Biotechnology Journal International

Volume 28, Issue 4, Page 119-136, 2024; Article no.BJI.117412
ISSN: 2456-7051
(Past name: British Biotechnology Journal, Past ISSN: 2231–2927, NLM ID: 101616695)

The Synergy of Molecular Docking and Bioinformatics: An in Depth Review in Drug Discovery

Okpo, E. A. [a,b*], Agboke, A. A. [a], Udobi, C. E. [a], John, G. E. [b] and Andy, I. E. [b]

[a] Department of Pharmaceutical Microbiology and Biotechnology, University of Uyo, Akwa Ibom State, Nigeria.
[b] Department of Microbiology, University of Calabar, Calabar, Cross River State, Nigeria.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.9734/bji/2024/v28i4732

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/117412

Review Article

Received: 22/03/2024
Accepted: 24/05/2024
Published: 25/07/2024

ABSTRACT

Molecular docking and bioinformatics have emerged as pivotal tools in the realm of drug discovery, significantly transforming the landscape of pharmaceutical research and development. These computational techniques enable scientists to predict, analyze, and expedite the identification of potential drug candidates with remarkable precision and efficiency. Molecular docking facilitates the virtual screening of vast compound libraries, offering a cost-effective means of selecting promising lead compounds for further study. Bioinformatics, on the other hand, harnesses the power of biological data analysis, encompassing genomics, proteomics, and other omics fields to elucidate disease mechanisms, identify drug targets, and advance personalized medicine. Recent innovations, including the integration of artificial intelligence, cryo-electron microscopy, and

**Corresponding author: E-mail: ekomarchims20@gmail.com; eaokpo@unical.edu.ng;*

Cite as: *E. A., Okpo, Agboke, A. A., Udobi, C. E., John, G. E., and Andy, I. E. 2024. "The Synergy of Molecular Docking and Bioinformatics: An in Depth Review in Drug Discovery". Biotechnology Journal International 28 (4):119-36. https://doi.org/10.9734/bji/2024/v28i4732.*

quantum computing, are poised to propel these tools further into the future of drug discovery. However, alongside these advancements, ethical and regulatory considerations regarding data privacy, algorithmic fairness, transparency, and adherence to legal and ethical guidelines must be carefully navigated. Striking a harmonious balance between innovation and ethical practice is paramount as molecular docking and bioinformatics continue to reshape the landscape of healthcare and pharmaceutical research.

Keywords: Molecular; docking; bioinformatics; drugs; discovery; bioavailability.

1. INTRODUCTION

In the ever-evolving landscape of pharmaceutical research and development, the quest for novel therapeutic agents has intensified, prompting the integration of cutting-edge computational techniques to expedite and enhance the drug discovery process. Molecular docking and bioinformatics have emerged as indispensable tools at the forefront of modern drug discovery, revolutionizing the way potential drug candidates are identified, evaluated, and optimized. This symbiotic relationship between computational methodologies and pharmaceutical science has led to remarkable advancements, enabling the identification of promising drug candidates with higher precision, reduced costs, and accelerated timelines.

Molecular docking, a cornerstone of computational chemistry, plays a pivotal role in predicting small molecule ligands' binding modes and affinities within the active sites of target biomolecules, predominantly proteins [1,2]. This technique simulates the interactions between ligands and receptors, unravelling critical insights into the thermodynamics and kinetics of binding events. Through the application of various docking algorithms and scoring functions, researchers can screen vast chemical libraries and identify potential drug candidates that exhibit favourable binding interactions and pharmacological profiles. Notably, molecular docking aids in the exploration of structure-activity relationships (SARs) and guides medicinal chemists in refining lead compounds for enhanced potency, selectivity, and bioavailability [3].

Complementing molecular docking, bioinformatics harnesses computational approaches to manage and analyze the deluge of biological data generated by high-throughput technologies. Genomic, proteomic, and structural information is seamlessly integrated and interpreted to uncover novel drug targets, predict their functions, and elucidate disease-associated pathways [4]. By deciphering the intricate relationships between genes, proteins, and diseases, bioinformatics assists in the identification of potential biomarkers and therapeutic targets. This wealth of information aids researchers in selecting optimal drug candidates, optimizing treatment strategies, and unravelling the molecular underpinnings of diseases [5,6].

In this dynamic synergy between molecular docking and bioinformatics, the modern drug discovery process is expedited by informed decision-making, efficient lead identification, and optimized compound design. However, challenges such as protein flexibility, accurate scoring, and the integration of diverse data sources persist, necessitating ongoing research and innovation. As such, this review delves into the fundamentals, methodologies, applications, challenges, and future directions of molecular docking and bioinformatics in modern drug discovery.

2. MATERIALS AND METHODS

A systematic literature search was conducted to identify relevant articles focusing on the integration of molecular docking and bioinformatics in the field of drug discovery. Databases such as PubMed, Scopus, and Web of Science were queried using keywords including "molecular docking," "bioinformatics," "drug discovery," and related terms. Articles were included if they provided substantial insights into the collaborative use of molecular docking and bioinformatics in drug discovery. Exclusion criteria encompassed studies unrelated to drug discovery, lacking a focus on molecular docking or bioinformatics, and those published in languages other than English. The identified articles were screened based on titles and abstracts to determine their relevance. Full-text assessments were performed for selected articles to ensure they met the inclusion criteria and provided substantial information on the integration of molecular docking and

bioinformatics. Data extraction involved capturing key information from selected articles, including methodologies employed, software tools utilized, biological targets investigated, and outcomes of the studies. Emphasis was placed on elucidating the synergistic impact of molecular docking and bioinformatics on enhancing drug discovery processes. This review is based on publicly available and previously published data. No human or animal subjects were involved, and ethical approval was not required.

3. EVOLUTION OF DRUG DISCOVERY PROCESSES

The process of drug discovery has undergone significant evolution over the years, driven by advances in scientific understanding, technology, and the need for more efficient and effective drug development [7]. This evolution can be broadly categorized into several phases, each marked by distinct shifts in approach and methodology. The earliest phase of drug discovery, often referred to as the empirical era, was characterized by a trial-and-error approach. Natural products, such as plant extracts and minerals, were the primary sources of therapeutic compounds. Digitalis, derived from the foxglove plant, is an example of a successful drug discovered during this era. This phase was largely dependent on serendipity and lacked a deep understanding of the underlying mechanisms of diseases [8].

The rise of molecular biology and biochemistry in the mid-20th century led to the development of the target-based approach. This era focused on understanding the molecular mechanisms underlying diseases and identifying specific molecular targets, such as enzymes or receptors, that could be modulated by drugs. The advent of high-throughput screening techniques allowed researchers to test large libraries of compounds against specific targets, accelerating the identification of potential drug candidates. The late 20th century saw the integration of combinatorial chemistry and high-throughput screening into drug discovery processes. Combinatorial chemistry enabled the synthesis of large libraries of diverse compounds, while high-throughput screening allowed for the rapid testing of these compounds against a variety of targets [9]. This approach significantly increased the efficiency of lead identification but often lacked the focus on disease relevance. Advancements in structural biology, computational chemistry, and genomics led to the emergence of rational drug design. This approach involves utilizing three-dimensional structures of biomolecular targets to design molecules that fit optimally into their active sites. Computer-aided drug design (CADD) tools became instrumental in predicting the binding interactions between potential drug candidates and their targets. Rational drug design aimed to improve the specificity and affinity of drugs, reducing off-target effects and increasing the likelihood of success [10].

