STM JOURNALS

International Journal of
Molecular Biotechnological Research

Volume 1, Issue 2, 2023
DOI (Journal): 10.37591/IJMBR

https://journals.stmjournals.com/ijmbr/

Review IJMBR

Illuminating the Frontier of Drug Discovery: Unleashing the Power of Bioinformatics for Unprecedented Breakthroughs

S. Luqman Ali[1], Awais Ali[1,*], Waseef Ullah[2], Kashif Adil[2], M. Usman2

Abstract

It takes a long time and a lot of effort to discover and develop new drugs, which necessitates extensive study and testing. With the help of computational techniques and data analysis, bioinformatics has grown to be a potent tool for drug discovery in recent years, allowing researchers to find new drugs faster. In this review, we examine the role of bioinformatics in drug discovery, including the use of ligand- and structure-based drug design, virtual screening based on pharmacophore models, de novo design based on pharmacophore models, and quantitative structure-activity relationship (QSAR) models and machine learning techniques. We also talk about how important data collection from different sources, like natural and synthetic databases, is for supporting drug discovery efforts. We highlight the potential of bioinformatics to revolutionise the field of drug discovery and to hasten the creation of new medications for the treatment of a variety of diseases through an analysis of recent research.

Keywords: Bioinformatics, Drug discovery, Ligand-based drug design, Structure-based drug design, Virtual screening, QSAR, Machine learning, Data assortment

INTRODUCTION

Drug Discovery by Bioinformatics

Drug discovery is a complex and time-consuming process that involves the identification and development of new therapeutic compounds. Bioinformatics, a multidisciplinary field that combines biology, computer science, and statistics, plays a crucial role in accelerating and streamlining the drug discovery process. One of the key contributions of bioinformatics in drug discovery is target identification. By analyzing genomic and proteomic data, researchers can identify potential drug targets, such as specific proteins or genes associated with a particular disease. This information provides a foundation for further investigation and development of therapeutic interventions. Additionally, bioinformatics enables virtual screening, which involves the rapid screening of vast libraries of compounds to identify potential drug candidates. Through computational algorithms and predictive models, researchers can assess the likelihood of a compound binding to a target protein and evaluate its potential therapeutic activity. This approach significantly expedites the process of identifying promising compounds for further experimental validation.

Bioinformatics also facilitates molecular docking simulations, which predict how small molecules, known as ligands, interact with target proteins. By examining the binding affinity and orientation of the ligand within the target's binding site, researchers can assess the potential

***Author for Correspondence**
Awais Ali
E-mail: Awaisalbio@gmail.com

[1]M.phill Scholars, Department of Biochemistry, Abdul Wali khan university Mardan, Pakistan
[2]Students, Department of Biochemistry, Abdul Wali khan university Mardan, Pakistan

Received Date: June 14, 2023
Accepted Date: June 23, 2023
Published Date: July 15, 2023

Citation: S. Luqman Ali, Awais Ali, Waseef Ullah, Kashif Adil, M. Usman. Illuminating the Frontier of Drug Discovery: Unleashing the Power of Bioinformatics for Unprecedented Breakthroughs. International Journal of Molecular Biotechnological and Research. 2023;1(2): 1–10p.

effectiveness of a compound as a drug candidate. Furthermore, bioinformatics plays a crucial role in predicting the Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADME-Tox) properties of potential drug candidates. These predictions help researchers assess the pharmacokinetic and safety profiles of compounds, aiding in the selection of promising drug candidates for further experimental validation. By integrating and analyzing diverse biological and chemical data, bioinformatics provides comprehensive insights into the mechanisms underlying diseases and helps identify potential drug targets. Overall, bioinformatics serves as a powerful tool in drug discovery, enabling researchers to efficiently navigate through vast amounts of data, accelerate the identification of potential drug candidates, and optimize the development of new therapeutic interventions.

Bioinformatics is a tool for the analysis of biological data, basically it is a field with tackle with mathematics, physics , computer science and biological data (Parikh et al., 2023) [9]. Bioinformatics plays an important role in the expansion of small molecules for over the decades through CAAD (computer-aided drug designing) also termed as *In Silico* approach or computational approach (Figure 1).

