of Different Approaches in Virtual Screening

Łukasz Milewski





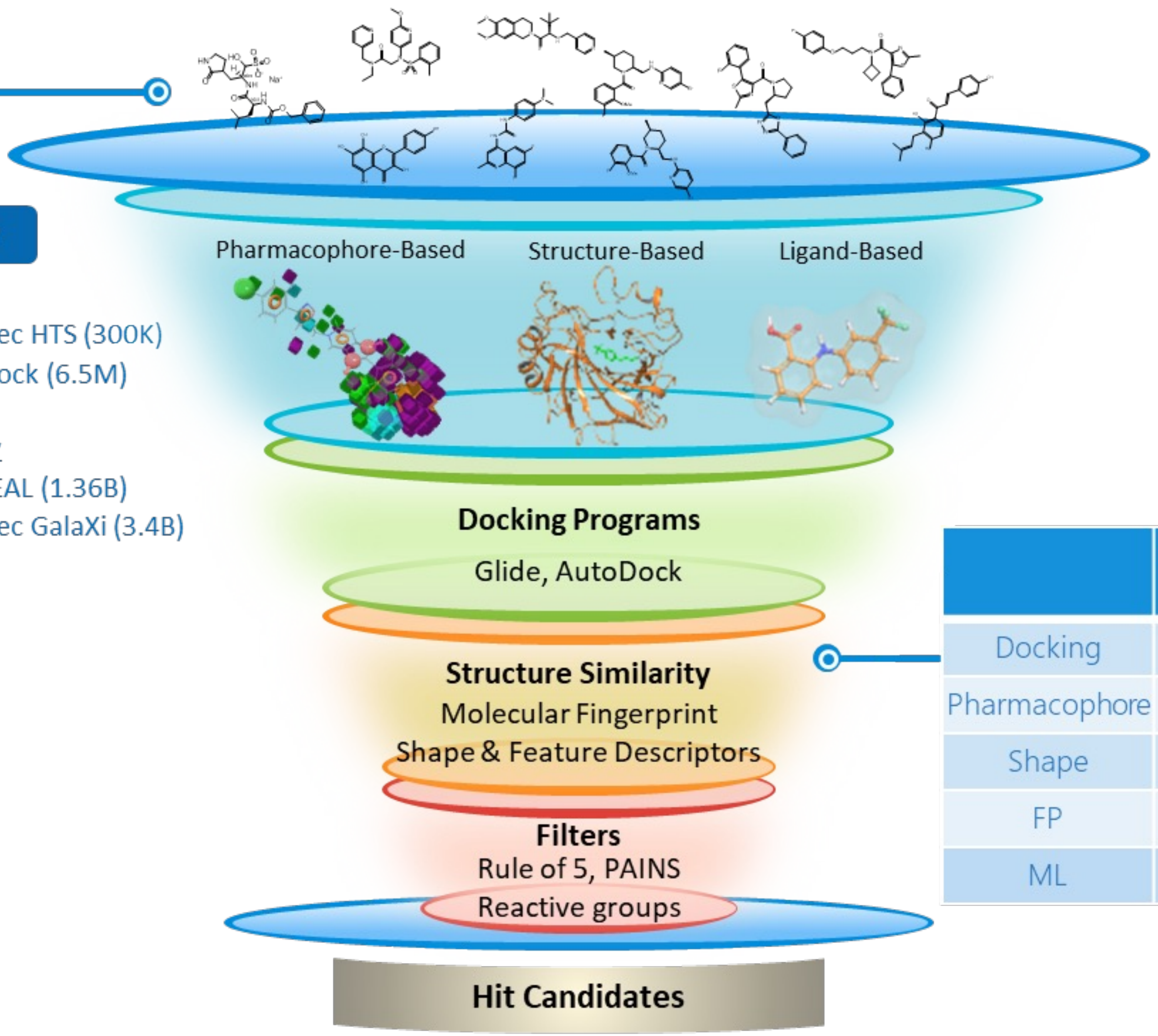
- - Virtual Screening (VS) is a computational technique used in drug discovery to identify potential drug candidates from large compound libraries.
- - It plays a crucial role in accelerating the drug discovery process by reducing the number of compounds that need to be synthesized and tested experimentally.