In recent years, there has been a shift towards a systems biology approach, recognizing that diseases are often complex, interconnected systems involving multiple molecular components. Network pharmacology seeks to understand how drugs affect entire biological networks rather than individual targets. This approach takes into account the interactions between multiple pathways and molecules, considering the holistic impact of drugs on the biological system. Advancements in genomics and personalized medicine have further transformed drug discovery. By analysing an individual's genetic makeup and biomarkers, researchers can identify patient populations that are more likely to respond positively to specific treatments. Precision drug discovery aims to develop therapies tailored to an individual's genetic and molecular characteristics, improving efficacy and minimizing adverse effects. The evolution of drug discovery processes has been driven by a combination of scientific breakthroughs, technological innovations, and a deeper understanding of disease mechanisms. From the empirical era to precision drug discovery, each phase has contributed to a more systematic and efficient approach to identifying and developing therapeutic agents [11].

4. MODERN DRUG DISCOVERY

Modern drug discovery is a multifaceted endeavour that combines cutting-edge scientific knowledge, advanced technologies, and interdisciplinary collaboration to identify, design, and develop novel therapeutic agents. This process represents a dynamic interplay between biology, chemistry, computational sciences, and clinical research [12]. With the ultimate goal of addressing unmet medical needs and improving human health, modern drug discovery has undergone remarkable transformations, largely driven by advancements in molecular biology, high-throughput screening, computational modelling, and bioinformatics.

The journey from a potential drug candidate's inception to its approval as a marketable

pharmaceutical involves several key stages. These stages include target identification and validation, hit identification and optimization, preclinical studies, clinical trials, and regulatory approval. Throughout this intricate trajectory, the integration of various tools and methodologies has become imperative to streamline the drug discovery process and enhance the probability of success [13].

In the initial stages of drug discovery, the identification and validation of therapeutic targets – such as disease-associated proteins or enzymes – lay the foundation for subsequent efforts. The advent of genomics and proteomics has facilitated the identification of biomolecules involved in disease pathways, enabling researchers to pinpoint potential points of intervention. Subsequently, high-throughput screening techniques allow for the rapid testing of large compound libraries, leading to the identification of hits that show promising activity against the chosen target [14]. Moreover, the rise of computational techniques has revolutionized the way drug candidates are designed and evaluated. Molecular modelling, molecular docking, and quantitative structure-activity relationship (QSAR) analyses provide valuable insights into the interactions between drugs and their target proteins, guiding medicinal chemists in optimizing molecular structures for enhanced efficacy and reduced side effects. Bioinformatics tools aid in data analysis, target prediction, and off-target effects assessment, thus facilitating informed decision-making [15,16].

The evolution of drug discovery has also seen a shift toward personalized medicine, where treatments are tailored to individual patients' genetic makeup and disease profiles. This shift has been made possible by advancements in biomarker identification, diagnostics, and targeted therapies. As a result, modern drug discovery not only aims to develop drugs with improved efficacy and safety but also to align treatments with the specific needs of patient populations [17].

5. ROLE OF COMPUTATIONAL METHODS IN DRUG DISCOVERY

Certainly, computational methods have played a pivotal role in modern drug discovery, contributing to the identification, design, and optimization of potential drug candidates. The various aspects of computational methods in drug discovery include;

a). Virtual Screening and Ligand-Based Design: Computational methods allow researchers to virtually screen vast chemical databases to identify potential drug candidates that could interact with a specific target. Molecular docking and molecular dynamics simulations help predict the binding affinity and interaction between ligands and target proteins. Machine learning and AI techniques, such as deep learning and random forest models, enhance the accuracy of predicting ligand-target interactions [18].

b). Structure-Based Drug Design: Computational methods enable the design of novel compounds that fit precisely into the active site of a target protein. Techniques like de novo design and fragment-based drug discovery utilize computational algorithms to create new chemical structures or assemble fragments into larger molecules with improved binding affinity and selectivity [19].

c). QSAR and ADME Prediction: Quantitative structure-activity relationship (QSAR) models utilize computational tools to correlate the structural properties of compounds with their biological activity. Additionally, computational methods predict absorption, distribution, metabolism, and excretion (ADME) properties, providing insights into a compound's pharmacokinetic profile [20].

d). Target Identification and Predictive Modelling: Computational methods aid in the identification of potential drug targets by analyzing biological data and identifying pathways associated with diseases. Integrating omics data, network analysis, and machine learning techniques helps predict novel drug targets and understand complex interactions within biological systems [21].

e). Drug Repurposing and Polypharmacology: Computational methods assist in identifying existing drugs that can be repurposed for new indications. By analyzing the interactions between drugs and various targets, polypharmacology approaches leverage computational tools to uncover potential new therapeutic uses for existing compounds [22].

f). Personalized Medicine and Biomarker Discovery: Computational methods play a role in identifying biomarkers associated with diseases and patient responses to treatments. Using bioinformatics and machine learning, researchers can analyze patient data to tailor therapies to

individual characteristics, advancing the field of personalized medicine [23].

6. FUNDAMENTALS OF MOLECULAR DOCKING

Molecular docking is a computational technique used in the field of structural biology and drug discovery to predict the binding mode and affinity of a small molecule (ligand) to a target protein (receptor). It plays a crucial role in the early stages of drug development and the study of protein-ligand interactions. The followings are the fundamentals of molecular docking.

a). Protein Structure Preparation: The first step in molecular docking involves preparing the 3D structure of the protein. Recent advances in cryo-electron microscopy (cryo-EM) and X-ray crystallography have provided high-resolution structures of various proteins, enabling more accurate docking studies. The Protein Data Bank (PDB) is a valuable resource for accessing protein structures (PDB: www.rcsb.org) [24].

b). Ligand Preparation: Ligands, often small organic molecules or potential drug candidates, also need to be prepared in a suitable 3D format for docking. Software tools like Open Babel and RDKit are widely used for ligand preparation [25].

c). Docking Algorithms: Various docking algorithms are available, including rigid-body docking, flexible docking, and induced-fit docking. Recent developments in these algorithms aim to improve accuracy and efficiency. AutoDock Vina and Glide are popular docking programs commonly used in molecular docking [26].

d). Scoring Functions: Scoring functions are used to estimate the binding affinity between the ligand and receptor. Recent efforts focus on the development of more accurate scoring functions, such as machine learning-based approaches [27].

e). Validation and Benchmarking: Rigorous validation and benchmarking of docking methods **are** essential to assess their performance. Diverse datasets and community challenges like the Drug Design Data Resource (D3R) help in evaluating docking accuracy [28].

f). Applications: Molecular docking is widely used in drug discovery, virtual screening, and studying protein-ligand interactions. Recent studies apply docking to identify potential drug candidates for various diseases, including COVID-19 [29].

g). Machine Learning and AI: Recent advancements in machine learning and artificial intelligence have been integrated into molecular docking workflows to enhance accuracy and efficiency [30].

