GMR
The Original

MACHINE LEARNING-ASSISTED COMPUTATIONAL CHEMISTRY FOR ACCELERATED MOLECULAR DESIGN AND SUSTAINABLE CHEMICAL INNOVATION

Dr. Anil Kumar Singh[1], Dr Kedar Prasad Meena[2], Balaji Maroti Pulchwad[3], Texin Joseph[4], Dr. Chayannika Singh[5], Dr. Dinesha H E[6]

[1]Associate Professor, Physical Chemistry, Department of Chemistry, Teerthanker Mahaveer University, Moradabad, UP - 244001, ORCID: https://orcid.org/my-orcid?orcid=0000-0003-2933-4058 Email Id: dr.anilkumarsingh2009@gmail.com
[2]Associate Professor, Pharmaceutics, Guru Ghasidas Vishwavidyalaya, Bilaspur (C.G), Email ID: meenapharmaceutics@rediffmail.com
[3]Lecturer, Department of Petrochemical Engineering, Dr. Babasaheb Ambedkar Technological University, Lonere, Raigad, M.S, Email ID: balajipulchuwad1@gmail.com
[4]Assistant Professor, Department of Chemistry, St. Paul's College, Kalamassery, Kerala, 683503, ORCID: 0009-0008-2444-2897, Email ID: texin@stpauls.ac.in
[5]Associate Professor, Department of Chemistry, Deen Dayal Upadhyaya College, University of Delhi, Delhi, 110078, ORCID: 0009-0009-3313-9491
[6]Assistant Professor, Department of Chemistry, Maharani's Science College for Women (Autonomous), Mysore, 570005, Email ID: dineshahe@gmail.com

ABSTRACT:
Machine learning-assisted computational chemistry has emerged as a transformative approach for accelerating molecular design and promoting sustainable chemical innovation by integrating data-driven algorithms with computational modeling techniques. This review provides an overview of the principles, methodologies, and applications of machine learning in computational chemistry, highlighting its role in molecular property prediction, quantitative structure–activity and structure–property relationship modeling, virtual screening, de novo molecular design, and structure-based drug discovery. The review further discusses key molecular representation strategies, including molecular descriptors, fingerprints, SMILES, SELFIES, graph-based representations, and biomolecular features, which serve as the foundation for predictive modeling. Major machine learning algorithms, including regression models, support vector machines, random forests, artificial neural networks, graph neural networks, transformer models, and generative artificial intelligence, are examined with respect to their contributions to computational chemistry workflows. In addition, the review explores the expanding applications of machine learning in molecular biology and genetics, including drug target identification, protein structure prediction, pharmacogenomics, biomolecular interaction analysis, disease-associated molecular discovery, and precision medicine. The role of artificial intelligence in advancing green molecular design, sustainable catalyst discovery, reaction optimization, eco-friendly material development, and sustainable pharmaceutical manufacturing is also highlighted. Finally, emerging trends such as generative AI, large language models, explainable artificial intelligence, multi-omics integration, and autonomous molecular discovery platforms are discussed alongside current challenges and future perspectives. Collectively, machine learning-assisted computational chemistry represents a powerful interdisciplinary framework for accelerating molecular discovery while supporting sustainable and biologically informed chemical innovation.

KEYWORDS: Machine learning, computational chemistry, molecular design, drug discovery, sustainable chemical innovation

1. INTRODUCTION

Computational chemistry is considered an important branch of science for studying structural, electronic, and physicochemical properties of molecules through theoretical models and computer simulations. Approaches like quantum chemistry, molecular mechanics, density functional theory (DFT), and molecular dynamics have contributed to better molecule studies, lowering down the cost and increasing speed of chemical studies (Ma, 2022). Though there has been advancement in conventional approaches of computation chemistry, such computational techniques have taken a lot of computational resources and time to study large chemical space and optimize complicated molecular systems (Keith et al., 2021). The quick development in machine learning (ML) field has made computational chemistry much easier through data driven prediction of molecular properties and reactions. Contrary to the algorithms used in conventional computational approaches, machine learning models help find out relationships between data in chemical databases and help improve accuracy while cutting down computational efforts. Recent development in deep learning, graph neural network, and computer aided synthesis planning has allowed predicting of chemical reactions and optimizing the structure of molecules, which has increased the efficiency of chemical research and development work (Cova & Pais, 2019; Coley et al., 2018).

