DARSGAH-E-AHLEBAIT

Corresponding Author:
dr.syedrizwan@kiu.edu.pk

Journal of Biotechnological Sciences

CHEMOINFORMATICS: UNRAVELLING NATURE'S CODE FOR DRUG DEVELOPMENT

Asadullah[1] and Syed Rizwan Abbas[2]
[1]Department of Animal Sciences; Zoology; Karakoram International University; Gilgit.
[2]Department of Plant Sciences; Karakoram International University; Gilgit.

ABSTRACT

Chemoinformatics, a multidisciplinary field involving physics, chemistry, biology, mathematics, and informatics, has emerged as a significant field of computational molecular sciences. From the inception of chemical structure representations and Structure Searching in the 1960s, to advances like molecular modeling, Computer Aided Drug Design, Computer-Assisted Structure Elucidation (CASE), and Quantitative Structure-Activity relationships (QSAR/QSPRs). This review explores Chemoinformatics implementation in hit identification, lead compound identification, and lead optimization at various phases of drug development. The revolutionary significance of artificial intelligence and machine learning in transforming several procedures, including virtual screening methods, de novo drug discovery, and candidate drug ranking, is highlighted in particular. However, there are some obstacles to overcome in integrating AI into drug research, from problems with data availability and quality to ethical and legal considerations. This study thoroughly summarizes the developments and difficulties in Chemoinformatics and artificial intelligence integration in drug discovery, providing insightful information for academics and professionals.

Keywords: Chemoinformatics; Drug development; De novo drug design; Molecular simulation; Lead optimization.

INTRODUCTION

A few years ago, the term "Chemoinformatics" emerged and quickly became popular. One can find numerous job postings in publications and attend workshops and symposia dedicated solely to Chemoinformatics **(Gasteiger, 2006)**. Frank Brown is credited with coining the term

"Chemoinformatics" when he stated that information technology and management are now essential components of the drug discovery process. To turn data into information and information into knowledge and make better judgments more quickly in the field of drug lead organization and identification, chemo-informatics combines several information resources **(Brown, 1998) .** Greg Paris developed a definition that was far more inclusive: "The structure, development, organization as well, management, retrieving, assessment, disseminating representation, and application of chemical data are all included in the general term "Chemoinformatics." **(Gasteiger, 2006)**. Clearly, not just in drug design but in every area of chemistry, the conversion of findings into information and that data into knowledge is a necessary task. Gasteiger *et al*., therefore stick to a definition that is significantly broader and share the belief that Chemoinformatics technologies are necessary in all fields of chemistry: Chemoinformatics is the use of informatics techniques to tackle problems related to chemistry **(Gasteiger, 2006)**. Natural substances originating from microorganisms, plants, and marine life have proven to be abundant in biologically active substances with a wide range of pharmacological uses. The application of bioinformatics techniques to natural product research has significantly improved the investigation of this enormous pool of chemical variety **(Singh, 2023).**

Chemo-informatics, also known as chemical informatics, chemical information, or chemo-informatics, is a word that has gained recognition as a separate field in computational molecular sciences in recent times. Because it integrates physics, chemistry, biology, mathematics, biochemistry, statistics, and informatics, cheminformatics is often called interface science **(Begam B. F., 2012).** Drug development, which includes verifying, testing, and launching a new drug, and drug discovery are the two most important tasks in pharmaceutical research. The process of discovering novel pharmaceutical medicines is known as drug discovery **(Romano, 2019)**. Natural products (NPs) have long been used in herbal cures and conventional medicine. Even in the current era of small-molecule drug development, they continue to be the most abundant source of motivation **(Newman, 2020) .** NPs are connected to small molecules to varying degrees in nearly two-thirds of all small-molecule medications licensed between 1981 and 2019 **(Chen, 2020)**. During this period, just 5% of new medicines have been approved for use; the remaining 28% are NP derivatives, and the remaining 35% mimic and/or contain an NP pharmacophore **(Newman, 2020).**