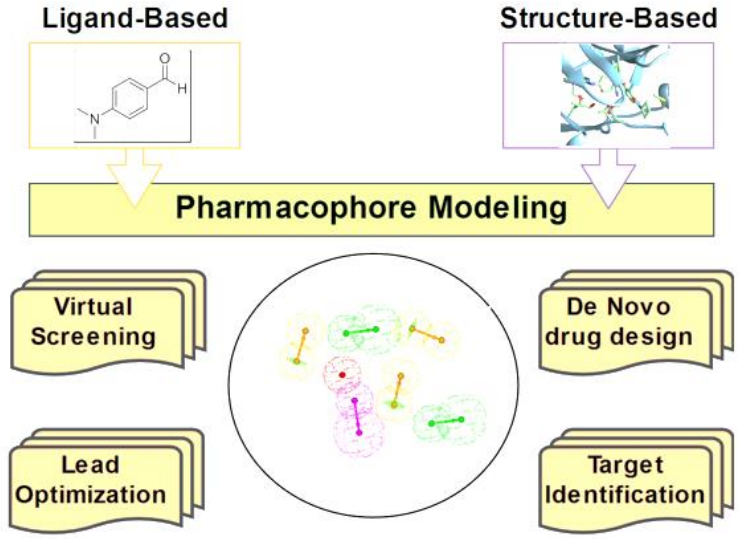

Figure 1. Main methods used for the discovery of drug by CAAD:

Ligand Base Drug Designing

Ligand-based drug development starts with either a single molecule or a set of compounds known to be potent against a target and based on the understanding of structure-activity relationships (SAR), potency and other important attributes are strengthened by producing appropriate analogs. Designing may be performed by Topliss technique or simple analog design based on structural similarities or features. Often, computational tools such as pharmacophore models or shape of the molecules could be beneficial for design objectives. Once a dataset of sufficient size becomes available, with a good range of potency, a Quantitative Structure-Activity Relationships (QSAR) models may be constructed and deployed if the models are powerful enough for prediction purposes. Similarly, if the target is well recognisable with a number of substances already recognised in public literature or public databases, then machine-learning based models may also be examined. If the machine-learning models are robust enough, they may be applied for filtering design ideas or virtual screening or scaffold-hopping hits. Jubilant has been effective in creating clinical candidate medicines for numerous targets for which the target structure was not known at the time. The computational chemistry team, coupled with the medicinal chemistry team, works particularly closely, to perform the projects required LBDD efforts (Ajjarapu et al., 2022) [1].

Jubilant Biosys employs Schrodinger software package as well as Cresset's scaffold-hopping tool, Spark, for driving ligand-based drug development programmes. The computational chemistry section

of Jubilant Biosys also employs multiple machine-learning (ML) methodologies to construct ML-based models, including regression and classification models, to support LBDD efforts (Sliwoski et al., 2014) [11].

If the target structure is not known, but structures of its closest homologues are known, then a homology-based model may be created using the experimental coordinates of the structure of the closest homologue. If the homology model is strong enough, then a structure-based design approach may be utilised.

Pharmacophore-model-based virtual screening

Pharmacophore-model-based virtual screening is a computational drug discovery technique used to identify potential drug candidates that have a high likelihood of binding to a specific target protein. This approach utilizes pharmacophore models, which are spatial arrangements of key molecular features necessary for the interaction between a drug and its target. In pharmacophore-model-based virtual screening, a virtual library of compounds is screened against a pharmacophore model generated from the known ligands or active compounds for the target protein (Sliwoski et al., 2014) [11]. The pharmacophore model represents the essential features required for a compound to interact with the target protein in a specific way. During the screening process, the virtual compounds are evaluated based on their ability to match the pharmacophore features and spatial arrangements. Compounds that closely match the pharmacophore model are considered potential hits and can be further investigated through subsequent experimental or computational studies.