AI has also transformed molecular design by speeding up the discovery of new molecules with desirable physicochemical and biological characteristics. Machine-learning enabled computational chemistry facilitates virtual screening, prediction of molecular properties, de novo molecule generation, and molecule optimization, greatly reducing the time required in the conventional molecular design process. Advanced molecule representation approaches used together with deep learning models have facilitated the development of novel drugs, polymers, catalysts, and materials with enhanced performance while minimizing the need for experimentation (Chen et al., 2020; Li et al., 2022). Beyond molecular design, computational chemistry supported by AI has been taking on increasing significance in molecular biology and drug development. By incorporating machine learning in molecular docking and molecular dynamics simulation, it becomes easier to predict protein-ligand interactions, biomolecular recognition, and other aspects of structure-function relationships important for developing therapies (Zhao et al., 2023). Computational methods help conduct rational drug design, target identification, and analysis of biomolecules, which enhances the association between computational chemistry and contemporary molecular biology (Singh et al., 2022; Hong et al., 2024). Besides that, computational methods have become increasingly helpful for developing molecular biotechnology and precision medicine by facilitating the comprehension of biological processes (Glick & Patten, 2022).

AI-assisted computational chemistry can play a part in fostering sustainable chemical innovation by facilitating the design of eco-friendly materials, catalysts, and energy-saving chemical reactions. AI models for computation will minimize experimental waste, optimize resource use, and speed up the design of sustainable and renewable materials, sustainable polymers, hydrogen storage systems, and eco-friendly production processes (Leonard et al., 2021; Osman et al., 2023). The examples above are indicative of the great promise that lies in combining computational intelligence with chemistry for the promotion of scientific and sustainable progress. This review is devoted to the exploration of machine learning-assisted computational chemistry, focusing on its application in accelerated molecular design and sustainable chemical innovation. In particular, this paper highlights the theoretical background of computational and machine learning algorithms, molecular modeling techniques, applications in molecular design and biomolecular research, sustainability-driven innovations, emerging technologies, existing challenges, and future research trends.

2. Fundamentals of Machine Learning-Assisted Computational Chemistry

Molecular design through computational chemistry assisted by machine learning involves a systematic approach in which data and computational and artificial intelligence approaches are brought together for chemical and biomolecular information. Through the combination of data from molecular databases, feature engineering, machine learning, and computational chemistry, this framework is capable of predicting molecular properties, conducting virtual screening, optimizing molecules, and discovering drugs based on structure (Keith et al., 2021; Coley et al., 2018). Figure 1 shows the workflow of this computational framework as illustrated below.

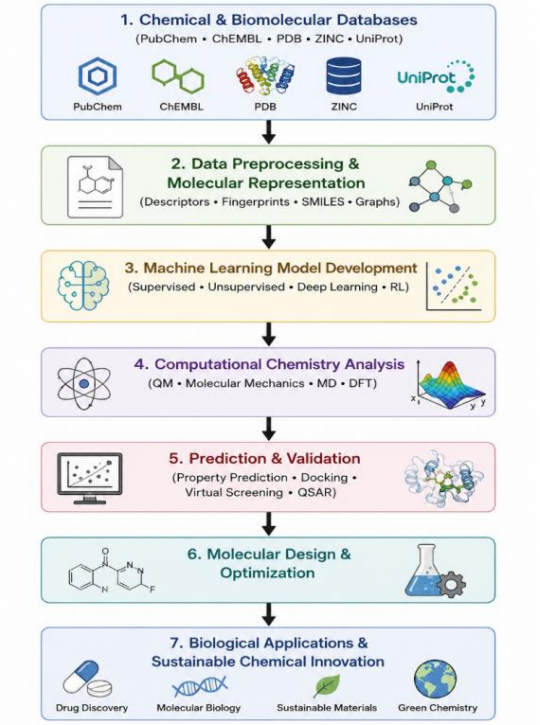

Figure 1. Workflow of Machine Learning-Assisted Computational Chemistry for Accelerated Molecular Design and Biological Applications

From Figure 1 below, it is evident that the combination of machine learning with computational chemistry enables a seamless and data-intensive process whereby molecular data are converted into insights to be applied in chemical and biological contexts.

2.1 Computational Chemistry: Concepts and Methods

Computational chemistry uses mathematical modeling and computer simulation techniques to study the structure, electron configuration, and other physical and chemical properties of molecules. While it supports experimentation in understanding the behavior, mechanism, and other properties of molecules through prediction, it reduces time and expenses. Due to its effectiveness in simulating chemical systems, computational chemistry has become vital in drug discovery, biomolecular sciences, materials, and sustainable chemistry (Ma, 2022; Keith et al., 2021).