A relatively new area of information technology called "cheminformatics" is devoted to gathering, storing, analyzing, and manipulating chemical data. Information about tiny molecule formulas, structures, characteristics, spectra, and activities (industrial or biological) is usually included in the scientific data of significance. Originally developed to aid in the drug discovery and development process, cheminformatics is becoming more and more significant in many branches of biology, chemistry, and biochemistry **(Wishart, 2007).**

HISTORY OF CHEMOINFORMATICS

Chemo-informatics is the use of informatics techniques to tackle problems related to chemistry. The origins of this field date back more than 40 years, although this word has only been recently used. The 1960s saw the beginning of work on computer-assisted structure elucidation and synthesis design, quantitative structure-activity connections, chemo-metrics, molecular modeling, and chemical structure representation and searching **(Gasteiger, 2006).**

CHEMICAL STRUCTURE REPRESENTATIONS

Several machine-understandable chemical structure representations were investigated in the early 1960s as a potential foundation for libraries of chemical compounds and activities **(Tate, 1967) .** Because they were brief at first, linear notations were used; the most popular of them was the Wiswesser line notation (WLN). When it came to representing structures, mathematicians investigated matrices; however, one drawback of them is that the quantity of elements grows as the square of the atom counts. In due course, connection tables—which depict molecules as lists of their atoms and bonds—became widely used. The Chemical Abstracts Registry System, which debuted in the latter part of the 1960s, also made use of connection tables **(Dyson, 1968) .** The SMILES notation, on the other hand, is one line **(Weininger, 1988)** and it continues to be widely used, especially for online chemical structural data exchange **(Gasteiger, 2006)**.

STRUCTURE SEARCHING

The Morgan algorithm was the first to count the particles in a molecule distinctly and clearly **(Morgan, 1965).** That is why a complete search structure was supplied for the foundation. 1957 saw the introduction of substructure searching's next phase **(Ray, 1957)** in the course of the late 1960s and the beginning of the 1970s, this greatly sparked research efforts **(Sussenguth, 1965).**

QUANTITATIVE STRUCTURE-ACTIVITY/PROPERTY RELATIONSHIPS

In 1964, the Free-Wilson analysis was additionally created to establish a connection between biological function and the existence or lack of specific substructures in molecules **(Free, 1964) .** Over the past 40 years, a vast quantity of work has been published connecting descriptors obtained from molecular structures with various physical, chemical, or biological data. During these studies, statistical models as well as descriptive categorization were developed. Quantitative structure-property relationships (QSPRs) and quantitative structure-activity relationships (QSARs) have been recognized as distinct areas by these investigations, complete with conferences, journals, and societies **(Society, n.d.)**.

COMPUTER-ASSISTED STRUCTURE ELUCIDATION (CASE)

Chemical structure elucidation was identified early on as a domain to explore with artificial intelligence methods. The DENDRAL project was started at Stanford University in 1964 **(Lindsay, 1980)** and sparked a lot of curiosity **(Gray, 1986)** at Toyohashi University of

Technology, Sasaki launched more methods for computer-assisted structural elucidation in the late 1960s **(Sasaki, 1968)** and at the University of Arizona by Munk **(Shelley, 1978).**

MOLECULAR MODELING

Langridge and others created techniques in the late 1960s for cathode ray tube screens to display three-dimensional molecular models **(Lesk, 1977) .** Marshall began displaying protein structures on graphic screens at the same time. Advancements in hardware and software technology, namely in the areas of graphics screens and graphics cards, have resulted in the development of extremely complex systems that enable the detailed imaging of intricate molecular structures. The modeling of complex organic structures, in particular proteins, has also advanced significantly thanks to techniques like molecular dynamics simulations, docking into protein binding sites, and homology modeling **(Gasteiger, 2006).**

APPLICATIONS OF CHEMOINFORMATICS IN DRUG DEVELOPMENT

New drug research is a time-consuming, costly, and hazardous process that frequently results in project failure. Each inconclusive study is a huge loss for the pharmaceutical industry since each drug takes about 2.5 billion dollars and 15–17 years to be fully launched, and only 13% of drugs acquire the necessary authorizations **(Wouters, 2020).** The initial step in the drug discovery process is target identification, which can be carried out using a range of technologies, such as proteomics and genomics. In this sense, the techniques of bioinformatics and Chemoinformatics complement each other. Optimization of leads and identification can be accomplished through a variety of methods, such as data mining, QSAR, and Insilco-ADME. The result is an active pharmaceutical compound with minimal to no adverse effects that provide a therapeutic response **(Raslan et al., 2023)**.