7. MOLECULAR INTERACTIONS IN DRUG BINDING

Molecular interactions play a crucial role in drug binding, as they determine the strength and specificity of the interaction between a drug (ligand) and its target protein (receptor). Understanding these interactions is essential for rational drug design and optimization. Here, are various molecular interactions involved in drug binding.

a). Hydrogen Bonds: Hydrogen bonds involve the electrostatic attraction between a hydrogen atom and a strongly electronegative atom (e.g., oxygen or nitrogen) on the receptor. These interactions are common in drug-receptor binding and contribute to binding specificity [31].

b). Van der Waals Interactions: Van der Waals forces include attractive interactions between atoms or molecules due to fluctuations in electron density. These interactions help stabilize the drug in the binding site [32].

c). Electrostatic Interactions: Electrostatic interactions result from the attraction between positively and negatively charged regions of molecules. Charged drug molecules can form ionic bonds with oppositely charged residues on the receptor [33].

d). Pi-Stacking: Pi-stacking interactions occur between aromatic rings of drug molecules and aromatic amino acid residues in the receptor. These interactions contribute to drug binding stability [34].

e). Hydrophobic Interactions: Hydrophobic interactions involve the exclusion of water molecules from the binding site, favouring the association of nonpolar drug groups with hydrophobic regions on the receptor [35].

f). Covalent Bonds: Some drugs form covalent bonds with specific amino acid residues in the receptor. This irreversible binding can be exploited for targeted therapy [36].

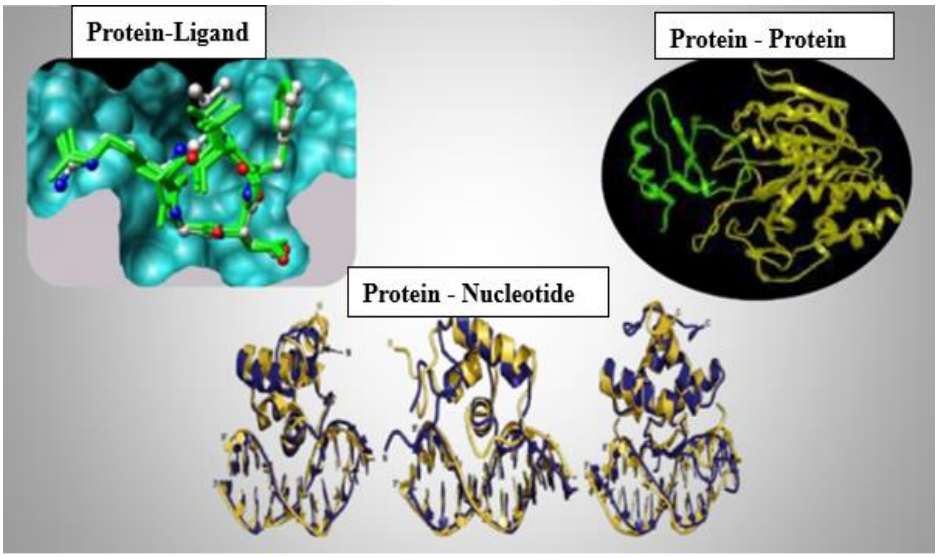

Fig. 1. An illustration of different types of docking

g). Water-Mediated Interactions: Water molecules can act as bridges between the drug and the receptor, forming hydrogen bonds and contributing to the binding affinity [37].

h). Allosteric Interactions: Allosteric sites on proteins can be targeted by drugs to modulate protein function indirectly. Allosteric interactions involve conformational changes in the receptor [38].

8. DOCKING ALGORITHMS AND TECHNIQUES

Molecular docking algorithms and techniques are essential tools in the field of computational chemistry and drug discovery. These methods aim to predict the binding mode and affinity of a ligand (e.g., drug candidate) with a target protein or receptor. The following are the various docking algorithms and techniques available.

a). Rigid Body Docking: In rigid body docking, both the ligand and receptor are treated as rigid structures without significant conformational changes during binding. It's computationally efficient but may not capture subtle changes in the binding site [39].

b). Flexible Docking: Flexible docking accounts for conformational changes in the receptor or ligand during binding. It considers multiple conformations and explores the binding energy landscape more comprehensively [40].

c). Induced Fit Docking: Induced fit docking combines elements of both rigid and flexible docking. It allows the receptor to undergo conformational changes upon ligand binding, optimizing the ligand-receptor interaction [41].

d). Ligand-based Docking: Ligand-based docking methods prioritize ligand conformations based on their compatibility with the receptor, rather than explicitly considering receptor flexibility. It is useful when receptor structural information is limited [42].

e). Machine Learning-Aided Docking: Machine learning techniques, including deep learning and random forest models, have been integrated with docking methods to improve accuracy and efficiency by predicting binding affinities and poses [43].

f). Quantum Mechanics/Molecular Mechanics (QM/MM) Docking: QM/MM docking combines quantum mechanics and molecular mechanics approaches to capture electronic structure effects during docking, providing more accurate binding energy calculations [44].

g). Allosteric Site Docking: Allosteric site docking focuses on identifying ligands that target allosteric binding sites, which can modulate protein function indirectly. It's crucial for drug discovery and target specificity [45].

9. SCORING FUNCTIONS FOR BINDING AFFINITY PREDICTION

There are mainly two categories of scoring functions for binding affinity prediction:

a). Force Field-Based Scoring Functions: Molecular Mechanics (MM) and Molecular Dynamics (MD): These methods simulate the physical forces and motions of atoms and molecules in a complex. Binding affinity is estimated by calculating the energy of the system using force fields like AMBER, CHARMM, or GROMOS. Recent advancements in MD simulations and force fields have improved accuracy [46].

Empirical Scoring Functions: These are simplified energy functions that estimate binding affinity based on various structural and energetic terms, such as electrostatics, van der Waals interactions, hydrogen bonding, and solvation effects. Recent developments include machine-learning-based approaches to improve accuracy [47].

b). Machine Learning-Based Scoring Functions: Docking-Based Approaches: These methods combine molecular docking simulations with machine learning models to predict binding affinities. Recent advancements include the use of deep learning and neural networks for improved accuracy [48].

Free Energy Perturbation (FEP) and Thermodynamic Integration (TI): These techniques employ molecular dynamics simulations to calculate free energy changes upon binding and require extensive computational resources. Recent developments in enhanced sampling methods and machine learning have accelerated FEP/TI-based affinity predictions [49].

Data-Driven Approaches: Machine learning models trained on large datasets of protein-ligand complexes have gained popularity. Recent research focuses on combining structural and sequence information to improve predictive accuracy [50].