This method is valuable in early-stage drug discovery as it allows for the rapid screening of a large number of compounds, significantly reducing the time and cost associated with traditional experimental screening methods. Additionally, pharmacophore-model-based virtual screening can provide insights into the binding mode and interaction patterns of potential drug candidates, aiding in the optimization and design of new compounds. It is important to note that while pharmacophore-model-based virtual screening is a powerful tool in drug discovery, it is a computational approach and should be followed by experimental validation to confirm the activity and effectiveness of the identified compounds.

Pharmacophore-based de novo design

Pharmacophore-based de novo design is a computational approach used in drug discovery and medicinal chemistry to generate novel chemical compounds with desired pharmacological properties. This method involves the identification and characterization of key pharmacophoric features necessary for a molecule to interact with its target receptor or enzyme.The process begins with the creation of a pharmacophore model, which represents the spatial arrangement of essential chemical features, such as hydrogen bond donors/acceptors, hydrophobic regions, and aromatic groups. These features are derived from known active ligands or structural information about the target protein.Once the pharmacophore model is established, it is employed to screen large compound libraries or to guide the generation of new chemical entities through virtual synthesis. During this de novo design process, various algorithms and scoring functions are employed to prioritize and optimize compounds based on their predicted fit to the pharmacophore model and their potential to exhibit desired pharmacological activities (Figure 2).

Pharmacophore-based de novo design offers several advantages in drug discovery. It allows researchers to explore chemical space beyond existing molecules, enabling the discovery of unique and patentable compounds. Additionally, this approach helps in reducing the reliance on serendipity and empirical synthesis, leading to a more rational and efficient drug design process. It is important to note that the success of pharmacophore-based de novo design relies heavily on the accuracy and relevance of the pharmacophore model. Therefore, rigorous validation and optimization of the model using experimental data are crucial to ensure the reliability and predictive power of the generated compounds.

