CADD in the 21st Century: Emerging Problems and Opportunities

Kiran Yadav[1]*, Sandhya Jaishwal[1], Narra Kishore[2], Sunil Kumar[3]

1-Chandigarh College of Pharmacy, Chandigarh group of Colleges, Landran, Mohali, Punjab, India- 140307
2- Centre of Molecular Medicine and Diagnostics (COMManD), Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India-600077
3- Department of Pharmaceutical Sciences, Indira Gandhi University, Meerpur, Rewari, Haryana, India- 122502

*corresponding author-Dr. Kiran Yadav (kiranyadav24@yahoo.co.in)

Abstract:

The use of computational tools in drug discovery has revolutionized the field of pharmaceutical research and development. Computer-Aided Drug Design (CADD) is a powerful approach that uses computational methods to identify and optimize potential drug candidates. However, as the complexity of drug targets increases, the challenges associated with CADD also grow. This review highlights the emerging problems and opportunities in CADD in the 21st century. We discuss the current state-of-the-art methods in CADD, including machine learning, data mining, and molecular dynamics simulations. We also identify the key challenges facing CADD, including the need for more accurate models and the integration of multi-omics data. Finally, we explore the opportunities presented by new technologies, such as quantum computing and artificial intelligence, and their potential impact on the future of CADD.

Table of Contents:

- Ethical and Legal Issues

IV. Emerging Technologies and Opportunities

- Quantum Computing
- Artificial Intelligence
- Big Data Analytics
- Virtual and Augmented Reality

V. Future Directions and Conclusion

- The Role of CADD in the 21st Century
- Potential Impact of Emerging Technologies
- Future Research Directions
- Conclusion

Introduction

The field of drug discovery has always been challenging, time-consuming, and costly. The traditional drug discovery process involves several stages, including target identification, lead discovery, lead optimization, preclinical testing, and clinical trials. This process can take up to 15 years and cost billions of dollars before a drug can be approved for market use [1-3]. CADD has revolutionized the drug discovery process by reducing the time and cost associated with drug development and improving the efficiency and accuracy of drug design. CADD uses a combination of computational methods, such as molecular docking, virtual screening, and molecular dynamics simulations, to predict the binding affinity of potential drug candidates to their target protein and optimize their properties [4]. The use of CADD in drug discovery has become even more important in the 21st century due to the increasing complexity of drug targets and the need for more personalized and precision medicine. The advent of multi-omics data, such as genomics, proteomics, metabolomics, and transcriptomics, has enabled researchers to identify novel drug targets and biomarkers for drug efficacy and safety [5]. The integration of multi-omics data with CADD can facilitate the design of more effective and targeted drugs.

However, the challenges associated with CADD have also increased with the growing complexity of drug targets. The accuracy and reliability of models used in CADD have become a major concern, as inaccuracies can lead to false-positive or false-negative predictions of drug efficacy and safety. The limited availability of experimental data for model validation is also a challenge in CADD. Additionally, ethical and legal issues related to the use of CADD, such as intellectual property and data sharing, must be addressed [6]. Despite these challenges, the emergence of new technologies, such as quantum computing and artificial intelligence, presents exciting opportunities for CADD. These technologies can provide more accurate and efficient computational models and enable the analysis of big data [6,7]. Virtual and augmented reality can also enhance drug design and facilitate the exploration of drug-target interactions.

CADD is a powerful tool that has transformed the drug discovery and development process. However, as the complexity of drug targets increases, the challenges associated with CADD also grow. The integration of multi-omics data and the emergence of new technologies offer exciting opportunities for the future of CADD in drug discovery and development.

II. State-of-the-Art Methods in CADD:

The use of computational methods in CADD has evolved significantly over the past few decades. The current state-of-the-art methods in CADD include machine learning, data mining, molecular dynamics simulations, and hybrid approaches.

1. Machine Learning:

Machine learning is a subfield of artificial intelligence that involves the development of algorithms and statistical models that enable computers to learn from and make predictions or decisions based on data. In the context of CADD, machine learning algorithms can be trained on large datasets of known ligand-protein complexes to predict the binding affinity of new ligands to their target protein [8, 9].