http://www2.riken.jp/zhangiru/en_research_3.html

Library

Real library
- WuXi AppTec HTS (300K)
- Mcule In-stock (6.5M)

Virtual library
- Enamine REAL (1.36B)
- WuXi AppTec GalaXi (3.4B)

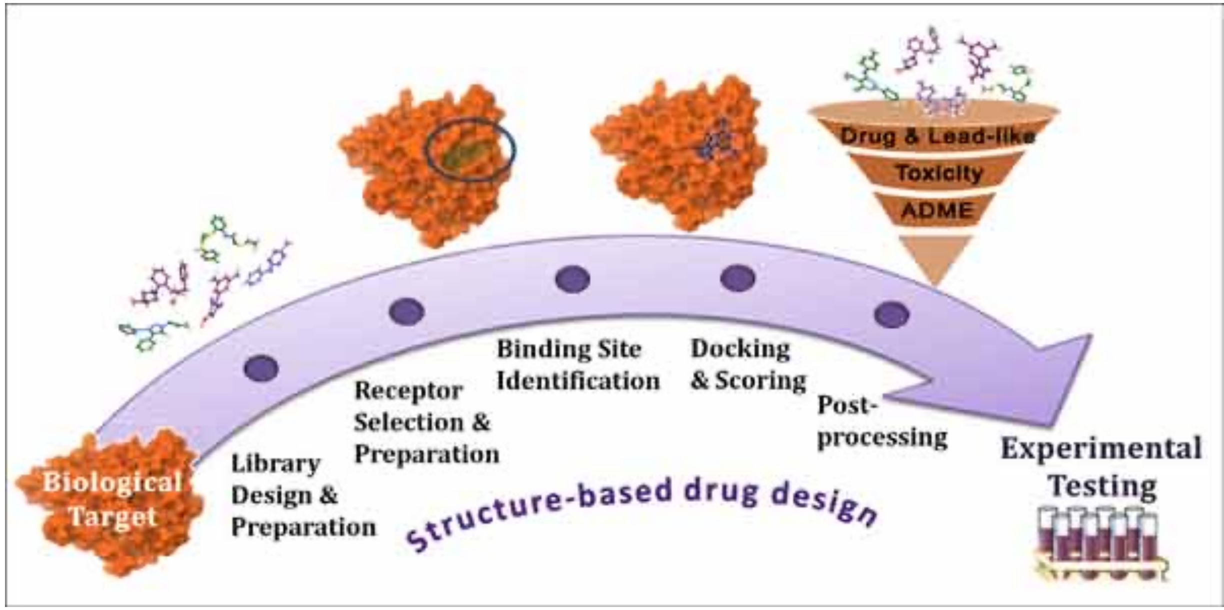


	Chemical Space	Time consuming
Docking	Millions	Days
Pharmacophore	Millions	Days
Shape	Billions	Hours
FP	Billions	Hours
ML	Billions	Hours

https://wuxibiology.com/

Structure-Based Virtual Screening (SBVS)

- - Docking simulations involve predicting the preferred orientation and conformation of a ligand when bound to a target protein. Programs such as AutoDock, DOCK, and Glide are commonly used for molecular docking.
- - Pharmacophore modeling identifies common features or spatial arrangements essential for biological activity among a set of active ligands. Software tools like Pharmer, Phase, and MOE are used for pharmacophore generation and screening.



Lionta et al., 2014

Ligand-Based Virtual Screening (LBVS)

- - Similarity searching compares the structural similarity of a query compound to known active compounds in a database. Programs such as ROCS, ChemMine, and ChemAxon are used for similarity-based screening.
- - Quantitative Structure-Activity Relationship (QSAR) modeling establishes mathematical relationships between chemical structure and biological activity. Software tools like Open3DQSAR, QSARINS, and Dragon are used for QSAR analysis.