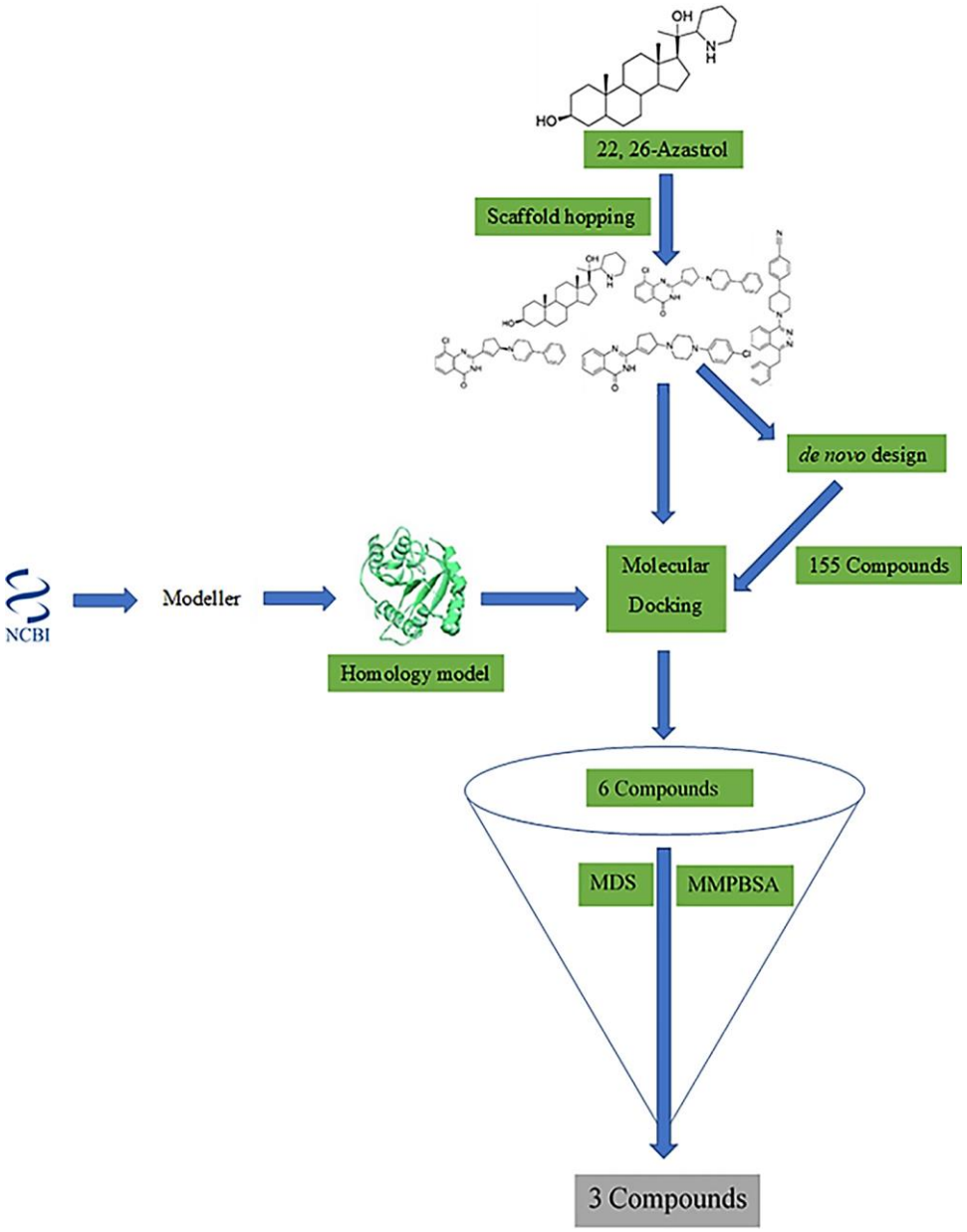

Figure 2. Denvo data Pharmacophore-based pathway.

Structure Base Drug Designing

Structure base drug designing include pharmacophore, docking of ligand, and the methods to design ligand. Meanwhile ligand base drug designing include pharmacophore on the base of ligand, molecular docking, and QSAR (quantitative structure activity relationship). Bioinformatics also plays an important role in the optimization of ligand structure by DFT (density functional thory method) method and toxicity prediction for the safe use of drug in future (Rastogi et al., 2022) [10].

CAAD is very proficient when it comes to increse the lead rate of drug compounds because it is a method that use targeted research than the combinatorial chemistry. CAAD not only expalins the molecular base activity but also helps in the prediction of derivatives that improves activity (Hemmerling & Piel, 2022) [5]. There are three main purpose of CAAD (Figure 3).

- Secreen the large set of libraries into smaller set for the prediction of acive compounds.
- Optimization of drug pharmacokinetics and metabolism including excretion, distribution, absorption and toxicity prediction by ADMET.

- Designing of compounds by growing functional groups or by together fragments into chemotypes.

One of the most common tool of CAAD is to screen the libraries of virtual compounds that termed as vHTS (virtual high throughput screening). vHTS plays an important role in similar chemical search by topology or fingerprints for the selection of compounds with good predicted biological activity by the model of QSAR and mapping of pharmacophore, as well as the virtual molecular docking of the compounds according to the target of interest termed as docking based on structure (Deng et al., 2022) [3].