The theory behind quantum chemistry helps describe molecular systems through quantum mechanical modeling of their electronic structures and energy levels. Though very accurate, these approaches are computationally expensive when used on complex molecules. Molecular mechanics is a computationally inexpensive method for modeling molecular structures using classical force fields, which makes it suitable for use in modeling of proteins, polymers, and other macromolecules. The molecular dynamics approach adds another dimension to molecular mechanics by modeling atom movements in time. It helps study protein folding, ligand binding, and conformations (Naqvi et al., 2018; Singh et al., 2022). Another approach that combines computational efficiency and accuracy is density functional theory (DFT). This approach allows the study of electron density and it is commonly used in catalytic processes, chemical reactions and material discovery (Keith et al., 2021; Wang et al., 2024).

2.2 Fundamentals of Machine Learning

Machine learning (ML) is an application that allows computational systems to learn from the data of chemicals and predict their properties without any explicit programming. Machine learning has many applications in computational chemistry, where it learns the molecular descriptors, fingerprints, and structures for the prediction of properties, optimization of reactions, and discovering molecules (Cova & Pais, 2019). Supervised machine learning algorithms are used extensively to predict the properties of molecular structures and biological activity using labelled data sets, while unsupervised learning methods are used to recognize the hidden patterns and clusters within the chemical data (Li et al., 2022; Stepišnik et al., 2021). Deep learning uses multilayer neural networks such as graph neural networks to learn complex features of molecular structures and enhance the molecular representation, virtual screening, and de novo molecular design (Raghunathan & Priyakumar, 2022). Reinforcement learning further enables the iterative optimization of molecular structures by generating compounds with desired physicochemical and biological characteristics, thereby supporting automated molecular design (Chen et al., 2020).

2.3 Integration of Machine Learning with Computational Chemistry

Combining machine learning approaches with computational chemistry has revolutionized the traditional methods of molecular modeling by making them data-driven. Machine learning, alongside quantum-mechanical calculations and molecular simulations, provides rapid prediction of the properties of molecules, their reactions, and bioactivity through reduced computation costs and experimental requirements (Keith et al., 2021). Machine learning is used for molecular representation, virtual screening, reaction prediction, retrosynthesis analysis, molecular docking, and lead optimization. More sophisticated deep learning models along with the new generation of artificial intelligence help find new molecules, catalysts, polymers, and materials (Coley et al., 2018; Gao et al., 2024).

2.4 Chemical and Biomolecular Databases

Chemical and biomolecular databases have high-quality datasets which are used to build and validate ML models for computational chemistry. Such databases include data on molecular structures, physicochemical properties, biological activities, protein sequences, and three-dimensional structures that are useful for molecular modeling, virtual screening, and drug discovery (Keith et al., 2021). One of the biggest public databases of chemical compounds is the PubChem database which offers vast amounts of data related to chemical compounds, their molecular structures, biological activities, and bioassay data for computational analysis. The ChEMBL database provides bioactive molecules and information on drug-target interactions, which are experimentally validated and can be used in QSAR modeling, lead optimization, and pharmacology studies. The Protein Data Bank stores three-dimensional structures of proteins, nucleic acids, and macromolecules experimentally determined by various techniques. It is a critical database for molecular docking, studying protein-ligand interactions, and structure-based drug design (Naqvi et al., 2018). Another database useful in computational chemistry is the ZINC database which includes millions of compounds ready for virtual screening and de novo molecular design. UniProt helps in this context by offering protein sequence and functional annotation information along with biological pathways, allowing for the integration of chemical and protein data for research purposes (Glick & Patten, 2022). Together, all these sources serve as the basis of computational chemistry through machine learning by providing standard and dependable data for modeling.

3. Molecular Representation and Feature Engineering

Chemical and biomolecular representation and feature engineering form the basic parts of machine learning-driven computational chemistry because they allow for transforming chemical and biomolecular data to the

numerical format that can be further computed by machine learning models. Good representations ensure preservation of the chemical, physicochemical, and biological features and allow for making reliable predictions about the properties, biological activities, and interaction between molecules. The improvements made to molecular encoding approaches have led to the better performance of machine learning algorithms in molecular design, virtual screening, drug discovery, and biomolecular studies (Wigh et al., 2022; Keith et al., 2021). The main molecular and biomolecular representation approaches used in machine learning-driven computational chemistry are described in Table 1 below.