Fragment-Based Virtual Screening (FBVS)

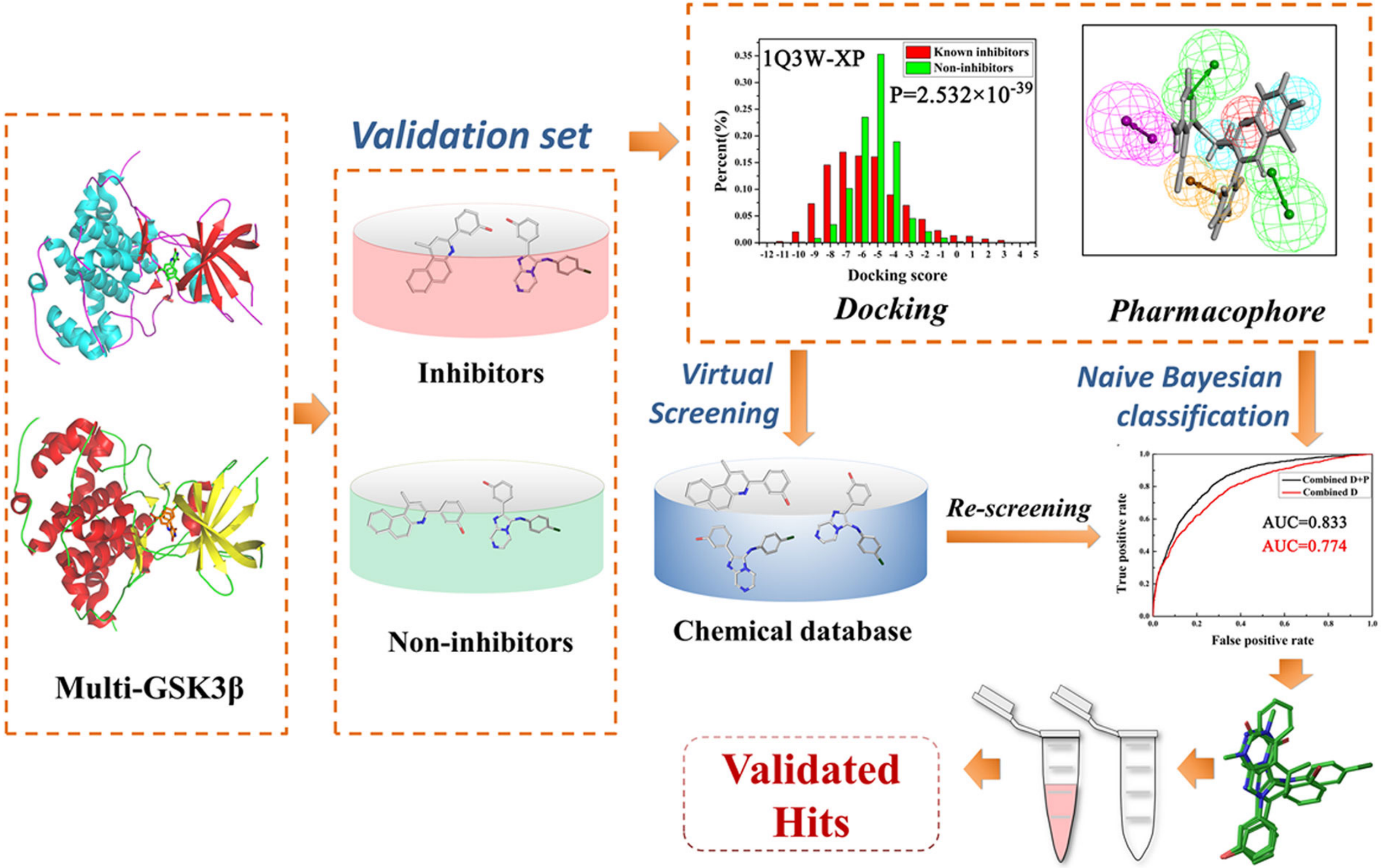
- - FBVS involves screening libraries of small molecular fragments rather than complete ligands. Programs such as FragFEATURE, SPROUT, and eLBOW are used for fragment library generation and screening.
- - Nuclear Magnetic Resonance (NMR) techniques are commonly used to identify fragment hits binding to target proteins. Software tools like SPRING, FTMap, and MEGADOCK are used for NMR-based fragment screening.

Machine Learning (ML) in Virtual Screening

- - ML techniques learn patterns from known ligand-target interactions to predict the activity of new compounds. Programs such as RDKit, scikit-learn, and TensorFlow provide libraries for ML model development.
- - Data preparation involves feature selection and featurization of chemical compounds using tools like PaDEL, Cheminformatics Toolkit, and RDKit.