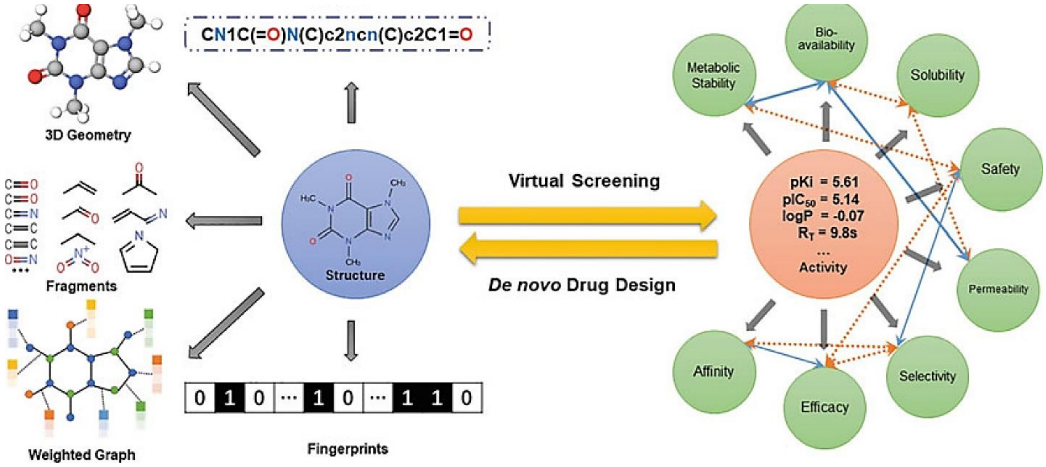

Figure 3. Different pathways and routs of Strucutre-based Pharmacoshpore.

Assortment of data

For discovery of drug or repurposing of drug data is collected from different database according to the need either it is from the natural sources or synthetic (Irham et al., 2022) [6]. Different kind of databases are available online for the assortment of data As shown in Table 1.

Table 1. Handy Science Databases and Tools for Researchers and Scholars.

Database	**Features**	**Access**
Arvix	Open access for almost 1,094,715 e-books of physics, biology and computer science	Access is Public
Compass ASTM (on and off campus access)	Give free access to 12500+ global and international standards	Public access
Biorxiv	It is a pre print server for biology and also some overlap fields of chemistry such as biochemistry and cell biology.	Public access
Bsol	Library of almost 90,000 international standards with the 24/7 availability as well as all the data revised every single day	Free of public access
Common chemistry CAS	An open source to access the information about chemicals. Almost it has 500,000 chemicals according to the community area of interest.	Open Access
Chemical entities and biological interest ChEBI	It is a sort of chemicals dictionary that is freely available and based on small molecules.	Free source
Chemrxiv	An online tool that cover the broad range of fields that comes in chemistry	Free source
Chemspider	A database that provides the free chemical structure almost the 35 million structure.	Open access
Cite in the right way	It is an essential tool for the referencing of source. This will help in reference the source and also to avoid the plagiarism	Open access
e-EROS	A great tool of referencing in organic synthesis. It has detailed information about almost the 4500 reagents as well as catalyst. Every year this cite almost add 200 new more articles and update it.	Open access
Lens	It is a trans-domain tool that give information about the patents and scholar works as a management source.	Open access

QSAR (Quantitative structure activity relationship)

From the last decade QSAR or pharmacopore modeling has been established as a major part of computational modeling and methdologies. In any research domain, QSAR model characterize by the collection of well described protocols as well as procedures that helps in the exploration of ever growing assortment of biologically active ligands, chemical compounds or structures. This article will provide critical pharmacophore or QSAR modeling that regard as a best practice in this field (Labjar et al., 2022) [7].

Workflow for the predictive QSAR model

With the rapid increase in the communication as well as information technologies in the last few years has changed our abilities of collecting, storing and analyzing all sorts of data. This whole procedure has a great influence on the research in many domains such as the rapid developent in the formation of new generation selective as well as effective medicines (Winkler, 2022) [13].

Usually large database that contain chemical compounds in millions further tested in numerous biolgical assesses such as PubChem is the largest database that available for the online collection of data and it also reveiwed by Tropsha and Opera. To find out the new lead drug, it is the necessity of efficient as well as robust protocol that use to screen the virtual libraries and chemical database with well known properties. To this point QSAR or pharmacophore model provide a well known method for both the exploitation and exploration relation ship between biological actions and chemical structures for the formation of innovative drug candidate Therefore, models that are the result of predictive QSAR could be use to select the structures for the validation of experimental value. As a fact the crucial increase in the validation of experimental value as the eventual declaration of model base prediction value.

There are many different softwares that used for the assortment of large database according to the structure-activity relationship. There are many development and advances has been done in recent years as shown in Table 2 and Table 3.

Table 2. Comprehensive List of Software for QSAR Modeling and Analysis.