Table 1. Molecular and Biomolecular Representations Commonly Used in Machine Learning-Assisted Computational Chemistry

Representation	Description	Major Applications	References
Molecular Descriptors	Numerical physicochemical, topological, electronic, and geometrical features extracted from molecular structures	QSAR/QSPR, molecular property prediction	Raghunathan & Priyakumar (2022); Ma (2022)
Molecular Fingerprints	Binary or integer vectors representing molecular substructures and connectivity patterns	Similarity search, virtual screening, compound classification	Stepišnik et al. (2021)
SMILES	Text-based molecular notation encoding atoms, bonds, and molecular connectivity	Molecular databases, deep learning, molecular generation	Wigh et al. (2022); Zhu et al. (2020)
SELFIES	Robust molecular string representation ensuring chemically valid molecular structures	Generative AI, de novo molecular design	Chen et al. (2020)
Graph-Based Representation	Molecules represented as atom–bond graphs for graph learning algorithms	Graph neural networks, reaction prediction, molecular property prediction	Li et al. (2022); Keith et al. (2021)
Protein/Biomolecular Features	Protein sequence, structure, physicochemical, and functional annotations	Protein function prediction, molecular docking, drug discovery	Glick & Patten (2022); Rapley & Whitehouse (2015)

Each representation of molecules depicted in Table 1 highlights different aspects of their structure and biology. Choosing one can be influenced by the particular task of machine learning since the graph-based approach and the representation based on biomolecular features have been shown to be more efficient when predicting molecular properties and protein-ligand interactions.

3.1 Molecular Descriptors

Molecular descriptors are numeric representations of the physicochemical, topological, electronic, and geometrical characteristics of a molecule. Examples of molecular descriptors include molecular weight, lipophilicity, hydrogen-bond donor potential, polar surface area, and electronic parameters utilized in QSAR/QSPR studies. Molecular descriptors are interpretable features of molecules that make it possible to predict their properties, classify compounds, and optimize leads (Raghunathan & Priyakumar, 2022).

3.2 Molecular Fingerprints

Fingerprints represent molecular structures using binary or integer vector representations indicating the presence or absence of particular substructures, functional groups, and connectivity patterns. The use of fingerprints allows efficient searches for similarities, virtual screening, clustering, and classification while decreasing computational complexity. Being suitable for conventional machine learning algorithms, fingerprints are one of the most popular approaches to representing molecules in computational chemistry (Sun et al., 2019; Stepišnik et al., 2021).

3.3 Smiles and Selfies

Simplified Molecular Input Line Entry System (SMILES) is a text string representing molecules by encoding atoms, bonds, and molecular connectivity. Although being very popular in cheminformatics, classical SMILES can produce an incorrect representation of molecules during the automated generation of molecules. SELFIES (Self-Referencing Embedded Strings) resolve this problem by ensuring syntactic validity of each generated string and thus are well suited for deep learning and generative molecular design (Wigh et al., 2022).

3.4 Graph-Based Molecular Representations

Graph-based molecular representations represent molecules as graphs where atoms form the nodes while chemical bonds form the edges of the graph. Graph-based learning has been successfully employed for improving prediction

of molecular properties, chemical reactions, protein-ligand interaction studies, and designing new molecules (Li et al., 2022).

3.5 Representation of Proteins and Other Biomolecules

In addition to small molecules, machine learning approaches are now requiring effective representations of proteins and other biomolecules. Protein feature representations consider amino acid sequence information, three-dimensional structure information, physicochemical information, evolutionary information, and functional annotations for characterization of biomolecular interactions. These feature representations improve prediction of protein functions, target identification, molecular docking, and structure-based drug discovery approaches (Hong et al., 2024; Glick & Patten, 2022).

4. Machine Learning Algorithms in Computational Chemistry

Machine learning algorithms have become integral to computational chemistry as they allow fast predictions of the molecular properties, reaction results, and biological activity from vast chemical data sets. Compared to traditional approaches to computational chemistry, machine learning algorithms increase predictive efficiency, decrease computational costs, and make the process faster (Keith et al., 2021; Cova & Pais, 2019). The choice of a proper algorithm is based on the nature of a chemical problem, the dataset properties, and required predictive efficiency. Machine learning algorithms impact various steps of computational chemistry including molecular property prediction, de novo molecular generation, and biological applications. Figure 2 represents the key machine learning algorithms and their roles in computational chemistry processes (Keith et al., 2021).