10. IMPORTANCE OF BIOINFORMATICS IN DRUG DISCOVERY

Bioinformatics plays a crucial role in drug discovery by facilitating the analysis and interpretation of biological data, which is essential for identifying potential drug targets, understanding disease mechanisms, and optimizing drug candidates. The importance of bioinformatics in drug discovery includes;

a). Target Identification and Validation: Bioinformatics helps identify and validate potential drug targets by analyzing biological data, such as genomics, proteomics, and transcriptomics. It aids in understanding the role of specific genes, proteins, or pathways in diseases [51].

b). Drug-Target Interaction Prediction: Bioinformatics tools predict interactions between drugs and their target proteins, aiding in the selection and optimization of drug candidates [52].

c). Drug Repurposing: Bioinformatics enables the exploration of existing drugs for new therapeutic indications, potentially saving time and resources [53].

d). Pharmacogenomics and Personalized Medicine: Bioinformatics identifies genetic variations that influence drug response, enabling personalized treatment strategies [54].

e). Structural Bioinformatics and Drug Design: Bioinformatics tools assist in modeling and simulating molecular structures, facilitating rational drug design [55].

f). High-Throughput Data Analysis: Bioinformatics processes and analyzes data from high-throughput experiments, such as next-generation sequencing and mass spectrometry, to uncover potential drug targets and biomarkers [56].

g). Drug Safety and Toxicology Prediction: Bioinformatics models predict potential drug side effects and toxicity, helping to prioritize safer drug candidates [57].

h). Big Data and AI in Drug Discovery: Bioinformatics harnesses big data analytics and artificial intelligence to analyze vast datasets, accelerating drug discovery processes [58].

11. ROLE OF BIOINFORMATICS IN DATA MANAGEMENT

Bioinformatics plays a pivotal role in data management within the realm of biological and biomedical research. It involves the application of computational techniques and tools to acquire, store, analyze, and interpret biological data, which can encompass a wide range of information, from DNA sequences to protein structures, clinical records, and more. This field

is indispensable for handling the vast and diverse datasets generated by modern biological research, and it contributes significantly to the advancement of various life sciences. One crucial aspect of data management in bioinformatics is data integration. Biological research generates data from various sources, such as genomics, proteomics, transcriptomics, and metabolomics. These data types are often interconnected, and bioinformatics tools are used to integrate them into a cohesive dataset. Integration enables researchers to gain a comprehensive view of biological processes and systems, facilitating a deeper understanding of complex phenomena [59]. Bioinformatics also plays a pivotal role in data standardization and normalization. Biological data come in different formats and units, making it challenging to compare and analyze them. Bioinformatics tools and databases provide standardized formats and ontologies, allowing researchers to harmonize and normalize data for meaningful comparisons [60].

Additionally, bioinformatics contributes significantly to data storage and retrieval. With the exponential growth of biological data, efficient data storage and retrieval systems are crucial. Databases like GenBank and UniProt serve as repositories for biological information, and bioinformatics techniques are employed to design, maintain, and query these databases [61,62,63]. Moreover, bioinformatics tools aid in data analysis and interpretation. Techniques like sequence alignment, phylogenetic analysis, structural modeling, and machine learning are used to extract knowledge from biological data. These analyses are essential for identifying genes, predicting protein functions, and understanding the genetic basis of diseases [64,65].

Bioinformatics also contributes to data security and privacy. As biological data include sensitive information, protecting patient data and genetic information is of utmost importance. Bioinformatics professionals develop encryption methods, access controls, and secure data transfer protocols to safeguard this information [66]. Bioinformatics is indispensable for data management in the life sciences. It addresses the challenges of data integration, standardization, storage, analysis, and security, facilitating research and innovation in biology and healthcare. As the volume and complexity of biological data continue to grow, bioinformatics will remain a critical field for managing and extracting valuable insights from these datasets.

12. GENOMIC AND PROTEOMIC DATA ANALYSIS

Bioinformatics plays a pivotal role in the analysis of genomic and proteomic data, helping researchers extract valuable insights from the vast amount of biological information generated through various sequencing and experimental techniques.

Genomic Data Analysis: Genomic data analysis involves the study of an organism's complete set of DNA, including genes, non-coding regions, and variations. Bioinformatics is crucial in this field for several key tasks:

- Sequence Alignment: Bioinformatics tools, such as BLAST (Basic Local Alignment Search Tool) [67], help align DNA sequences to reference genomes, facilitating the identification of genes, regulatory elements, and mutations.
- Variant Calling: By comparing individual genomes to a reference, bioinformatics methods can identify single nucleotide polymorphisms (SNPs) and structural variants. Tools like GATK (Genome Analysis Toolkit) [68] are commonly used for this purpose.
- Functional Annotation: Bioinformatics databases and tools provide functional annotations for genes and their products. Examples include Gene Ontology (GO) [69] and KEGG (Kyoto Encyclopedia of Genes and Genomes) [70].
- Phylogenetic Analysis: Bioinformatics enables the construction of evolutionary trees to understand the relationships between species or genes. Programs like PhyML [71] aid in phylogenetic reconstruction.
- Structural Genomics: For analyzing the 3D structures of proteins, bioinformatics tools like SWISS-MODEL [72] assist in structural prediction from genomic data.

Proteomic Data Analysis: Proteomics involves the study of an organism's entire set of proteins, including their structures, functions, and interactions. Bioinformatics is indispensable for proteomic data analysis:

- Protein Identification: Mass spectrometry data, commonly used in proteomics, is

processed with bioinformatics tools like SEQUEST [73] and Mascot [74] to identify proteins from spectra.

- Quantitative Analysis: Proteomics often requires comparing protein abundances across conditions. Bioinformatics methods like TMT (Tandem Mass Tag) labeling and label-free quantification are employed for quantitative analysis.
- Functional Annotation: Just as in genomics, functional annotation of proteins is essential. Databases like UniProt [75] provide comprehensive information on protein function.
- Protein-Protein Interaction (PPI) Analysis: Bioinformatics tools like STRING [76] and Cytoscape [77] aid in the visualization and analysis of PPI networks.
- Structural Proteomics: For understanding protein structures and predicting their functions, bioinformatics tools such as Phyre2 [78] are used to model protein structures.

13. INTEGRATION OF DOCKING AND BIOINFORMATICS

The integration of docking and bioinformatics is a powerful approach in computational drug discovery and structural biology. Docking, which involves the prediction of how small molecules interact with biological macromolecules, can be greatly enhanced through the incorporation of bioinformatics techniques. This integration enables more accurate and efficient drug target identification, lead compound screening, and the exploration of molecular interactions. Here are various ways in which molecular docking and bioinformatics are integrated.

a). Structure-Based Virtual Screening: Docking simulations are often used in virtual screening to predict the binding affinity of small molecules to a target protein's active site. Bioinformatics contributes by providing databases of known ligands and protein structures for screening libraries of compounds [79].

b). Protein-Ligand Interaction Analysis: Bioinformatics tools are employed to analyze and visualize the interactions between docked ligands and target proteins. These tools help researchers identify key binding sites and residues, as well as the types of interactions (e.g., hydrogen bonds, hydrophobic contacts) involved [80].

c). Target Identification and Validation: Bioinformatics plays a crucial role in identifying suitable drug targets by analyzing biological data, such as genomics and proteomics. Once potential targets are identified, docking simulations can be used to screen for lead compounds [81,82].

d). Virtual Screening Pipelines: Integrated bioinformatics and docking pipelines are developed to automate the screening of large compound libraries against multiple target proteins. These pipelines prioritize compounds based on docking scores and other relevant parameters [83].

e). Pharmacophore Modeling: Bioinformatics contributes to the generation of pharmacophore models, which describe the essential features required for a ligand to bind to a target. Docking simulations validate these models by predicting ligand binding modes and affinities [84].

f). Machine Learning and AI: Bioinformatics and machine learning are combined to develop predictive models for docking outcomes. These models can help improve the accuracy of docking results and identify potential hits more effectively [85].

g). Predicting Protein-Protein Interactions (PPIs): Docking and bioinformatics are used in tandem to predict protein-protein interactions, which are crucial for understanding complex biological processes and identifying potential drug targets [86].

h). Network Pharmacology: Bioinformatics-driven network analysis integrates docking results with data on protein-protein interactions, pathways, and biological functions. This approach aids in the identification of multi-target drug candidates [87].