Machine learning algorithms have been increasingly used in CADD due to their ability to analyze large datasets and identify complex patterns that may not be apparent to human researchers. They can learn from diverse data sources, such as chemical structures, protein sequences, and functional annotations, and can integrate multiple sources of data to improve the accuracy of predictions [10].

Several machine learning algorithms have been developed for CADD, including support vector machines, random forests, and deep neural networks. Support vector machines are a type of supervised learning algorithm that separates data into distinct classes, while random forests are an ensemble learning method that combines multiple decision trees to improve the accuracy of predictions. Deep neural networks are a type of artificial neural network that are inspired by the structure and function of the human brain. They can learn complex non-linear relationships between ligand and protein features and provide highly accurate predictions [9,11].Deep neural networks have shown promising results in CADD. For example, a recent study demonstrated that a deep neural network could accurately predict the binding affinity of small molecules to their target protein, outperforming other machine learning algorithms such as random forests and support vector machines [11].

The use of machine learning in CADD holds great promise for improving the efficiency and accuracy of drug discovery. By leveraging large datasets and complex algorithms, researchers can identify new drug candidates and optimize their properties, potentially leading to faster and more effective drug development.

2. Data Mining:

Data mining is a process of extracting hidden knowledge or patterns from large datasets through statistical and computational methods. In the field of computer-aided drug design (CADD), data mining plays a vital role in identifying new drug targets, predicting the efficacy and safety of drugs, and analyzing the structure-activity relationships of ligands.

One of the main applications of data mining in CADD is the identification of new drug targets. Data mining techniques, such as clustering, can be used to group similar proteins or genes that are involved in specific diseases or pathways [12]. By analyzing these clusters, researchers can identify potential drug targets that can be used to develop new therapies.

Another application of data mining in CADD is the prediction of drug efficacy and safety. Machine learning algorithms can be trained on large datasets of drug molecules and their properties to predict the therapeutic outcomes and side effects of new drug candidates. This approach can help reduce the time and cost required for drug development and clinical trials [13].

Data mining can also be used to analyze the structure-activity relationships (SAR) of ligands. SAR analysis involves studying the chemical and biological properties of molecules to understand how their structure affects their activity. By applying data mining techniques such as association rule mining, researchers can identify key chemical features that are associated with drug efficacy or toxicity [14-16].

3. Molecular Dynamics Simulations:

Molecular dynamics (MD) simulations are a computational technique used to investigate the behavior of biological molecules, such as proteins and nucleic acids, at an atomic level. In the field of computer-aided drug design (CADD), molecular dynamics simulations can be used to study the binding affinity and stability of ligand-protein complexes, and to optimize the properties of potential drug candidates [17,18]. MD simulations involve the calculation of the motion of atoms and molecules over time, using classical mechanics. The simulations take into account the interactions between atoms and molecules, such as electrostatic and van der Waals forces, to simulate the behavior of the system in question. By simulating the dynamics of the ligand-protein complex, MD simulations can provide insights into the binding affinity, stability, and conformational changes of the complex [17, 19].

In CADD, MD simulations can be used to study the binding of ligands to their protein targets. By simulating the motion of the ligand and protein over time, researchers can gain insights into the binding affinity and stability of the complex. MD simulations can also be used to identify key protein residues that are involved in ligand binding, which can be used to optimize the design of potential drug candidates [19].

Another application of MD simulations in CADD is the optimization of the properties of potential drug candidates. By simulating the motion of the ligand and protein over time, researchers can gain insights into the interactions between the ligand and protein, and identify areas for improvement. For example, MD simulations can be used to optimize the binding affinity and selectivity of potential drug candidates, or to improve the pharmacokinetic properties of the drug [18, 20].

Hence, molecular dynamics simulations are a powerful technique in CADD that can be used to study the behavior of biological molecules at an atomic level. By simulating the motion of ligand-protein complexes, researchers can gain insights into the binding affinity and stability of the complex, and optimize the properties of potential drug candidates.