HIT IDENTIFICATION

Because fragments have a low affinity by nature, hit validation is a crucial step in FBDD. Before reaching a lead compound, a hit fragment structure must undergo multiple variations of progress, merging, and/or linking. Without structure information, the entire process can prove to be quite challenging in comparison to the molecular development of small compound hits found by HTS studies **(Kasper P. Lundquist, 2021)** . The earliest and most crucial phase of small-molecule drug development is the recognition of hits **(Yan, 2020)** Using virtual chemical libraries in various virtual screening techniques has emerged as a viable strategy for finding new hit compounds. In this context, several researchers are employing diverse in silico techniques to create novel de novo chemical and on-demand libraries **(Walters, 2018).**

LEAD COMPOUND IDENTIFICATION AND TARGET VALIDATION

To find possible lead substances for a particular biological target, bioinformatics techniques including machine learning algorithms and quantitative structure-activity relationships (QSAR) are employed **(Begam B. F., 2012).** By examining the connection between chemical structure and biological activity, these techniques make it possible to forecast the activity of novel substances **(Lavecchia, 2015).**

LEAD OPTIMIZATION

A time-consuming and costly step in the drug discovery process is lead optimization. The scientists working on it too have a poor understanding of this component. Little progress was made for several decades, despite early Chemoinformatics and QSAR scientists making good contributions in putting the foundations of lead optimization on a mathematical basis. However, more effective approaches and tools to support effective lead optimization have been developed as a result of the availability of improved datasets as well as the use of statistical optimization techniques employed by other industries **(Green, 2013)**.

In the critical lead optimization stage of the drug development process, preliminary hits are improved to produce optimized compounds with improved pharmacokinetic, potent, and selectivity characteristics. Conventional techniques rely on empirical strategies, including hand chemical changes and iterative experimentation. AI and ML approaches have completely transformed lead optimization, which uses big data and predictive modeling to identify and improve lead candidates quickly. AI-driven methods navigate chemical space, anticipate ideal binding positions, and adjust force fields in molecular mechanics using sophisticated algorithms and predictive models **(Niazi et al., 2024).** Toxicology prediction is a key component of lead optimization. Whenever a lead chemical is put through clinical studies, bioinformatics methods can be utilized to forecast the possible harm related to it. In the early stages of medical development, this aids scientists in identifying any safety concerns **(Mustafa, 2023).**

DOCKING

Docking is mostly accomplished using two kinds of methods. Using computer simulations, one method estimates the energy profile of ligand targets coupled conformers. In contrast, the second strategy makes use of a method that determines the complementarity of surfaces between the ligand and the target **(Agarwal, 2016) .** To predict a tiny molecule's affinity and activity, docking is frequently utilized to predict how therapeutic compounds including small molecules would align concerning their protein targets. Docking is essential to sensible medication design **(Raval, 2022).**

A computer method for predicting the binding manner and affinity of small molecules (ligands) to target proteins is called protein-ligand docking, and it is employed in drug discovery. It entails figuring out a ligand's ideal orientation and shape within a protein binding site. Target protein preparation, ligand preparation, docking type selection, scoring function, and validation make up the process. Flexible docking takes into account conformational changes that occur during binding, whereas rigid body docking treats ligands and targets as rigid bodies. Commonly used docking tools include AutoDock, Vina, and LeDock, each with unique algorithms and force fields. Many criteria, including ligand flexibility, computational capacity, and the requirement for speed in virtual screening, influence the choice of a docking program **(Muhammed, 2024).**