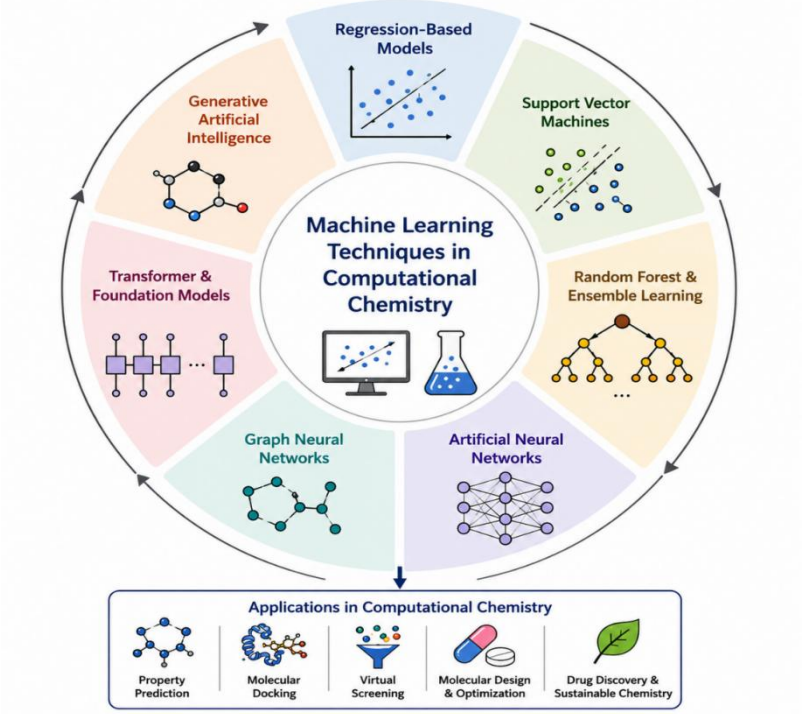

Figure 2. Machine learning techniques applied across computational chemistry workflows

Figure 2 depicts how machine learning algorithms have complementary capabilities in the computational chemistry process. While algorithms like regression, SVM, and Random Forest are reliable prediction models, deep learning, graph neural network, transformers, and generative AI offer advanced molecular representation, generation, and smart molecular design and hence boost the discovery and sustainable development of chemicals.

4.1 Regression-based Models

Regression models represent one of the first sets of machine learning algorithms to be employed in computational chemistry to predict continuous molecular properties including solubility, toxicity, binding affinity, and energies of reactions. Linear and non-linear regressions create relationships between molecular descriptors and target properties offering efficient and interpretable predictions. This method is still widely utilized in QSAR and QSPR analysis (Raghunathan & Priyakumar, 2022).

4.2 Support Vector Machines

Support vector machines (SVMs) are supervised learning models that can efficiently classify and predict the molecular properties using optimal decision boundaries in multidimensional space chemical data. Nonlinear nature of relations handled by using kernel functions in SVMs have enabled their application in compound classification, toxicological assessment, virtual screening, and bioactivity evaluation (Stepišnik et al., 2021).

4.3 Random Forest and Ensemble Learning

Random Forest is one of the ensemble methods used in machine learning, where a set of decision trees are used in order to enhance the accuracy of predictions. Ensembles methods have been widely applied in the field of molecular property prediction, reaction prediction, and compounds selection due to their robustness in capturing complex relationships between molecular descriptors (Keith et al., 2021).

4.4 Artificial Neural Networks

Artificial Neural Networks (ANN) are composed of interconnected computational neurons able to learn the non-linear dependencies from large data sets. The use of ANNs has greatly improved the prediction of properties of molecules, optimizations of reactions, and QSAR modeling. The versatility of ANNs makes them ideal for describing complex
chemical systems that cannot be easily described using traditional statistical methods (Cova & Pais, 2019).

4.5 Graph Neural Networks

Graph Neural Networks (GNN) model molecules directly as atom-bond graphs, enabling the learning of structural information about them without requiring any additional features engineering. Such approach has greatly improved the molecular property prediction, the analysis of protein-ligand interactions, reaction prediction, and de novo molecule design (Li et al., 2022; Hong et al., 2024).

4.6 Transformer and Foundation Models

Transformer models employ attention-based systems to capture long-range dependencies in molecular sequences and chemical representations. Foundation models pre-trained on big datasets of chemicals enable the prediction of molecular properties, reactions, molecular design, and analysis of protein sequences and show good generalization capabilities for many computational chemistry applications (Yue et al., 2025).