4. Hybrid Approaches:

Hybrid approaches in computer-aided drug design (CADD) involve the combination of multiple computational methods to improve the accuracy and efficiency of drug design. One example of a hybrid approach is the use of machine learning algorithms to predict the binding affinity of potential drug candidates, combined with molecular dynamics simulations to study the stability and dynamics of the ligand-protein complex.

Machine learning algorithms can be trained on large datasets of known ligand-protein complexes to identify patterns and relationships in the data that can be used to predict the binding affinity of new ligands. These predictions can be combined with molecular dynamics simulations to study the behavior of the ligand-protein complex over time and gain insights into the stability and dynamics of the complex [21, 22].

Hybrid approaches can also involve the integration of experimental data, such as crystallographic data, with computational methods to improve the accuracy of the models. For example, crystallographic data can be used to generate a high-resolution structure of the ligand-protein complex, which can be used as a starting point for molecular dynamics simulations. The resulting simulations can provide insights into the stability and dynamics of the complex, which can be used to optimize the design of potential drug candidates [21].

Another approach to hybrid modeling in CADD involves the integration of multi-omics data with computational methods. For example, genomic, proteomic, and metabolomic data can be used to identify potential drug targets and to optimize the design of drug candidates. By combining computational methods with multi-omics data, researchers can gain a more comprehensive understanding of the biological system and improve the accuracy and efficiency of drug design [21, 23]. The current state-of-the-art methods in CADD include machine learning, data mining, molecular dynamics simulations, and hybrid approaches. These methods offer powerful tools for drug discovery and development and have the potential to accelerate the drug discovery process. Hence, hybrid approaches in CADD involve combining multiple computational methods to improve the accuracy and efficiency of drug design. These approaches can involve the integration of experimental data, such as crystallographic data, with computational methods, as well as the integration of multi-omics data with computational methods. The current state-of-the-art methods in CADD offer powerful tools for drug discovery and have the potential to accelerate the drug discovery process.

III. Challenges in CADD:

While computational methods in CADD offer powerful tools for drug discovery and development, there are several challenges that need to be addressed to improve their accuracy and reliability. Some of the major challenges in CADD include:

1. Accuracy and Reliability of Models:

One of the major challenges in computer-aided drug design (CADD) is to develop accurate and reliable models for predicting the efficacy and safety of drug candidates. The accuracy of the models depends on several factors such as the quality of the input data, the complexity of the ligand-protein interactions, and the limitations of the computational methods [24, 25].

Second main challenges in developing accurate CADD models is obtaining high-quality input data. This includes accurate structural data for both the protein target and potential drug candidates, as well as experimental data on the activity and toxicity of compounds. Obtaining high-quality data can be particularly challenging when working with complex biological systems, such as multi-protein complexes, or when working with less well-characterized targets [24-26].

Another challenge in CADD is the complexity of the ligand-protein interactions. The binding of a ligand to a protein target can involve a range of interactions, including electrostatic, hydrogen bonding, hydrophobic, and van der Waals interactions. Predicting the binding affinity of a ligand to a protein target requires accurate modeling of all of these interactions, which can be challenging given the complexity of the underlying physics [23, 26].

In addition, the accuracy of CADD models can also be limited by the computational methods used. For example, molecular dynamics simulations can provide valuable insights into the dynamics of ligand-protein complexes, but are computationally expensive and may not be feasible for large-scale virtual screening studies. Machine learning algorithms can provide a more efficient alternative, but may be limited by the quality and quantity of training data [23-28].

To improve the accuracy and reliability of CADD models, researchers are developing new algorithms and computational methods that can more accurately model ligand-protein interactions. For example, advances in quantum mechanics-based methods have enabled more accurate modeling of electrostatic interactions, while machine learning algorithms are being developed that can more accurately predict the binding affinity of potential drug candidates [27, 29]. In addition, the integration of experimental data can also help to improve the accuracy of CADD models. For example, crystallographic data can be used to generate high-resolution structures of the ligand-protein complex, which can be used to improve the accuracy of molecular docking simulations. Similarly, experimental data on the activity and toxicity of compounds can be used to validate and refine CADD models [29]. So, the development of accurate and reliable models for predicting the efficacy and safety of drug candidates is a major challenge in CADD. Improving the accuracy and reliability of these models requires the development of new algorithms and computational methods, as well as the integration of experimental data to improve the accuracy of the models.