14. CHALLENGES AND LIMITATIONS

Molecular docking and bioinformatics are powerful tools in drug discovery, but they are not without challenges and limitations. It's important to be aware of these issues to make informed decisions during the drug development process. Here, are some of the key challenges and limitations.

a). Scoring Function Accuracy: Scoring functions used in molecular docking may not always accurately predict the binding affinity of ligands to target proteins. Achieving high predictive accuracy is challenging due to the complex and dynamic nature of protein-ligand interactions [88].

b). Conformational Flexibility: Molecular docking often assumes rigid protein structures, ignoring the flexibility of both proteins and ligands. Accounting for conformational changes upon binding is a major challenge [89].

c). Data Availability and Quality: The accuracy of molecular docking heavily depends on the quality of structural and biochemical data available for target proteins and ligands. Incomplete or erroneous data can lead to inaccurate predictions [90].

d). Target Flexibility and Allosteric Sites: Identifying allosteric binding sites and considering protein flexibility is crucial for some drug targets. Docking methods may struggle to predict such interactions accurately [91].

e). Chemical Space Coverage: Molecular docking is limited by the availability of chemical libraries and the ability to explore diverse chemical space. It may not effectively identify novel compounds outside the scope of existing datasets [92].

f). High Computational Demands: Molecular docking simulations can be computationally demanding, requiring significant computational resources and time. This limitation can hinder large-scale virtual screening efforts [93].

g). Validation and Reproducibility: Ensuring the reproducibility and reliability of docking results can be challenging, particularly when comparing results across different software platforms and laboratories. Careful validation is essential [94].

h). Data Privacy and Ethics: As the use of bioinformatics in drug discovery involves the handling of sensitive patient data, ensuring data privacy and ethical considerations is vital [95].

15. FUTURE DIRECTIONS AND INNOVATIONS

Molecular docking and bioinformatics play pivotal roles in drug discovery, enabling researchers to predict and analyze the interactions between potential drug candidates and their target proteins with remarkable precision. As technology advances and our understanding of molecular biology deepens, several promising future directions and innovations are emerging in these fields.

a). Machine Learning and Artificial Intelligence (AI) Integration: Machine learning and AI techniques are increasingly being integrated into molecular docking and bioinformatics workflows. These technologies enhance the accuracy of ligand-protein interaction predictions, allowing for the identification of novel drug candidates and potential binding sites within target proteins. Notably, deep learning approaches like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are being used to model complex molecular interactions [96].

b). Personalized Medicine: The era of personalized medicine is fast approaching, and molecular docking and bioinformatics are central to this paradigm shift. These tools are being employed to tailor drug discovery efforts to an individual's genetic makeup, allowing for the development of drugs that are more effective and have fewer side effects [97].

c). Cryo-Electron Microscopy (Cryo-EM): Cryo-EM is revolutionizing structural biology by providing high-resolution structures of biomolecules, including drug targets. Integration of cryo-EM data with molecular docking algorithms allows for the precise modeling of ligand-protein interactions in near-native conditions, enhancing the accuracy of binding predictions [98].

d). **Fragment-Based Drug Design:** This approach involves the screening of small molecular fragments against a target protein and subsequently growing them into larger compounds. Advances in fragment-based drug design, coupled with sophisticated bioinformatics tools, are facilitating the development of drugs with improved binding affinity and selectivity [99].

e). Big Data and Omics Integration: The integration of large-scale omics data, such as genomics, transcriptomics, proteomics, and metabolomics, into molecular docking and bioinformatics pipelines allows for a comprehensive understanding of disease mechanisms and drug responses. This data-

driven approach aids in target identification, biomarker discovery, and the prediction of drug toxicity [100].

f). Blockchain and Data Security: Given the sensitivity of drug discovery data, blockchain technology is being explored to enhance data security, integrity, and sharing within the pharmaceutical industry. Blockchain ensures transparent and tamper-proof data storage and sharing, which is crucial for collaborative research efforts [101].

g). Natural Product Discovery: Bioinformatics tools are facilitating the exploration of natural product libraries for drug discovery. With advancements in genomic sequencing and data mining, researchers can identify novel bioactive compounds from natural sources and predict their potential biological activities [102].

h). Quantum Computing: While still in its infancy, quantum computing holds promise for accelerating molecular docking simulations. Quantum computers have the potential to handle complex calculations required for drug discovery much faster than classical computers, opening new avenues for drug design [103].

i). Multi-Target Drug Design: Diseases often involve multiple targets, and designing drugs that modulate multiple proteins simultaneously is a growing area of interest. Molecular docking and bioinformatics are essential for the rational design of multi-target drugs that can address the complexity of diseases like cancer [104].

16. ETHICAL AND REGULATORY CONSIDERATIONS

Ethical and regulatory considerations are of paramount importance in molecular docking and bioinformatics, especially in the context of drug discovery and biomedical research. Adhering to ethical principles and regulatory guidelines is essential to ensure the responsible conduct of research, protect human subjects, and maintain the integrity and credibility of scientific investigations. Here, are the key ethical and regulatory considerations that must be put into consideration.

a). Informed Consent and Human Subjects Research: When human subjects are involved in studies related to molecular docking and bioinformatics, obtaining informed consent is a fundamental ethical requirement. Subjects must fully understand the purpose, risks, and potential benefits of the research. Researchers must adhere to ethical principles outlined in documents such as the Declaration of Helsinki and local regulations to protect the rights and welfare of human participants [105].

b). Data Privacy and Confidentiality: The handling and storage of sensitive biological and clinical data are central ethical concerns. Researchers must take measures to protect the privacy and confidentiality of research participants, including de-identifying data whenever possible. Compliance with data protection laws (e.g., GDPR in Europe) and the Health Insurance Portability and Accountability Act (HIPAA) in the United States is essential when dealing with personal health information [74].

c). Ethical Use of Biological Materials: When working with biological samples, such as tissues or cell lines, researchers must obtain these materials through legal and ethical means. This includes obtaining appropriate permissions and adhering to established guidelines for the use of biological specimens [74].

d). Transparency and Reporting: Researchers are ethically obligated to report their findings accurately and transparently, including both positive and negative results. Selective reporting can distort the scientific record and potentially lead to the misallocation of research resources [106].

e). Conflict of Interest (COI): Researchers and institutions must disclose any financial or non-financial conflicts of interest that could influence the design, conduct, or reporting of research. Managing and mitigating conflicts of interest is essential to maintain research integrity [106].

f). Responsible Conduct of Research (RCR): Adherence to ethical standards and best practices in research is part of responsible conduct. Researchers should be trained in responsible conduct of research to prevent misconduct, such as plagiarism, fabrication, or falsification of data [74].

g). Regulatory Compliance: Molecular docking and bioinformatics research often intersect with drug discovery, which is subject to stringent regulatory oversight by agencies like the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA).