CREATION OF VIRTUAL LIBRARIES

Chemical informatics instruments produce virtual libraries containing chemicals, which are subsequently examined for their biological activity. With this method, scientists can comb

over a large chemical universe and find substances that show promise for future research **(Begam B. F., 2012).**

MOLECULAR SIMULATIONS

All theoretical and computational techniques used to investigate the activity of a molecule or an assembly of compounds are collectively referred to as "molecular simulation" or "molecular modeling" **(Katiyar, 2018).** AI is being used to reduce the need for physical testing of potential therapeutic compounds by enabling high-fidelity molecular simulations that can be performed entirely on computers (i.e., in silico) without suffering the prohibitive costs of traditional chemical techniques **(Yang H. S., 2018).** Molecular dynamics (MD) represents one such method of simulation. It seeks to derive claims on a molecular system's structural, dynamical, and thermodynamical characteristics. The latter is usually a biomolecule (solute) that is submerged in an aqueous solvent (water or electrolyte), such as an enzyme, protein, or group of lipids that form a membrane. For proteins and enzymes, MD simulations begin with the experimental protein structure that has been deposited in the Protein Data Bank (PDB) **(Salo-Ahen, 2020).**

PREDICTION OF DRUG PROPERTY

Certain AI systems are used to predict key characteristics like as toxicity, bioactivity, and physicochemical qualities of drugs, hence reducing the need to simulate the testing of medication candidates **(Ye, 2021)**. Artificial intelligence is a significant tool in medical chemistry used for the prediction of toxins as well as the efficiency/ efficacy of a potential drug candidate. In addition, the use of machine learning (ML), one of artificial intelligence's most well-known and effective techniques, has emerged to help with drug development issues **(Alhatem, 2024) .** Machine Learning (ML) algorithms help in the identification of patterns and trends that are not easily detectable to researchers while the analysis of large data sets. This activity boosts up the process of recognition of new biologically active drugs as well as small synthetic compounds with the least side effects, unlike traditional methods **(Thenuwara, 2023).**

DE NOVO DRUG DESIGN

De novo drug design is a technique that involves assembling components or atoms to create unique ligands in the designated target site **(Bai, 2022)** . AI is also transforming the traditional method of medication discovery, which entails screening a vast number of possible molecules. Certain systems possess the capability to entirely generate innovative and auspicious therapeutic substances **(Aja, 2021).** Proposing new compounds with a desired biological impact, i.e., affinity to a target macromolecule, is the main goal of de novo design. All de novo design algorithms consider this goal, even though scoring functions vary greatly in how they evaluate biological activity. That is why, it can be known as the primary drawback of de novo design **(Hartenfeller, 2011) .** Through training both the chemical and physical characteristics of medications, deep learning approaches offer yet another novel avenue for de novo drug design. The crucial stage in de novo drug design using deep learning techniques is extracting the required chemical and physical characteristics **(Yang S. Q.-Z., 2021).**

CANDIDATE DRUG PRIORITIZATION

Artificial intelligence (AI) is utilized to evaluate and prioritize these molecules for further study when multiple promising lead therapeutic compounds have been identified, outperforming conventional ranking techniques **(Zhou, 2020).**

SYNTHESIS PATHWAY GENERATION

Artificial Intelligence is being utilized to build synthesis pathways for producing imaginary medical compounds and designing speculative medications. These routes even propose compound modifications to increase their ease of production in certain instances. As AI systems advance, fully automated end-to-end drug discovery is more likely to become a reality **(Wen, 2017).**

OBSTACLES TO THE USE OF MACHINE LEARNING IN DRUG DEVELOPMENT

PROBLEMS WITH DATA AVAILABILITY AND QUALITY

ML models are very dependent on plentiful and high-quality data. Nevertheless, such data is frequently inaccessible and its reliability might be impacted by several issues, such as data noise, missing values, and discrepancies **(Udegbe, 2024).**

MODEL INTERPRETABILITY AND INTEGRATION

Understanding the underlying mechanics and decision-making processes of machine learning (ML) models can be tough due to their interpretability. Furthermore, incorporating ML techniques into current workflows can be difficult and necessitate considerable adjustments to conventional drug discovery procedures **(Udegbe, 2024)**.