2. Integration of Multi-Omics Data:

CADD involves the use of a variety of techniques and data types to facilitate the discovery and development of new drugs. This can include data from genomics, proteomics, metabolomics, and structural biology, among others. Integrating data from multiple sources can help to provide a more complete understanding of the complex biological processes underlying drug efficacy and safety [30-32]. However, integrating data from different sources can be challenging. One major challenge is the development of algorithms that can effectively integrate data from multiple sources while accounting for differences in data types and data quality [33]. Another challenge is the need for standardization of data formats and ontologies, so that data can be effectively shared and compared across different research groups and institutions. In addition, the integration of multi-omics data requires large-scale data sharing and collaboration, which can be hindered by issues related to data privacy and

intellectual property [34]. These challenges are being addressed through the development of new computational methods and the establishment of standards and guidelines for data sharing and collaboration. Overall, the integration of multi-omics data holds great promise for advancing drug discovery and development.

3. Limited Availability of Experimental Data:

The availability and quality of experimental data are critical factors that impact the accuracy and reliability of CADD models. The limited availability of high-quality experimental data is a major challenge in drug discovery, particularly for new or emerging drug targets. This limitation is due to various factors such as the complexity of biological systems, ethical considerations, and high costs associated with experimental data generation [35]. To address this challenge, researchers are developing new experimental techniques and technologies for generating high-quality data. For example, cryo-electron microscopy (cryo-EM) and X-ray free-electron laser (XFEL) have revolutionized the field of structural biology by enabling the determination of high-resolution structures of biomolecules. Similarly, advancements in next-generation sequencing (NGS) technologies have made it possible to generate large-scale omics data. Moreover, the development of data sharing and collaboration platforms is essential for facilitating the sharing and integration of experimental data. Several public databases and platforms, such as the Protein Data Bank (PDB) and the Gene Expression Omnibus (GEO), provide access to large-scale experimental data. Collaborative efforts such as the Cancer Genome Atlas (TCGA) and the Human Protein Atlas (HPA) have enabled the generation and sharing of multi-omics data [36-38].

4. Ethical and Legal Issues:

CADD raises several ethical and legal issues, particularly with regard to the use of human data and the protection of intellectual property rights. The use of human data in drug discovery and development raises ethical concerns regarding privacy, confidentiality, and informed consent. Additionally, the development of CADD models and drug candidates raises legal issues regarding the protection of intellectual property rights, particularly in the case of collaborations between academic and industry partners.

The challenges in CADD include improving the accuracy and reliability of models, integrating multi-omics data, addressing the limited availability of experimental data, and addressing ethical and legal issues. Addressing these challenges requires a multidisciplinary approach that involves collaborations between researchers, clinicians, and policymakers [39].

IV. Emerging Technologies and Opportunities:

Advancements in emerging technologies have the potential to revolutionize CADD, providing new opportunities for drug discovery and development. Some of the emerging technologies that are gaining popularity in CADD include:

Quantum Computing:

Quantum computing is based on the principles of quantum mechanics, which allow quantum bits (qubits) to exist in multiple states simultaneously. This property allows quantum computers to perform certain types of calculations much faster than classical computers. In CADD, the ability to perform fast and accurate simulations of large molecules and their

interactions is crucial for the design of effective drugs [40]. Quantum computing can be used to simulate the behavior of large molecules and their interactions with drug candidates through quantum chemistry calculations, which can provide more accurate predictions of the properties of molecules and their interactions with other molecules. For example, quantum computing can be used to simulate the behavior of a protein-ligand complex, which can provide information on the binding affinity, conformational changes, and stability of the complex[40,41].

Quantum machine learning algorithms can also be used in CADD to analyze large datasets of molecular structures and predict their properties. These algorithms use quantum computing to perform complex calculations that are beyond the capabilities of classical computing, such as the analysis of high-dimensional data and the identification of non-linear relationships between variables.While quantum computing has the potential to revolutionize CADD, it is still in its early stages and faces several challenges, including the development of hardware and software, the optimization of quantum algorithms, and the integration of quantum computing with classical computing. However, with continued research and development, quantum computing has the potential to greatly accelerate drug discovery and development [40-42].