Name	Approaches
SiRMS	Simplex representation of the structures of the molecules, which is a flexible representation of chemical compounds and structures.
Mixtures Representation	Tool for QSAR model as well as predictive properties of chemicals.
Quasi mixture representation	It improves the over all performace of QSAR modeling as well as descriptors used by machine learning approach.
Structural interpreation	Structural interpretation also plays an important role in QSAR model interpretation regardless of descriptors used.
Pc interpretation by QSAR modeling	For the interpretation of QSAR model according to PC properties, there the chemical structures show no dependance on the methods of machine learning.
3D pharmacophore modeling	The visualization of 3D pharmacophore model for the discovery of similar pharmacophores as well as 3D-pharmacophore model.
Structure generation by fragments	The method for fragment base generation of synthetically available as well as valid structures.

Table 3. Software used for QSAR chemoinformatics tool.

Name	Features	Developer	Platform
Spci	A tool to retive the structure activity relationship by large database.	Accelrys, Inc.	Windows
Sirms	A tool that is used to generate the 2D descriptors of single compounds as well as chemical reactions.	Accelrys, Inc.	Windows
Pmapper	A tool that is used to generate hashes of 3D pharmacophore.	Tripos International	Any
Psearch	A tool that is used for ligand base 3D pharmacophore model.	Accelrys, Inc.	Windows
Crem	A tool that is used for the fragment generation of hemical structures.	Accelrys, Inc.	Windows

MACHINE LEARNING IN DRUG DESIGN

The drug discovery process comprises the use of hybrid methodologies for discovery and synthesis of novel bioactive compounds, which may be candidate to new drugs. One essential strategy is the

accurate prediction of biological activity of chemical compounds using a set of atomic and molecular descriptors, and this is known as quantitative structure-activity relationships or QSAR research. It is generally recognised that physicochemical and structural features of chemical compounds are vital to comprehend many parts of chemical and biological interactions in drug design projects and many other chemical processes. QSAR approaches are utilised to build statistical models relating chemical structure and biological activity, as well as they are valuable in describing the mechanisms of the chemical–biological interaction in various biomolecules. One of the most fundamental aspects of QSAR models is their predictive potential.

As shown above, QSAR studies utilises extrathermodynamically derived and computational-based descriptors in order to correlate them with biological activity. From the discovery of a fairly small number of connected or unrelated chemicals with recognised biological activities, many descriptors for these molecules may be developed. These basic molecular descriptors regularly utilised in QSAR research may be grouped roughly into four primary classes: electronic, steric, hydrophobic and topological. The next step in QSAR studies is developing a suitable statistical model, employing the descriptors obtained previously for a compound set, and the main application of this model is the activity prediction of new compounds and, consequently, to use such model for understanding the possible mechanisms of action for a particular drug (Labjar et al., 2022) [7].

The quality and efficacy of a QSAR model relies totally on the validity of input data, selection of suitable descriptors and statistical approaches, and most critically validation of the created model. It is crucial to stress that a QSAR model is valid only for analogous molecule structures applied to construct the model. The validation process may be conducted using numerous strategies: (a) internal validation (or cross-validation); (b) external validation (test compound set, not employed in the model training); (c) data randomization (or Y-scrambling), and others.

In the previous decades, numerous statistical methodologies have increased the arsenal of instruments that may be employed to QSAR study. A certain number of computational algorithms have been established effective for the development of these correlations such as multiple linear regression (MLR) and partial least squares (PLS). Recently, there is also an increased interest in the use of artificial neural networks (ANNs) and support vector machines (SVM) in the area of QSAR, as well as other molecular modeling techniques have been recognised as significant tools in drug discovery.

The most common machine learning techniques (MLT) can be classified as supervised (such as Multilayer Perceptrons, Bayesian Neural Network, and Support Vector Machines), unsupervised (for example, Self-Organizing Maps) and hybrid models, such as Counter Propagation Neural Network (CPN) , which comprises the advantages of supervised and unsupervised learning techniques.