REGULATORY AND ETHICAL ISSUES

The application of algorithmic learning in drug development raises ethical and legal questions. Securing regulatory compliance and addressing ethical concerns around bias in ML algorithms and data security are critical to the successful application of machine learning (ML) in drug development **(Udegbe, 2024).**

INTERDISCIPLINARY COOPERATION AND DATA-SHARING

The application of machine learning (ML) in drug development necessitates cooperation between scientists with various specialties, such as biology, chemistry, and computer science. Leveraging ML's strengths and tackling its limitations require interdisciplinary collaboration and effective data-sharing methods **(Udegbe, 2024)**.

LIMITATIONS OF ARTIFICIAL INTELLIGENCE IN DRUG DISCOVERY

Drug research could benefit from artificial intelligence, but some various obstacles and restrictions need to be carefully considered. The availability of appropriate data is one of the

main obstacles. Large datasets are usually necessary for the efficient training of AI-driven techniques **(Blanco-Gonzalez, 2023)** . Unfortunately, the availability of data is frequently restricted, inconsistent, or of inferior quality, which jeopardizes the accuracy and dependability of the findings. The issue of ethics also poses a difficulty **(Prem, 2023).**

Modern AI-based approaches shouldn't be used in place of traditional experimental techniques, nor can they take the place of the invaluable knowledge and experience that human researchers provide **(Dwivedi, 2023) .** AI can only make predictions based on data that is currently accessible; human researchers are still needed for result validation and interpretation. However, Integrating AI with traditional experimental approaches may enhance the medication development procedure. Computational intelligence has the potential to enhance the drug discovery process and accelerate the creation of novel treatments by combining its predictive powers with the knowledge and experience of human researchers in a synergistic manner **(Khan *et al*., 2024).**

OVERCOMING CHALLENGES

Using intelligent technology (AI) in the pharmaceutical industry is a prudent strategy to overcome current roadblocks. This will help ensure that AI in drug research continues to advance. The optimization of data inputs is prioritized, focusing on diverse and high-quality datasets as the cornerstone of resilient artificial intelligence models. This addresses problems with data representativeness and correctness. Establishing ethical standards and governance frameworks is critical, and implementing sustainable and ethical AI use acts as an inspiration in this process. Authorization and data security are two examples of these variables. Fusing domain-specific knowledge with computational skills promotes a synergistic partnership engaging with regulatory organizations based on integrity and validity standards can help based on artificial intelligence drug discovery techniques become more widely accepted and regulated. AI studies aim to promote open collaboration and data sharing that supports a mutually beneficial culture in the drug development industry **(Khan et al., 2024)**. To bridge the knowledge gap and ensure a workforce capable of navigating the intersection of medical sciences and artificial intelligence, dedication to training and development activities is recommended. To sum up, these tactics build a thorough framework for resolving current issues and maximizing AI's contribution to the advancement of drug discovery techniques **(Sagar, 2024).**

CONCLUSION

The combination of artificial intelligence (AI) and Chemoinformatics has completely changed the drug development industry by providing strong instruments for organizing, interpreting, and maximizing chemical data. This multidisciplinary strategy has greatly sped up the drug development process by combining disciplines including biology, chemistry, pharmacology, and computer science. Chemoinformatics allows researchers to forecast the characteristics and behavior of drug candidates, optimize lead molecules, and quickly identify possible therapeutic targets. AI-driven methods, such as molecular simulations and machine learning, allow for the quick exploration of large chemical regions, which saves money and time in finding new therapeutic chemicals. But even with these encouraging developments, there are

still many obstacles standing in the way of using AI in medication development. To completely utilize AI in drug discovery, concerns including data quality and availability, model interpretability, regulatory compliance, and interdisciplinary collaboration must be resolved. However, these obstacles can be addressed by emphasizing the ethical and responsible use of AI, encouraging interdisciplinary cooperation, and funding skill development and education. The application of AI in drug research has the potential to completely transform the industry and ultimately result in the creation of more potent treatments for a variety of illnesses.