Artificial Intelligence:

Artificial intelligence (AI) has been increasingly used in drug discovery and development, including CADD. AI has been utilized in various stages of drug discovery, such as lead optimization, hit identification, and toxicity prediction. Machine learning, deep learning, and natural language processing are examples of AI techniques used in CADD. Deep learning algorithms, in particular, have shown great promise in predicting the binding affinity of drug candidates to target proteins, as well as identifying potential new drug targets. These algorithms can analyze vast amounts of data and identify patterns and relationships that may not be apparent through traditional methods. Furthermore, AI can be used to optimize drug dosages by analyzing patient data, such as genetics and medical history, and developing personalized treatment plans. This personalized approach has the potential to improve treatment outcomes and reduce adverse effects [43].

Big Data Analytics:

Big data analytics in CADD involves the use of advanced computational and statistical methods to analyse large and complex datasets, such as genomic, proteomic, and metabolomic data, to identify patterns and correlations that can inform drug discovery and development. With the vast amount of data available, big data analytics can help to accelerate the drug discovery process and improve the accuracy of drug design.

In addition to identifying new drug targets and predicting drug toxicity, big data analytics can also be used to optimise drug dosages for improved efficacy and safety. For example, pharmacogenomics data can be analysed to identify genetic variations that affect an individual's response to a particular drug, allowing for personalised dosing regimens.So, big data analytics is a powerful tool for CADD that has the potential to transform the drug discovery process. By leveraging the power of advanced computational and statistical

methods, researchers can gain deeper insights into complex biological systems and develop more effective drugs [44].

Virtual and Augmented Reality:

Virtual and augmented reality (VR/AR) technologies are becoming increasingly popular in drug discovery and development due to their ability to provide a more intuitive and interactive visualization of complex molecular structures. In CADD, researchers can use VR/AR to design and optimize drug candidates, simulate drug-target interactions, and visualize drug mechanisms of action.

One key advantage of VR/AR in CADD is the ability to manipulate molecular structures in real-time, allowing researchers to observe the effects of structural changes on drug activity and selectivity. This can help researchers to design more effective drugs with fewer side effects. Another advantage of VR/AR in CADD is the ability to simulate drug-target interactions in a more realistic way than traditional computational methods. This can help researchers to identify potential drug candidates that are more likely to bind to the target protein and have a therapeutic effect [45-47].

One example of the use of VR/AR in CADD is the development of VR-based drug design platforms that allow researchers to interact with and manipulate molecular structures in real-time. For example, the Molecular Rift platform allows users to visualize and manipulate protein-ligand complexes in virtual reality, providing a more intuitive and interactive approach to drug design.Hence, emerging technologies such as quantum computing, artificial intelligence, big data analytics, and virtual and augmented reality offer new opportunities for drug discovery and development in CADD. These technologies have the potential to revolutionize the field and accelerate the pace of drug discovery and development, providing new treatments for a wide range of diseases [46-49].

V. Future Directions and Conclusion:

The field of CADD is rapidly evolving, with new technologies and methodologies emerging at an unprecedented pace. As we move further into the 21st century, CADD will continue to play a critical role in drug discovery and development. In this section, we discuss the potential impact of emerging technologies and future research directions in CADD.

1. The Role of CADD in the 21st Century:

In the past few decades, CADD has become an integral part of drug discovery and development, and it is expected to continue to grow in importance in the 21st century. The use of computational methods in drug discovery has increased dramatically, with many pharmaceutical companies now using CADD as a key component of their drug discovery programs. With the development of new technologies such as quantum computing and artificial intelligence, CADD has the potential to revolutionize the field, making drug discovery faster, more efficient, and more cost-effective.

One of the main advantages of CADD is the ability to screen large libraries of compounds for potential drug candidates. This approach can significantly reduce the time and cost required

for drug discovery, as it enables researchers to focus their efforts on compounds with the highest potential for success. Additionally, CADD can be used to optimize drug candidates, predict their efficacy and toxicity, and identify new drug targets.