MLT are excellent for QSAR research as there is a list of compounds with known biological activity available for the training stage. There are countless potential parameters for each molecule and the contribution of each parameter is not known a priori. Below, characteristics of numerous machine learning algorithms will be presented in greater depth.

The interest in the application of machine learning approaches (MLT) as drug design tools is growing in the recent decades. The reason for this is tied to the fact that the drug design is highly sophisticated and needs the use of hybrid approaches . A basic review of certain MLT such as self-organizing maps, multilayer perceptron, bayesian neural networks, counter-propagation neural network and support vector machines is presented in this paper. A comparison between the performance of the provided techniques and other basic statistical methods (such as partial least squares and multiple linear regression) reveals that MLT provides major improvements . Nowadays, the number of study in medicinal chemistry that employ these methodologies has greatly risen, in

particular the utilisation of support vector machines. The state of the art and the future improvements of MLT applications entail the use of these technologies to construct more trustworthy QSAR models. The models developed by MLT may be applied in virtual screening tests as well as filters to develop/discovery novel compounds. An significant difficulty in the drug design industry is the prediction of pharmacokinetic and toxicological aspects, which may avert failures in the clinical phases. Therefore, this work gives a critical point of view on the major MLT and suggests their probable capabilities as a useful tool in drug development (Gertrudes et al., 2012; Neves et al., 2018) [4, 8].

Molecular Docking

Molecular docking is a type of computational chemistry that simplify the process of prediction for preffered orientation of one binding ligand to another receptor. When both of these interact with each other for the formation of stable complex structure, so acquired information could be obtained by the orientation of preffered bound molecules for the prediction of energy stability (binding constant as well as binding affinity), energy strength, as well as energy profiling. This can be obtained by the use of scoring function of docking. This gives the raw data for structure base drug designing also termed as rational drug designing. Docking performed by different tools such as autoDock tool, discovery studio, Molegro virtual docker, and schrodinger etc as shown in Table 4, (Crampon et al., 2022) [2].

Table 4. Tools for Molecular Docking in In Silico Vaccine Design for Epitope and B-cell Studies.

Tool	Description	Availability	Supported Platforms
AutoDock	Widely used tool for molecular docking.	Open-source	Windows, Linux, macOS
Discovery Studio	Comprehensive software suite for drug discovery and design.	Commercial	Windows, Linux
Molegro Virtual Docker	Advanced docking tool for structure-based drug design.	Commercial	Windows, macOS
Schrödinger	Leading suite of computational chemistry software.	Commercial	Windows, Linux, macOS

Molecular Dynamics

Molecular dynamic is the most used study for the structures whose properties based on the microscopic scale. The possible application range for md simulation is pretty broad and constantly growing as the advances in computer acessibility happens. After the binding affinity of ligand-protein interaction these proteins again re-evaluate for further screening that which of these protein is the best for future use of drug (Ugbaja et al., 2022) [12]. There are many softwares for molecular dynamics simulation as shown in Table 5 (Yao et al., 2022) [14].

Table 5. Cutting-Edge Software Tools for Molecular Analysis and Drug Discovery.

Software Name	Main Features	Supported Platforms	Pricing	Website
iMods	Protein flexibility analysis, prediction of protein binding sites	Web-based	Free	https://imods.iqfr.csic.es/
Carbs-flex	Analysis of flexibility and dynamics of carbohydrates and glycoconjugate	Linux, macOS, Windows	Free	http://biocomp.chem.uw.edu.pl/CABSflex/
Gromacs	Molecular dynamics simulations, energy minimization, free energy calculations	Linux, macOS, Windows	Free	https://www.gromacs.org/
Schrodinger	Drug discovery, protein structure prediction, molecular dynamics simulations	Linux, macOS, Windows	Commercial, pricing varies	https://www.schrodinger.com/
Toolbox	Minecraft modding tool, texture and resource pack creator, command generator	Windows	Free	https://toolbox-for-minecraft

CONCLUSION

bioinformatics has emerged as a revolutionary tool in drug discovery, accelerating the process of identifying and developing new therapeutic compounds. By leveraging ligand-based and structure-based approaches, virtual screening, pharmacophore modeling, and machine learning techniques, researchers are able to navigate vast data sources and generate valuable insights. The integration of computational methods and data analysis not only expedites drug discovery but also enhances our understanding of structure-activity relationships and enables the design of novel chemical compounds. Through the continuous advancement of bioinformatics, we are poised to witness remarkable breakthroughs in the creation of medications for a wide range of diseases. By embracing these cutting-edge approaches, we can drive innovation and transform the field of drug discovery, ultimately improving the lives of countless individuals worldwide.