REFERENCES

Agarwal, S. &. (2016). An overview of molecular docking. *JSM chem, 4*(2), 1024-1028.

Aja, P. M. (2021). Prospect into therapeutic potentials of Moringa oleifera phytocompounds against cancer upsurge: de novo synthesis of test compounds, molecular docking, and ADMET studies. *Bulletin of the National Research Centre, 45*(1), 99.

Alhatem, A. W. (2024). Revolutionizing diagnostic pathology: the emergence and impact of artificial intelligence-what doesn't kill you makes you stronger? *Clin. Dermatology*. doi:10.1016/j.clindermatol.2023.12.020

Bai, Q. L.-L.-S. (2022). Application advances of deep learning methods for de novo drug design and molecular dynamics simulation. *Wiley Interdisciplinary Reviews: Molecular Science, 12*(3), 1581.

Begam, B. F. (2012). A study on cheminformatics and its applications on modern drug discovery. *Procedia engineering, 38*, 1264-1275.

Blanco-Gonzalez, A. C.-G.-T.-R.-F. (2023). The role of AI in drug discovery: challenges, opportunities, and strategies. *Pharmaceuticals, 16*(6), 891.

Brown, F. K. (1998). Chapter 35. Chemoinformatics: What is it and how does it impact drug discovery. *Annual Reports in Medicinal Chemistry, 33*, 375-384.

Chen, Y. &. (2020). Cheminformatics in natural product-based drug discovery. *Molecular informatics, 39*(12), 2000171.

Dwivedi, Y. K. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, 102642.

Dyson, G. M. (1968). A modified IUPAC-Dyson notation system for chemical structures. *Information Storage and Retrieval, 4*(1), 27-83.

Free, S. M. (1964). A mathematical contribution to structure-activity studies. *Journal of medicinal chemistry, 7*(4), 395-399.

Gasteiger, J. (2006). Chemoinformatics: a new field with a long tradition. *Analytical and bioanalytical chemistry, 384*, 57-64.

Gray, N. A. (1986). Computer-assisted structure elucidation. *John Wiley & Sons, Inc.*.

Green, D. V. (2013). Chemoinformatics in lead optimization. *Chemoinformatics for drug discovery*, 149-178.

Hartenfeller, M. &. (2011). Enabling future drug discovery by de novo design. *Wiley Interdisciplinary Reviews: Computational Molecular Science, 1*(5), 742-759.

Khan, M. K., Raza, M., Shahbaz, M., Hussain, I., Khan, M. F., Xie, Z., Shoaib, S., Shah, A., Tareen, A. K., & Bashir, Z. (2024). *The recent advances in the approach of arti fi cial intelligence (AI) towards drug discovery*. *May*, 1–10. https://doi.org/10.3389/fchem.2024.1408740

Kasper P. Lundquist, D. V. (2021). Fragment-Based Drug Discovery for RNA Targets. *chemistry Europe, 16*(17), 2588-2603.

Katiyar, R. S. (2018). Molecular simulations in drug delivery: Opportunities and challenges. *Wiley Interdisciplinary Reviews: Computational Molecular Science, 8*(4), 1358.

Lavecchia, A. (2015). Machine learning methods in QSAR modeling: a historical look, current state-of-the-art and future prospects. *Expert Opinion on Drug Discovery, 10*(5), 537-561.

Lesk, A. M. (1977). Macromolecular marionettes. *Computers in Biology and Medicine, 7*(2), 113-129.

Lindsay, R. K. (1980). Applications of artificial intelligence for organic chemistry: the DENDRAL project.

Morgan, H. L. (1965). The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service. *Journal of chemical documentation, 5*(2), 107-113.

Muhammed, M. T.-Y. (2024). Molecular docking: principles, advances, and its applications in drug discovery. *Letters in Drug Design & Discovery, 21*(3), 480-495.

Mustafa, M. M. (2023). Integrating Bioinformatics in Drug Discovery.

Newman, D. J. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of natural products, 83*(3), 770-803.