CADD is poised to play an increasingly important role in drug discovery and development in the 21st century. The integration of new technologies such as quantum computing and artificial intelligence is expected to further accelerate the pace of drug discovery, leading to the development of new and more effective treatments for a wide range of diseases.

Potential Impact of Emerging Technologies:

Quantum computing, for example, can be used to simulate the behavior of large molecules and their interactions with drug candidates, allowing researchers to design more effective drugs in less time. It can also aid in predicting the behavior of molecules and optimizing drug dosages. Artificial intelligence can assist in designing new drug candidates, predicting their toxicity and efficacy, and optimizing drug dosages. Deep learning algorithms can predict the binding affinity of drug candidates to target proteins with high accuracy.

Big data analytics is a powerful tool for analyzing large and complex datasets in CADD, allowing researchers to identify patterns and correlations that would be difficult to detect using traditional statistical methods. It can identify new drug targets, predict drug toxicity, and optimize drug dosages. In addition, virtual and augmented reality technologies offer new opportunities for drug discovery and development by allowing researchers to visualize and manipulate complex molecular structures in real-time. VR/AR can be used to design and optimize drug candidates, simulate drug-target interactions, and visualize drug mechanisms of action. With the continued development and advancement of these emerging technologies, CADD is expected to become more efficient and effective in discovering new drugs and developing them into successful treatments. These technologies can also help overcome some of the challenges that the field currently faces, such as limited availability of experimental data, accuracy and reliability issues with models, and ethical and legal issues related to the use of human data and protection of intellectual property rights [50].

Future Research Directions:

In addition to the areas of research mentioned, future research in CADD may also focus on the integration of emerging technologies such as quantum computing, artificial intelligence, big data analytics, and virtual and augmented reality to further enhance drug discovery and development processes. This may involve the development of new tools and platforms that leverage these technologies, as well as the exploration of novel applications and use cases.

Furthermore, there is a growing interest in the use of machine learning and deep learning algorithms to analyze complex biological data, identify new drug targets, and design more effective drug candidates. This may involve the integration of multiple sources of data, such as genomics, proteomics, metabolomics, and structural biology data, to enable a more comprehensive analysis of biological systems [51].

Future research in CADD is likely to be highly interdisciplinary, involving collaborations between researchers, clinicians, data scientists, and other experts from a range of fields. This will be essential for addressing the complex challenges facing drug discovery and

development, and for translating new discoveries and insights into effective treatments for a wide range of diseases.

Conclusion:

In conclusion, CADD is a rapidly evolving field that is poised to play an increasingly important role in drug discovery and development. Emerging technologies such as quantum computing, artificial intelligence, big data analytics, and virtual and augmented reality offer new opportunities for drug discovery and development, and are likely to have a profound impact on the field in the coming years. As we move further into the 21st century, CADD is likely to continue to grow and evolve, providing new treatments for a wide range of diseases and improving the health and well-being of people around the world.

References-

1. Mullard A. 2017 FDA drug approvals. Nat Rev Drug Discov. 2018 Feb;17(2):81-85. doi: 10.1038/nrd.2018.4. Epub 2018 Jan 19. Erratum in: Nat Rev Drug Discov. 2018 Feb 1;17 (2):150. PMID: 29348678.
2. Morgan S, Grootendorst P, Lexchin J, Cunningham C, Greyson D. The cost of drug development: a systematic review. Health Policy. 2011;100(1):4-17. doi: 10.1016/j.healthpol.2010.12.002.
3. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. J Health Econ. 2003;22(2):151-185. ISSN 0167-6296 doi: 10.1016/s0167-6296(02)00126-1
4. Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. Nature Reviews Drug Discovery, 3(11), 935-949. doi: 10.1038/nrd1549
5. Wang, Y., & Bryant, S. H. (2018). Developing integrative computational approaches to prioritize cancer driver genes and mutations. Frontiers in Physiology, 9, 1670. doi: 10.3389/fphys.2018.01670
6. Schneider, G., & Fechner, U. (2005). Computer-based de novo design of drug-like molecules. Nature Reviews Drug Discovery, 4(8), 649-663. doi: 10.1038/nrd1799
7. Bajorath, J. (2019). New horizons in computer-aided drug discovery. Future Medicinal Chemistry, 11(19), 2467-2470. doi: 10.4155/fmc-2019-0244
8. Clémence Réda, Emilie Kaufmann, Andrée Delahaye-Duriez, (2020) Machine learning applications in drug development,Computational and Structural Biotechnology Journal, 8,241-252.doi: 10.1016/j.csbj.2019.12.006.
9. Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E., & Svetnik, V. (2015). Deep neural nets as a method for quantitative structure–activity relationships. Journal of chemical information and modeling, 55(2), 263-274. doi: 10.1021/ci500747n
10. Ramsundar, B., Kearnes, S., Riley, P., Webster, D., Konerding, D., & Pande, V. (2015). Massively multitask networks for drug discovery. arXiv preprint arXiv:1502.02072.
11. Wu, Z., Ramsundar, B., Feinberg, E. N., Gomes, J., Geniesse, C., Pappu, A. S., & Pande, V. (2018). MoleculeNet: a benchmark for molecular machine learning. Chemical science, 9(2), 513-530. doi: 10.1039/c7sc02664a