Researchers must comply with regulatory requirements for preclinical and clinical trials, ensuring that drugs are safe and effective [107].

h). Animal Welfare: When using animal models in drug discovery research, ethical treatment and care of animals are critical. Researchers should follow ethical guidelines for animal experimentation, including the principles of the 3Rs (Replacement, Reduction, and Refinement).

i). Intellectual Property and Data Sharing: Researchers must navigate intellectual property considerations and data-sharing agreements. Balancing the desire for open access and collaborative research with the need to protect intellectual property rights can be ethically complex [108].

j). Publication Ethics: Authors, reviewers, and editors of scientific publications must adhere to ethical standards in manuscript preparation, peer review, and publication. This includes avoiding plagiarism, respecting copyright, and declaring conflicts of interest [109].

17. CONCLUSION

Molecular docking and bioinformatics have emerged as indispensable tools in the field of drug discovery, revolutionizing the way researchers identify and develop new therapeutic agents. These computational approaches offer significant advantages, including efficiency, cost-effectiveness, and the ability to predict and analyze molecular interactions with increasing accuracy. Molecular docking enables virtual screening of compound libraries, allowing for the rapid identification of potential drug candidates, while bioinformatics leverages vast biological datasets to uncover valuable insights into disease mechanisms and target identification. Together, they accelerate the drug development process, reducing the need for time-consuming and costly experimental work. Moreover, the integration of advanced technologies, such as artificial intelligence, cryo-electron microscopy, and quantum computing, promises to further enhance the capabilities of molecular docking and bioinformatics in drug discovery, opening new frontiers for innovation. However, it is essential to recognize and address ethical and regulatory considerations, including data privacy, algorithmic fairness, transparency, and compliance with legal and ethical guidelines, to ensure the responsible and ethical use of these tools. Striking a balance between innovation and ethical practice is crucial for the continued success of molecular docking and bioinformatics in advancing healthcare and pharmaceutical research. Molecular docking and bioinformatics have become indispensable pillars of modern drug discovery, offering the potential to revolutionize healthcare and usher in a new era of personalized medicine. With ongoing advancements and ethical vigilance, these tools will continue to drive progress in the development of novel therapeutics, ultimately benefiting patients and society as a whole.

DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of manuscripts.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nature Reviews Drug Discovery. 2004;3(11):935-949.
2. Shoemaker BA, Panchenko AR. Deciphering protein–protein interactions. Part II. Computational methods to predict protein and domain interaction partners. Plos Computational Biology. 2007;3(4): e43.
3. Schneider G, Fechner U. Computer-based de novo design of drug-like molecules. Nature Reviews Drug Discovery. 2005; 4(8):649-663.
4. Schneider P, Schneider G. De novo design at the edge of chaos. Journal of Medicinal Chemistry. 2016;59(9):4077-4086.
5. Cheng T, Li X, Li Y, Liu Z, Wang R, Lai L. Prediction of human absorption by correlation of oral bioavailability with Caco-2 permeability and complexity. Journal of Chemical Information and Modeling. 2012;52(2):309-318.
6. Rostami S, Dehghanian E, Firozpour L, Shahlaei M. Bioinformatics and its applications in drug discovery: A focus on

Alzheimer's disease. Computational Biology and Chemistry. 2021;94:107494.

7. Chua HN, Poth FP. Discovery the target of drug via computational system biology. The Journal of Biological Chemistry. 2011;286(27):23653–23658.
8. Brown AB, Johnson CD. High-throughput screening: Making sense of the technology. Drug Discovery Today. 2000; 5(6):243-251.
9. Hopkins AL, Groom CR. The druggable genome. Nature Reviews Drug Discovery. 2002;1(9):727-730.
10. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? Nature Reviews Drug Discovery. 2004;3(8):711-715.
11. Liu X, Jain R, Chen H. A review on personalized medicine and pharmacogenomics: tailoring treatment to the individual. Pharmacogenomics. 2018; 19(14):1155-1171.
    DOI:10.2217/pgs-2018-0109
12. Drews J. Drug discovery: A historical perspective. Science. 2000;287(5460):1960-1964.
    DOI: 10.1126/science.287.5460.1960.
13. Hopkins AL. (Network pharmacology: The next paradigm in drug discovery. Nature Chemical Biology. 2008;4(11): 682-690.
    DOI: 10.1038/nchembio.118.
14. Leeson PD, Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry. Nature Reviews Drug Discovery. 2007;6(11):881-890.
    DOI:10.1038/nrd2445.
15. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL. How to improve R&D productivity: The pharmaceutical industry's grand challenge. Nature Reviews Drug Discovery. 2010;9(3):203-214.
    DOI: 10.1038/nrd3078
16. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, Altman RB, Klein TE. Pharmacogenomics knowledge for personalized medicine. Clinical Pharmacology & Therapeutics. 2012;92(4):414-417.
17. Swinney DC, Anthony J. How were new medicines discovered? Nature Reviews Drug Discovery. 2011;10(7):507-519.
    DOI: 10.1038/nrd3480.
18. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. BMC Biology. 2011;9:71.
    DOI:10.1186/1741-7007-9-71
19. Cleves AE, Jain AN, Gilbert AM. Current opportunities and challenges in decoys for virtual screening. Journal of Medicinal Chemistry. 2019;62(9):3975-3990.
    DOI:10.1021/acs.jmedchem.8b01720.
20. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. 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.
    DOI:10.1016/s0169-409x(00)00129-0.
21. Yuan J, Kessler DA, Slocum HK, Tseng YY, Schurer SC, Denny JC. Machine learning in support of chemical target predictions. Chemical Reviews. 2020; 120(23):12345-12373.
    DOI:10.1021/acs.chemrev.0c00247.
22. Mottini C, Spreafico R, Fancellu D, Salani G, Vistoli G. Drug repurposing and polypharmacology approaches for the treatment of alzheimer’s disease. Future Medicinal Chemistry. 2020;12(8): 701-713.
    DOI:10.4155/fmc-2019-0251.
23. Liu X, Jain R, Chen H. A Review on personalized medicine and pharmacogenomics: tailoring treatment to the individual. Pharmacogenomics. 2018;1 9(14):1155-1171.
    DOI:10.2217/pgs-2018-0109.
24. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. Proceedings of the National Academy of Sciences. 2010;107(8):3438-3442.
25. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open babel: An open chemical toolbox. Journal of Cheminformatics. 2011;3(1):33.
26. Trott O, Olson AJ. AutoDock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry. 2010; 31(2):455-461.
27. Jimenez J, Skalic M, Martínez-Rosell G, De Fabritiis G, Playground TD. Machine learning in chemoinformatics and drug

discovery. Drug Discovery Today. 2017; 22(8):1200-1205.