Funding

This study received no funds or grants.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

All authors contributed equally to this study.

REFRENCES

1. Ajjarapu, S. M., Tiwari, A., Ramteke, P. W., Singh, D. B., & Kumar, S. (2022). Ligand-based drug designing. In *Bioinformatics* (pp. 233–252). Elsevier.
2. Crampon, K., Giorkallos, A., Deldossi, M., Baud, S., & Steffenel, L. A. (2022). Machine-learning methods for ligand–protein molecular docking. *Drug Discovery Today*, *27*(1), 151–164.
3. Deng, J., Yang, Z., Ojima, I., Samaras, D., & Wang, F. (2022). Artificial intelligence in drug discovery: applications and techniques. *Briefings in Bioinformatics*, *23*(1).
4. Gertrudes, J. C., Maltarollo, V. G., Silva, R. A., Oliveira, P. R., Honorio, K. M., & Da Silva, A. B. F. (2012). Machine learning techniques and drug design. *Current Medicinal Chemistry*, *19*(25), 4289–4297.
5. Hemmerling, F., & Piel, J. (2022). Strategies to access biosynthetic novelty in bacterial genomes for drug discovery. *Nature Reviews Drug Discovery*, *21*(5), 359–378.
6. Irham, L. M., Adikusuma, W., & Perwitasari, D. A. (2022). Genomic variants-driven drug repurposing for tuberculosis by utilizing the established bioinformatic-based approach. *Biochemistry and Biophysics Reports*, *32*, 101334.
7. Labjar, H., Labjar, N., & Kissi, M. (2022). QSAR Anti-HIV Feature Selection and Prediction for Drug Discovery Using Genetic Algorithm and Machine Learning Algorithms. In *Computational Intelligence in Recent Communication Networks* (pp. 191–204). Springer.
8. Neves, B. J., Braga, R. C., Melo-Filho, C. C., Moreira-Filho, J. T., Muratov, E. N., & Andrade, C. H. (2018). QSAR-based virtual screening: advances and applications in drug discovery. *Frontiers in Pharmacology*, *9*, 1275.
9. Parikh, P. K., Savjani, J. K., Gajjar, A. K., & Chhabria, M. T. (2023). Bioinformatics and Cheminformatics Tools in Early Drug Discovery. *Bioinformatics Tools for Pharmaceutical Drug Product Development*, 147–181.
10. Rastogi, S. C., Rastogi, P., & Mendiratta, N. (2022). *Bioinformatics: Methods and Applications-Genomics, Proteomics and Drug Discovery*. PHI Learning Pvt. Ltd.
11. Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacological Reviews*, *66*(1), 334–395.
12. Ugbaja, S. C., Mtambo, S. E., Mushebenge, A. G., Appiah-Kubi, P., Abubakar, B. H., Ntuli, M. L., & Kumalo, H. M. (2022). Structural Investigations and Binding Mechanisms of Oseltamivir Drug Resistance Conferred by the E119V Mutation in Influenza H7N9 Virus. *Molecules*, *27*(14).

https://doi.org/10.3390/molecules27144376
13. Winkler, D. A. (2022). The impact of machine learning on future tuberculosis drug discovery. *Expert Opinion on Drug Discovery*, *17*(9), 925–927.
14. Yao, H., Liu, J., Xu, M., Ji, J., Dai, Q., & You, Z. (2022). Discussion on molecular dynamics (MD) simulations of the asphalt materials. *Advances in Colloid and Interface Science*, *299*, 102565.