12. Jain, A. N. (2007). Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. Journal of computer-aided molecular design, 21(5), 281-306.
13. Chen, B., & Butte, A. J. (2016). Leveraging big data to transform target selection and drug discovery. Clinical Pharmacology & Therapeutics, 99(3), 285-297.
14. Wang, J., Yang, Y., & Yan, R. (2019). A novel data mining framework for drug discovery using protein–ligand–DNA/RNA complex structures. Briefings in bioinformatics, 20(1), 164-176.
15. Duan, J., Dixon, S. L., & Lowrie, J. F. (2010). Advances in computational methods for predicting the toxicity of chemical compounds. Journal of chemical information and modeling, 50(4), 572-584.
16. Medina-Franco, J. L. (2012). Mining SAR knowledge from large-scale compound data sets. Drug discovery today, 17(23-24), 1305-1312.
17. Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. BMC Biology, 9(1), 71.
18. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. Journal of computational chemistry, 32(10), 2319-2327.
19. Shaw, D. E. (2014). Molecular dynamics simulations of proteins and drug discovery. In Annual Reports in Computational Chemistry (Vol. 10, pp. 67-84). Elsevier.
20. Gilson, M. K., & Zhou, H. (2007). Calculation of protein-ligand binding affinities. Annual review of biophysics and biomolecular structure, 36, 21-42.
21. Torng, W., & Altman, R. B. (2018). Hybrid methods for computational drug discovery. Wiley Interdisciplinary Reviews: Computational Molecular Science, 8(1), e1342.
22. Amaro, R. E., & Mulholland, A. J. (2018). Multiscale methods in drug design bridge chemical and biological complexity in the search for effective new drugs. Nature Reviews Chemistry, 2(10), 0148.
23. Zou, J., & Xie, L. (2018). Computational approaches for drug design and discovery: A practical guide for molecular modelers. John Wiley & Sons.
24. Gilson, M. K. (2020). Challenges in the discovery of ligands that target protein-protein interactions. Journal of Medicinal Chemistry, 63(14), 7195-7204.
25. Ganesan, A. (2019). The impact of artificial intelligence in drug discovery. ACS Medicinal Chemistry Letters, 10(6), 857-859.
26. Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. Nature reviews Drug discovery, 3(11), 935-949.
27. Li, J., & Zhu, F. (2020). Machine learning in drug discovery. Nature Methods, 17(1), 114-122.
28. Lin, L., & Li, Y. (2019). Challenges and opportunities in computer-aided drug discovery. Current opinion in pharmacology, 48, 44-50.
29. Roumenina, L. T., Daugan, M. V., Noé, R., Petitprez, F., & Vano YA. (2019). Overcoming barriers to drug discovery and development: the challenge of the industry-academia gap. Drug discovery today, 24(9), 1831-1837.