Naive Bayes and Bayesian Networks

- \- Naive Bayes classifiers are probabilistic models based on Bayes' theorem, assuming independence between features. Programs such as Weka, Orange, and KNIME provide implementations of Naive Bayes classifiers.
- \- Bayesian networks represent probabilistic relationships among a set of variables using directed acyclic graphs. Tools like Hugin, Samlam, and GeNIe are used for Bayesian network modeling.



Zhu et al., 2020

Decision Trees (DTs)

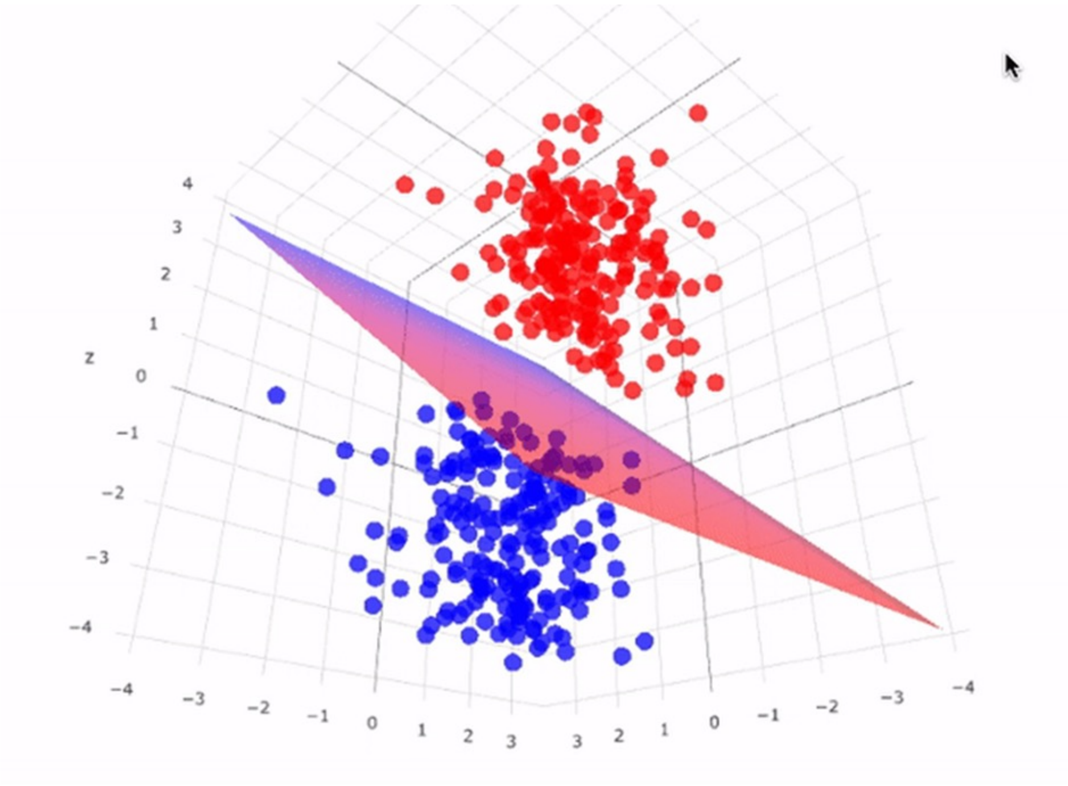
- - Decision trees are hierarchical structures that recursively partition data based on feature attributes. Programs such as scikit-learn, Weka, and C4.5 provide implementations of decision tree algorithms.
- - Random Decision Forests (RDFs) combine multiple decision trees to improve predictive accuracy and robustness. Tools like scikit-learn, Weka, and RandomForest package in R are commonly used for RDF-based virtual screening.

K-Nearest Neighbors (KNN)

- - KNN is a simple and intuitive algorithm that classifies data points based on the majority class of their k-nearest neighbors. Libraries such as scikit-learn, Weka, and MATLAB provide implementations of KNN classifiers.
- - It is commonly used in virtual screening for similarity searching and classification tasks.

Support Vector Machines (SVMs)

- - SVMs are supervised learning models that find the optimal hyperplane separating different classes in feature space. Libraries such as LIBSVM, scikit-learn, and MATLAB provide implementations of SVM classifiers.
- - They are widely used in virtual screening for their ability to handle high-dimensional data and nonlinear relationships.