Niazi, S. K., Mariam, Z., & Magoola, M. (2024). *Optimizing Lead Compounds : The Role of Artificial Intelligence in Drug Discovery Optimizing Lead Compounds : The Role of Artificial Intelligence in Drug Discovery*. https://doi.org/10.20944/preprints202404.0055.v1

Prem, E. (2023). From ethical AI frameworks to tools: a review of approaches. *AI and Ethics, 3*(3), 699-716.

Raslan, M. A., Raslan, S. A., Shehata, E. M., Mahmoud, A. S., & Sabri, N. A. (2023). *Advances in the Applications of Bioinformatics and Chemoinformatics*.

Raval, K. &. (2022). Basics, types and applications of molecular docking: A review. *IP International Journal of Comprehensive and Advanced Pharmacology, 7*(1), 12-16.

Ray, L. C. (1957). Finding chemical records by digital computers. *Science, 126*(3278), 814-819.

Romano, J. D. (2019). Informatics and computational methods in natural product drug discovery: a review and perspectives. *Frontiers in genetics, 10*, 442-506.

Sagar, R. U. (2024). A comparative study on morphology dependent performance of neodymium–graphene as an anode material in lithium-ion batteries. *Journal of Energy Storage, 77*, 109854.

Salo-Ahen, O. M. (2020). Molecular dynamics simulations in drug discovery and pharmaceutical development. *Processes, 9*(1), 71.

Sasaki, S. A. (1968). Automated structure elucidation of several kinds of aliphatic and alicyclic compounds. *Analytical Chemistry, 40*(14), 2220–2223.

Shelley, C. A. (1978). An approach to automated partial structure expansion. *Analytica Chimica Acta, 103*(2), 121-132.

Singh, L. (2023). Bioinformatics Approaches in Natural Product Research Unraveling Nature's Chemical Diversity. *International Journal of Advanced Research in Medicinal Chemistry, 5*(2), 1-5.

Society, Q. a. (n.d.). *Molecular Graphics and Modeling Society*. Retrieved from http://www.qsar.org: http://mgms.org

Sussenguth, E. H. (1965). A graph-theoretic algorithm for matching chemical structures. *Journal of Chemical Documentation, 5*(1), 36-43.

Tate, F. A. (1967). Handling chemical compounds in information systems. *Annual Review of Information Science and Technology, 2*(4), 285-309.

Thenuwara, G. C. (2023). Advances in diagnostic tools and therapeutic approaches for gliomass: a comprehensive review. *Sensors, 23*(24), 9842. doi:10.3390/s23249842

Udegbe, F. C. (2024). Machine Learning in Drug Discovery: A critical review of applications and challenges. *Computer Science & IT Research Journal, 5*(4), 892-902.

Walters, W. P. (2018). Virtual chemical libraries: miniperspective. *Journal of medicinal chemistry, 62*(3), 1116-1124.

Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences, 28*(1), 31-36.

Wen, M. Z. (2017). Deep-learning-based drug–target interaction prediction. *Journal of proteome research, 16*(4), 1401-1409.

Wishart, D. S. (2007). Introduction to cheminformatics. . *Current protocols in bioinformatics, 18*(1), 14-1. doi:https://doi.org/10.1002/0471250953.bi1401s18

Wouters, O. J. (2020). Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *Jama, 323*(9), 844-853.

Yan, X. C. (2020). Augmenting hit identification by virtual screening techniques in small molecule drug discovery. *Journal of chemical information and modeling, 60*(9), 4144-4152.

Yang, H. S. (2018). In silico prediction of chemical toxicity for drug design using machine learning methods and structural alerts. *Frontiers in chemistry, 6*, 30.

Yang, S. Q.-Z. (2021). Current advances in ligand-based target prediction. *Wiley Interdisciplinary Reviews: Computational Molecular Science, 11*(3), 1504.

Ye, Z. &. (2021). Prediction of small-molecule compound solubility in organic solvents by machine learning algorithms. *Journal of cheminformatics, 13*(1), 98.

Zhou, Y. W. (2020). Artificial intelligence in COVID-19 drug repurposing. *The Lancet Digital Health, 2*(12), e667-e676.