30. Chaudhary, K., Poirion, O. B., Lu, L., & Garmire, L. X. (2018). Deep learning-based multi-omics integration robustly predicts survival in liver cancer. Clinical Cancer Research, 24(6), 1248-1259.
31. Cho, D. Y., Kim, Y. A., & Przytycka, T. M. (2019). Chapter twelve - network biology approach to drug target identification. In Advances in Protein Chemistry and Structural Biology (Vol. 115, pp. 355-385). Academic Press.
32. Fang, J., & Wang, W. (2020). Multi-omics integration: a review of computational methods and applications in systems biology. Current Genomics, 21(6), 443-452.
33. Rappaport, N., Twik, M., Plaschkes, I., Nudel, R., Iny Stein, T., Levitt, J., ... & Lancet, D. (2017). MalaCards: a comprehensive automatically-mined database of human diseases. Current protocols in bioinformatics, 58(1), 1-18.
34. Singh, A., Gautam, A., & Chaudhary, K. (2021). Multi-omics data integration for identification of key signaling pathways in cancer. Frontiers in Genetics, 12, 585949.
35. Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., ... & Mahajan, S. (2018). Opportunities and obstacles for deep learning in biology and medicine. Journal of The Royal Society Interface, 15(141), 20170387.
36. Hulsen, T., de Vlieg, J., & Alkema, W. (2008). BioVenn–a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC genomics, 9(1), 488.
37. Newman, J., Sayle, R. A., & Crippen, G. M. (2007). A new approach to predicting binding affinities in computer-aided drug design. Protein Engineering, Design and Selection, 20(10), 567-573.
38. Seok, S., & Lee, H. S. (2021). Machine learning in drug discovery. Archives of Pharmacal Research, 1-13.
39. Kumar, A., Zhang, K., & Zhu, W. (2019). Challenges and opportunities in computer-aided drug discovery: A data-driven and machine learning perspective. Current opinion in systems biology, 13, 89-97.
40. Sayle, R. A. (2020). Quantum computing in computational chemistry. Journal of Chemical Information and Modeling, 60(6), 2652-2665.
41. Babbush, R., McClean, J. R., & Wecker, D. (2018). Chemical basis of Trotter-Suzuki errors in quantum chemistry simulation. Physical Review A, 98(2), 022322.
42. Lee, J., Chen, Z., Liu, C., Peng, Y., Cai, Z., & Lu, C. (2018). Quantum machine learning in chemical compound space. Chemical Science, 9(8), 2284-2291.
43. Aliper, A., et al. (2016). Deep Learning Applications for Predicting Pharmacological Properties of Drugs and Drug Repurposing Using Transcriptomic Data. Molecular pharmaceutics, 13(7), 2524-2530.
44. Koh YW, Kim H, Kwak D, et al. Big data analytics for drug discovery: From gene to drug target. Prog Biophys Mol Biol. 2020;154:23-33. doi:10.1016/j.pbiomolbio.2019.12.005
45. Gilson, M. K., Liu, T., Baitaluk, M., Nicola, G., Hwang, L., & Chong, J. (2016). BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic acids research, 44(D1), D1045-D1053.
46. Chong, J., & Gilson, M. K. (2015). Advances in the use of molecular docking to address key problems in computational drug discovery. Current opinion in structural biology, 35, 1-9.

47. Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., ... & Collins, J. J. (2020). A deep learning approach to antibiotic discovery. Cell, 180(4), 688-702.
48. Pereira, J. C., Caffarena, E. R., & dos Santos, C. N. (2016). Boosting docking-based virtual screening with deep learning. Journal of chemical information and modeling, 56(12), 2495-2506.
49. Zhang, L., Tan, J., Han, D., Zhu, H., Lin, H., & Yang, F. (2021). Virtual and augmented reality in drug discovery and development. Drug discovery today, 26(2), 595-602.
50. Hajduk, P., Hruby, V., Greer, J., Kaminski, R., & Johnson, T. (2017). Discovery of Potent and Selective Inhibitors for GPR139. ACS medicinal chemistry letters, 8(7), 807-812. doi: 10.1021/acsmedchemlett.7b00147
51. Cherkasov A, Muratov EN, Fourches D, et al. QSAR modeling: where have you been? Where are you going to? J Med Chem. 2014;57(12):4977-5010. doi: 10.1021/jm4004285.