4.7 Generative Artificial Intelligence

Generative artificial intelligence allows for designing new molecular compounds possessing desired physicochemical and biological properties. Machine learning models that are based on the architectures of variational autoencoders, generative adversarial networks, and diffusion systems allow one to design novel molecules, lead optimization, polymers, and functional materials in a highly efficient way without any need for experimental screening (Chen et al., 2020; Gao et al., 2024).

5. Machine Learning-Assisted Molecular Design

Machine learning has revolutionized molecular design through fast molecular predictions, optimizations, and designs based on specific physical and biological characteristics. Through machine learning integration of chemical databases and algorithms, this process is optimized through less experimentation and faster time to market for both the drug discovery process and materials discovery. Such machine learning techniques enhance traditional computational chemistry techniques in terms of developing accurate and efficient molecular design techniques (Keith et al., 2021; Chen et al., 2020). Applications of these techniques in molecular design include molecular property prediction, structure-based drug discovery, and biomolecular interactions. Major applications of these techniques have been presented in Table 2.

Table 2. Applications of Machine Learning in Accelerated Molecular Design

Application	Primary Objective	Representative Use	References
Molecular Property Prediction	Predict physicochemical and biological properties	Solubility, toxicity, binding affinity	Li et al. (2022); Keith et al. (2021)
QSAR/QSPR Modeling	Relate molecular structure to biological activity or physicochemical properties	Drug discovery, toxicity prediction	Raghunathan & Priyakumar (2022); Cova & Pais (2019)
Virtual Screening	Identify promising compounds from large chemical libraries	Hit identification	Coley et al. (2018); Naqvi et al. (2018)
De Novo Molecular Design	Generate novel molecules with desired characteristics	Drug and material discovery	Chen et al. (2020); Yue et al. (2025)
Drug Discovery & Lead Optimization	Improve potency, selectivity, and pharmacokinetics	Therapeutic candidate optimization	Zhang et al. (2023); Hong et al. (2024)
Protein–Ligand Interaction Prediction	Predict molecular recognition and binding affinity	Target identification	Linkuviene et al. (2018); Hong et al. (2024)
Molecular Docking & Molecular	Simulate ligand binding and molecular behavior	Structure-based drug design	Salmaso & Moro (2018); Singh et al. (2022)

Dynamics			

From Table 2, it can be seen that machine learning has extended the process of molecule design to include rapid prediction, optimization, and synthesis of new molecules through computational chemistry, contributing to molecular discovery, efficient drug development, and the design of new materials and chemicals.

5.1 Molecular Properties Prediction

One of the major uses of machine learning in computational chemistry is molecular properties prediction. Machine learning algorithms predict various physical, electronic, and biological properties of the molecule including solubility, toxicity, stability, and binding affinity from its structure alone. Accurate prediction helps to filter promising molecules for experiments prior to testing, thus lowering costs of development (Li et al., 2022).

5.2 QSAR and QSPR Modeling

The models of QSAR and QSPR create relationships between the structural features and bioactivity or physical and chemical properties of molecules. Machine learning provides further improvement in creating more complicated and non-linear relations, allowing more accurate predictions about bioactivity, toxicity, pharmacokinetics, and performance of molecules (Raghunathan & Priyakumar, 2022).

5.3 Virtual Screening

Virtual screening allows using machine learning algorithms to quickly analyze large chemical libraries and find potential hits with desired bioactivity. Using machine learning virtual screening is an improved variant of conventional approaches as it is able to detect more hits and avoid false-positive predictions (Coley et al., 2018).

5.4 De Novo Molecular Design

The method of de novo molecular design creates totally new molecular structures using generative machine learning models. Using advanced generative AI algorithms allows exploring chemical space much more efficiently and discovering new molecular structures such as drugs, polymers, catalysts, and functional materials (Chen et al., 2020; Yue et al., 2025).

5.5 Drug Discovery and Lead Optimization

Machine learning facilitates drug discovery through the prediction of biological activity, optimization of lead compounds, and selection of potential drugs. AI-guided lead optimization contributes to an increase in the efficiency of pharmaceutical development due to molecular potency, selectivity, pharmacokinetics, and increased safety (Hong et al., 2024).

5.6 Protein-Ligand Interactions Prediction

The prediction of protein-ligand interactions is important for elucidation of molecular recognition and treatment efficacy. ML algorithms allow predicting binding affinity and interaction patterns based on structural and biochemical properties of substances (Hong et al., 2024; Linkuviene et al., 2018).

5.7 Molecular Docking and Molecular Dynamics Simulations

ML applications become more popular in combination with molecular docking and MD simulations, allowing an improvement in docking precision, acceleration of conformational sampling, and enhanced predictions of binding affinities (Salmaso & Moro, 2018; Singh et al., 2022).