Medium.com

Artificial Neural Networks (ANNs)

- \- ANNs are computational models inspired by biological neural networks, consisting of interconnected nodes organized in layers. Frameworks such as TensorFlow, Keras, and PyTorch provide tools for building and training neural networks.
- \- ANNs excel in capturing complex relationships in chemical data and have shown promise in virtual screening.

Deep Learning (DL) Techniques

- - DL techniques, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), excel in processing sequential and spatial data. Frameworks such as TensorFlow, PyTorch, and Keras provide implementations of DL models.
- - Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) are used for generating novel molecular structures.

Evolutionary Algorithms

- - Evolutionary algorithms mimic the process of natural selection to optimize solutions to complex problems. Libraries such as DEAP, PyGMO, and ECJ provide implementations of evolutionary algorithms.
- - Genetic Algorithms (GAs) and Particle Swarm Optimization (PSO) are commonly used in drug design.

Molecular Docking Algorithms

- - Molecular docking algorithms predict the preferred conformation and binding affinity of a ligand to a target protein. Programs such as AutoDock, DOCK, and Glide are commonly used for molecular docking.
- - Incremental construction methods iteratively build ligand conformations within the protein binding site.

Consensus Scoring

- - Consensus scoring combines predictions from multiple virtual screening methods to improve accuracy and reliability. Tools like MOE, Maestro, and Discovery Studio provide features for consensus scoring.
- - Rescoring and ranking strategies are used to refine initial screening results and improve hit rates.

Case Studies and Applications

- - Successful case studies demonstrate the utility of virtual screening in drug discovery across various therapeutic areas. Examples include the discovery of HIV protease inhibitors using AutoDock and the identification of kinase inhibitors using Glide.
- - Challenges, such as compound promiscuity and off-target effects, highlight the importance of careful validation and experimental validation.

Regulatory Considerations

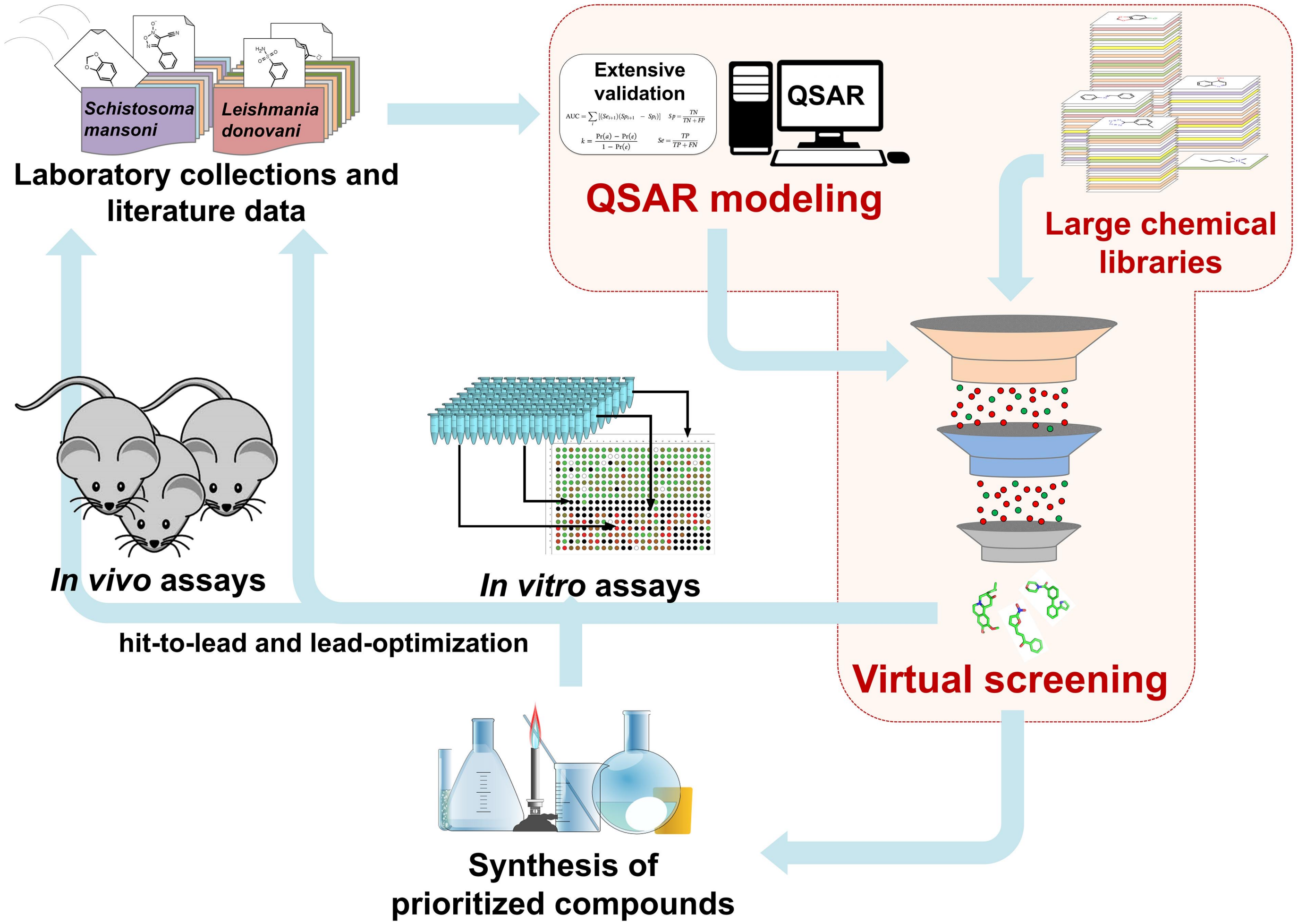
- - Regulatory agencies, such as the FDA and EMA, provide guidelines for the validation and reliability of computational models used in drug discovery. Tools like Derek Nexus, TOPKAT, and ADMET Predictor are used for predicting ADMET properties.
- - OECD principles emphasize the importance of transparency, reproducibility, and scientific rigor in model development and validation.