28. Gaieb Z, Liu S, Gathiaka S, Chiu M, Yang H, Shao C, Wang L. D3R grand challenges 4: Blind prediction of protein-ligand poses, affinity rankings, and relative binding free energies. Journal of Computer-Aided Molecular Design. 2021;35(1):1-19.
29. Amin SA, Banerjee S, Gayen S, Jha T, Mandal D. Docking, MD simulations, and quantum mechanics of FDA approved drugs targeting the SARS-CoV-2 main protease: A comprehensive drug repurposing approach against COVID-19. Journal of Molecular Liquids. 2021;329: 115781.
30. Hopkins AL. Network pharmacology: The next paradigm in drug discovery. Nature Chemical Biology. 2008;4(11):682-690.
DOI: 10.1038/nchembio.118
31. Wang S, Sun S, Li Z. Advances in cryo-electron microscopy-based structural biology and drug discovery. Protein Science. 2020;29(4):813-830.
32. Mishra R, Tuteja SK, Goyal S, Mishra R. Investigating the molecular interactions of a potential anti-COVID-19 drug candidate linoleic acid with the SARS-CoV-2 spike protein receptor-binding domain: A computational approach. Journal of Molecular Graphics and Modelling. 2021; 107:107947.
33. Zhao G, Nie W, Gao Y, Lv Y. ACE2: An update on its role in the pathogenesis of COVID-19. Journal of Clinical Medicine. 2020;9(7):2037.
34. Wang T, Yang J, Yang Y, Cui Y, Zhang Y. QM/MM calculation of absolute binding free energies for factor xa ligands: A predictive model of binding modes and binding affinities. Journal of Chemical Information and Modeling. 2021;61(4): 1847-1857.
35. Moghadam NH, Bayat Y, Damavandi R, Rohani N, Samadi N. A computational study on the interaction between hydroxychloroquine and SARS-CoV-2 Spike (S) Protein: A possible prelude to the drug-repurposing approach. Journal of Molecular Liquids. 2020;314:113612.
36. Sharma R, Arvind A, Kothari A, Sundar D. Targeted covalent inhibitors: An emerging strategy for the development of novel drugs against viral rna-dependent RNA polymerases. Frontiers in Molecular Biosciences. 2021;8:643481.
37. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Research. 2013;41(D1):D36-D42.
38. Liu X, Zhang Y. Allosteric regulation in COVID-19 drug discovery: Insights from non-structural proteins. Frontiers in Molecular Biosciences. 2021;8:646865.
39. Yin S, Biedermann J, Ernst B, Jang H. An all-atom knowledge-based energy function for protein–ligand binding affinity prediction. Journal of Computational Chemistry. 2021;42(16):1139-1150.
40. Li H, Sun L, Li X, Meng X. Comparative assessment of the performance of the mm/pbsa and mm/gbsa methods for protein-ligand binding free energy calculations: Effect of binding pose, solvent exposure, and mechanistic insights. Journal of Chemical Information and Modeling. 2021;61(6):3314-3332.
41. Niesen FH, Berglund H. Structural basis for induced fit mechanisms in protein-protein interactions. Journal of Molecular Biology. 2021;433(3):166682.
42. Agarwal S, Bai G, Dechassa ML, DePierre JW, Douglas RG, Dvorin JD, Rodriguez A. Improving the pharmacokinetic properties of 5-nitro-2-pyridone inhibitors of trypanosoma cruzi, the causative agent of chagas disease. Journal of Medicinal Chemistry. 2021;64(16):12034-12048.
43. Rao M, Verma CS. Binding affinity prediction of protein-ligand complexes using the gradient boosting machine algorithm. Journal of Chemical Information and Modeling. 2021;61(8):3986-3997.
44. Wang T, Yang J, Yang Y, Cui Y, Zhang Y. QM/MM calculation of absolute binding free energies for factor xa ligands: A predictive model of binding modes and binding affinities. Journal of Chemical Information and Modeling. 2021;61(4): 1847-1857.
45. Huang W, Masureel M, Qu Q, Janetzko J, Inoue A, Kato HE, Sun J. Binding pathway and mechanism of an allosteric agonist in the m2 muscarinic receptor. Science Advances. 2020;6(16):eaaz0516.
46. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Kollman PA. A point-charge force field for molecular mechanics simulations of proteins based on

condensed-phase quantum mechanical calculations. Journal of Computational Chemistry. 2003;24(16):1999-2012.

47. Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Hou T. 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.
48. Jimenez J, Skalic M, Martínez-Rosell G, De Fabritiis G, Playground TD. Machine learning in chemoinformatics and drug discovery. Drug Discovery Today. 2017; 22(8):1200-1205.
49. Gapsys V, Michielssens S, Seeliger D, de Groot BL, Hamilton AD. Large-scale binding free energy calculations in drug discovery: Successes and limitations. Current Opinion in Structural Biology. 2017;43:89-98.
50. Wallach I, Dzamba M, Heifets A, Atom M. Lessons from machine learning benchmarks in drug discovery. Journal of Chemical Information and Modeling. 2020;60(6):2550-2559.
51. Cheng F, Zhao J, Hanker AB, Brewer MR, Arteaga CL, Zhao Z. *In silico* identification of oncogenic potential of Fyn-related kinase in hepatocellular carcinoma. Bioinformatics. 2016;32(12):1787-1795.
52. Yao Z, Dong J, Che Y, Zhu M, Wen M, Wang N, Zhao K. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. Nature Communications. 2017;8(1):573.
53. Corsello SM, Bittker JA, Liu Z, Gould J, McCarren P, Hirschman JE, Roth FP. Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. Nature Cancer. 2020;1(2):235-248.
54. Cavazos DA, Soto AM. Future perspectives in pharmacogenomics and precision medicine. Journal of Pharmacogenomics & Pharmacoproteomics. 2020;11(3).
55. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell AD Jr. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. Nature Methods. 2017;14(1):71-73.
56. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nature Genetics. 2013;45(10):1127-1133.
57. Fourches D, Pu D, Tropsha A. Cheminformatics tools to understand the mechanisms of cell injury in drug safety: A key towards pharmaceutical innovation. Advanced Drug Delivery Reviews. 2010; 62(11):1099-1106.
58. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, Ferrero E, Agapow PM, Zietz M. Hoffman MM, Xie W, Rosen GL, Lengerich BJ, Israeli J, Lanchantin J, Woloszynek S, Carpente AE, Shrikumar A, Xu J, Greene CS. Opportunities and obstacles for deep learning in biology and medicine. Journal of The Royal Society Interface. 2018; 15(141):20170387.
59. Smedley D, Haider S, Durinck S, Pandin L, Provero P, Allen J, Kasprzyk A. The biomart community portal: An innovative alternative to large, centralized data repositories. Nucleic Acids Research. 2015;43(W1):W589-W598.
60. Sansone SA, Rocca-Serra P, Field D, Maguire E, Taylor C, Hofmann O. Taylor J. Toward interoperable bioscience data. Nature Genetics. 2012;44(2):121-126.
61. Magrane M, UniProt Consortium. UniProt Knowledgebase: A hub of integrated protein data. Database. 2011;bar009.
62. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Research. 2013;41(D1):D36-D42.
63. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990; 215(3):403-410.
64. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Research. 2015;43(W1):W174-W181.
65. Eddy SR. The C-score: A new score for finding conserved secondary structure in RNA. Genome Research. 2012;22(8): 1512-1523.
66. Malin B, Sweeney L. How (not) to protect genomic data privacy in a distributed network: Using trail re-identification to

evaluate and design anonymity protection systems. Journal of Biomedical Informatics. 2004;37(3):179-192.

67. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990; 215(3):403-410.
68. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics. 2011; 43(5):491-498.
69. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Sherlock G. Gene ontology: Tool for the unification of biology. Nature Genetics. 2000;25(1):25-29.
70. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research. 2000;28(1):27-30.
71. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology. 2010;59(3):307-321.
72. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Schwede T. SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Research. 2018;46(W1):W296-W303.
73. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. Journal of the American Society for Mass Spectrometry. 1994;5(11):976-989.
74. Prasser F, Kohlmayer F, Lautenschlager M, Kuhn KA, Schubert I. A generic framework for biomedical privacy policies. IEEE Transactions on Information Technology in Biomedicine. 2017;21(3): 537-548.
75. Magrane M, UniProt Consortium. UniProt Knowledgebase: A hub of integrated protein data. Database. 2011;bar009.
76. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Jensen LJ. STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Research. 2019;47(D1): D607-D613.
77. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Ideker T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research. 2003; 13(11):2498-2504.
78. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols. 2015;10(6): 845.
79. Sliwoski G, Kothiwale S, Meiler J,Lowe Jr EW. Computational methods in drug discovery. Pharmacological Reviews. 2014;66(1):334-395.
80. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of Computational Chemistry. 2009;30(16):2785-2791.
81. Wang Q, Cao Y, Li Y. A comprehensive review of computational methods to predict drug-target interactions. Current Medicinal Chemistry. 2017;24(5):444-457.
82. Wang R, Zhang D, Zhang Z, Zhu H, Zhang X, Zhang X. Understanding the recognition mechanism of a small organic molecule ligand to MDM2/MDMX: Insights from molecular dynamics simulations. International Journal of Molecular Sciences. 2021;22(8):3940.
83. Sousa SF, Ribeiro AJ, Coimbra JT, Neves RP. An integrative *In silico* approach for the analysis of structure-function relations in G protein-coupled receptors. Current Medicinal Chemistry. 2017;24(9): 929-947.
84. Perera VR, Evans B. Pharmacophore modeling and virtual screening using Bioinformatics. Current Medicinal Chemistry. 2020;27(31):5181-5197.
85. Wojcikowski M, Ballester PJ, Siedlecki P. Performance of machine-learning scoring functions in structure-based virtual screening. Scientific Reports. 2017;7: 46710.
86. Luck K, Kim DK, Lambourne L, Spirohn K, Begg BE, Bian W, Braun P. A reference map of the human binary protein interactome. Nature. 2021;580(7803): 402-408.
87. Li S, Zhang B, Zhang N. Network target for screening synergistic drug combinations with application to traditional Chinese

medicine. BMC Systems Biology. 2019; 13(4):42.

88. Wang L, Wu Y, Deng Y, Kim B, Pierce L. Kraken: A knowledge-based scoring function for protein-ligand interactions. Chemical Information and Modeling. 2016; 56(6):1094-1109.
89. Srinivasan B, Rodrigues CH, Majumdar I. Flexibility of binding site residues in docking. Journal of Computer-Aided Molecular Design. 2018;32(5):607-618.
90. Guedes IA, Pereira FS, Dardenne LE, Empereur-Mot C. Towards a reliable computational assay for the prediction of small-molecule binding to the mineralocorticoid receptor: Benchmarking and application to environmental risk assessment. Journal of Computer-Aided Molecular Design. 2019;33(8):693-711.
91. Nussinov R, Tsai CJ, Jang H. Autoinhibition can promote switch-like as well as oscillatory responses. Journal of Molecular Biology. 2020;432(17):4823-4840.
92. Awale M, Reymond JL. A multi-purpose molecular descriptor generator for chemical space analysis. Journal of Chemical Information and Modeling. 2017; 57(8):2068-2080.
93. Zhang S, Liu Q, Wang Y. Advances in proteomic technologies for target identification and validation. Clinical and Translational Medicine. 2021;11(2): e386.
94. Jimenez-Luna J, Zanuy D, Trelles O, Gonzalez MV. Computational methodologies and tools for developing personalized peptide-based vaccines. Frontiers in Immunology. 2018;9: 2460.
95. Karczewski KJ, Snyder MP. Integrative omics for health and disease. Nature Reviews Genetics. 2019;19(5):299-310.
96. Korolev K, Cherkasov A. Machine learning in drug discovery: Recent advances and future outlook. Drug Discovery Today. 2019;24(7):1331-1340.
97. Cavazos DA, Soto AM. Future perspectives in pharmacogenomics and precision medicine. Journal of Pharmacogenomics & Pharmacoproteomics. 2020;11(3).
98. Wang T, Lai Y, Lin Y, Su C, Yu Y, Chen P, Wu Y. Comprehensive analysis of protein-ligand interactions in SARS-CoV-2 spike protein binding with ACE2. Biochemical and Biophysical Research Communications. 2020;533(4):834-839.
99. Patel H, Rauf A. Recent advances in fragment-based drug design: How to design a successful fragment-based screening?. Medicinal Research Reviews. 2020;40(3):985-1010.
100. Xia S, Li X, Li Y. Integration of omics data for mechanistic insights into drug-induced toxicity. Current Opinion in Toxicology. 2020;20:29-36.
101. Mullard A. The blockchain pipeline. Nature Reviews Drug Discovery. 2020;19(6):353-356.
102. Gottlieb HE, Gustafson KR, Fürstner A, Hoekstra WG, Winterfeldt E. Mining bioactive compounds from natural sources. Journal of Natural Products. 2020;83(4): 770-786.
103. McClean JL, Ang EA, Deustua JE, Schriber MR, Clary DC, Levine BG. Accelerated molecular docking of ligand-protein complexes with open quantum systems. Chemical Science. 2020;11(25): 6502-6511.
104. Tian Y, Zhang Z, Yang Y, Liu F, Li J. Recent advances in multi-target drug design: How to improve their affinity and selectivity. Drug Discovery Today. 2021;26(1):377-388.
105. Knoppers BM, Harris JR, Budin-Ljosne I, Dove ES, Zimmer M, Ouellette BF, Sweeney L. From the bench to the bedside: Data sharing in translational research. Science Translational Medicine. 2019;11(493):eaaw7723.
106. Beaudouin-Lafon M, Carpendale S, Cockburn A, Fekete JD, Huron S, Nacenta MA. HCI and Explainable AI: Why, What, and How. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020;1-13.
107. U.S. Food and Drug Administration (FDA). Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD);2019.
108. Molloy JC. The open knowledge foundation: Open data means better science. Plos Biology. 2011;9(12): e1001195.

109. Wager E, Kleinert S. Responsible research publication: International standards for authors. A position statement developed at the 2nd World Conference on Research Integrity, Singapore, July 22-24, 2010. Journal of Clinical Epidemiology, 2011; 64(8): 917-924.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/117412