6. Applications in Molecular Biology and Genetics

The merging of machine learning enabled computational chemistry with molecular biology and genetics has greatly broadened the scope of molecular research through the ability to conduct in-depth studies on genes, proteins, and biomolecules associated with diseases. Computational methods using artificial intelligence assist in identifying targets, characterizing proteins, conducting pharmacogenomics studies, modeling biomolecular networks, and practicing precision medicine, thus helping to fast-track the process of drug discovery and advance molecular biology and biomedicine research (Keith et al., 2021; Glick & Patten, 2022). Integration of computational chemistry, machine learning, and molecular biology has formed an integrated approach for drug discovery, bridging molecular information with computational prediction models. Figure 3 illustrates this interdisciplinary workflow.

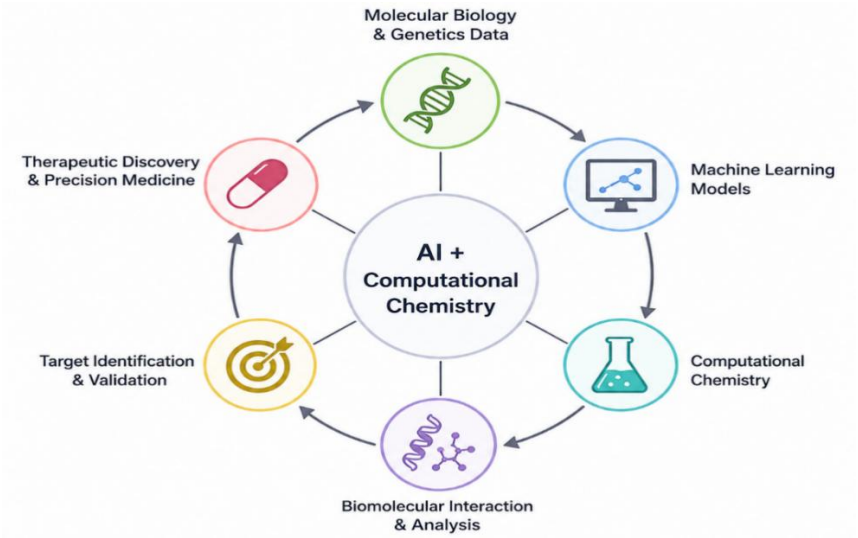

Figure 3. Integration of Computational Chemistry with Molecular Biology and Genetics for Therapeutic Discovery

Figure 3 illustrates how machine learning can be utilized as an intermediary that connects computational chemistry and molecular biology through the combination of genetic and biomolecular data with predictive computational methods. It results in improved identification of drug targets, proteins, biomarkers, and precision medicine that leads to the development of effective therapeutic solutions.

6.1 Artificial Intelligence for Identification of Drug Targets

Artificial intelligence speeds up the process of drug target identification through the combination of chemical, genomic, proteomic, and biological data to identify disease-related proteins and signal pathways. Through machine learning, possible drug targets are chosen based on their interaction with drugs, which reduces the number of experiments conducted at an early stage of drug discovery (Hong et al., 2024).

6.2 Prediction and Annotation of Protein Structure

The process of protein structure prediction and annotation has greatly benefited from machine learning through the use of amino acid sequence, structural, and evolutionary data. Such techniques help improve protein function understanding and facilitate drug design (Glick & Patten, 2022).

6.3 Computational Chemistry in Pharmacogenomics

The combination of computational chemistry and machine learning is helpful for pharmacogenomics in predicting the impact of genetic variations on drug metabolism, pharmacodynamics, and adverse effects on patients. Thus, computational tools contribute to the personalized approach in terms of selecting drugs that would be more effective and reduce the risk of adverse reactions (Patwardhan & Sharma, 2022; Song et al., 2025).

6.4 Biomolecular Interaction Networks Analysis

Using machine learning algorithms helps analyze biomolecular interaction networks that include different proteins, genes, metabolites, signaling pathways, and other elements. Such network-based computational models can help find the relationships between molecular processes, regulation, and disease-associated biomolecular interactions (Silverman et al., 2020; Sit et al. 2024).

6.5 Machine Learning for Disease-Associated Molecular Discovery

With the help of machine learning techniques, researchers can discover genes and proteins associated with diseases as well as other molecular biomarkers. This can contribute to the faster process of discovering new treatments (Koc et al., 2020; Patwardhan & Sharma, 2022).