Multidisciplinary Approach

- - Virtual screening requires a multidisciplinary approach integrating computational, chemical, and biological expertise. Collaboration between interdisciplinary teams is essential for successful drug discovery and development.
- - Software tools like Schrödinger Suite, OpenEye Toolkits, and Cresset provide integrated solutions for virtual screening, lead optimization, and ADMET prediction.

Future Outlook

- - Integration of experimental data, such as omics data and high-throughput screening results, will enhance the predictive power of virtual screening models. Cloud computing and big data analytics will enable the analysis of large-scale chemical datasets and accelerate drug discovery workflows.
- - The shift towards AI/ML-driven approaches will revolutionize virtual screening and lead to the discovery of novel therapeutics.



Neves et al., 2018

Challenges and Limitations

- - Scoring accuracy remains a challenge in virtual screening, particularly in accurately predicting ligand binding affinities. Handling ultra-large compound libraries poses computational challenges and requires efficient algorithms and parallel computing resources.
- - Interpretability of ML models is crucial for understanding predictions and making informed decisions in drug discovery.

Best Practices and Recommendations

- \- Establishing standardized protocols and guidelines for model development, validation, and reporting is essential for ensuring reproducibility and transparency. Continuous updates and improvements to virtual screening methods and algorithms are necessary to keep pace with advancements in computational and experimental techniques.
- \- Collaboration and knowledge sharing within the scientific community facilitate the adoption of best practices and drive innovation in drug discovery.

Conclusion

- - Virtual screening is a powerful tool in drug discovery, enabling the efficient identification and optimization of lead compounds. A diverse array of computational techniques, including machine learning and molecular docking, has revolutionized virtual screening workflows.
- - Addressing challenges and embracing multidisciplinary collaboration will drive the future of virtual screening and advance the discovery of life-saving therapeutics.

Literature

- http://www2.riken.jp/zhangiru/en_research_3.html
- Oliveira et al., Virtual Screening Algorithms in Drug Discovery: A Review Focused on Machine and Deep Learning Methods, Drugs and drug candidates MDPI, 2023
- Lionta et al., Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances, Current Topics in Medicinal Chemistry, 2014
- https://wuxibiology.com/drug-discovery-services/hit-finding-and-screening-services/virtual-screening/
- Zhu et al., Integrating Machine Learning-Based Virtual Screening With Multiple Protein Structures and Bio-Assay Evaluation for Discovery of Novel GSK3β Inhibitors, Frontiers in Pharmacology, 2020
- https://medium.com/low-code-for-advanced-data-science/support-vector-machines-svm-an-intuitive-explanation-b084d6238106
- Neves et al., QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery, 2018