6.6 Applications in Cancer and Precision Medicine

In oncology, machine learning-assisted computational chemistry supports the identification of cancer biomarkers, prediction of therapeutic response, and optimization of targeted drug design. Integrating computational modeling with molecular and genetic information enables personalized treatment strategies, improves drug efficacy, and advances precision medicine for cancer management (Hong et al., 2024; Glick & Patten, 2022).

7. Machine Learning for Sustainable Chemical Innovation

The contribution of machine learning to sustainable chemical innovations consists in the possibility to develop eco-friendly molecules, optimize the process, and discover sustainable materials through machine learning techniques. It is possible to save resources and minimize the use of experimental waste, making the chemical

manufacturing process more effective by combining computational chemistry and AI algorithms (Leonard et al., 2021; Osman et al., 2023). These approaches are useful for the creation of green technology and the implementation of sustainable industry and pharmaceutical practices. Machine learning assists sustainable chemical innovations through environmental molecular design, catalysis, reactions, and sustainable materials discovery. Major applications of machine learning to sustainable chemical innovations are presented in Table 3.

Table 3. Machine Learning Applications Supporting Sustainable Chemical Innovation

Application	Machine Learning Contribution	Sustainability Benefit	References
Green Molecular Design	Predicts environmentally friendly molecular structures	Reduced chemical toxicity and waste	Prabhakaran et al. (2024); Leonard et al. (2021)
Sustainable Catalyst Discovery	Identifies efficient catalyst materials	Improved energy efficiency and reaction selectivity	Wang et al. (2024); Zhang et al. (2022)
Reaction Prediction & Synthesis Optimization	Optimizes reaction pathways and reaction conditions	Reduced reagent consumption and waste generation	Coley et al. (2018); Meuwly (2021)
Eco-Friendly Material Discovery	Screens renewable and advanced functional materials	Supports clean energy and sustainable materials	Gu et al. (2019); Osman et al. (2024)
Sustainable Pharmaceutical Development	Optimizes drug discovery and lead selection	Lower development cost and environmentally responsible manufacturing	Leonard et al. (2021); Osman et al. (2023); Uddin et al. (2023)

As presented in Table 3, machine learning helps ensure sustainability in various stages of chemical innovation through efficient molecular design, chemical process optimization, and faster development of environmentally friendly materials and drugs. The above applications together help develop sustainable technologies, conserve resources, and practice sustainable production.

7.1 Green Molecular Design

Machine learning enables green molecular design through molecular property prediction and synthesis of compounds with higher efficiency and low environmental effect. AI-based molecular optimization helps reduce hazardous chemical synthesis and development of environmentally friendly molecules (Prabhakaran et al., 2024; Du et al., 2022).

7.2 Sustainable Catalyst Discovery

Catalyst is needed to ensure reaction efficiency with reduced energy costs and less waste production. Machine learning allows for faster catalyst discovery and analysis through prediction of catalytic activity, stability, and selectivity, which helps identify sustainable catalyst materials (Wang et al., 2024; Koc et al., 2020).

7.3 Reaction Prediction and Synthesis Optimization

Reaction pathway prediction, reaction condition optimization, and automated synthesis planning help predict reaction results, increase chemical yield, and minimize the reagents used (Coley et al., 2018; Meuwly, 2021).

7.4 Eco-Friendly Material Discovery

Machine learning-enabled computational chemistry has made the process of identifying sustainable materials easier, such as renewable polymers, energy storage materials, hydrogen storage materials, and renewable energy materials. Predictive modeling facilitates the quick identification of properties of materials without being overly dependent on expensive experimentation (Gu et al., 2019; Osman et al., 2024).

7.5 Sustainable Pharmaceutical Development

Machine learning plays an important role in sustainable pharmaceuticals by aiding in lead optimization, avoiding experimental errors, and minimizing chemical waste in pharmaceutical research and development (Leonard et al., 2021).

8. Emerging Trends, Challenges, and Future Perspectives

The field of computational chemistry is evolving due to the developments in generative artificial intelligence (AI), large language models, explainable AI, and multi-omics approaches. Generative models allow the quick development of new compounds with the required physicochemical and biological properties, while large language models allow efficient knowledge extraction, reaction prediction, and automation of the scientific workflow. Multi-omics approach combining genomics, proteomics, metabolomics, and molecular modeling allows efficient biomarker discovery, drug target identification, and precision medicine. Moreover, autonomous

platforms for molecule discovery that use machine learning, computational chemistry, and lab automation allow faster molecular design and screening.

However, there are still some challenges that limit the application of machine learning-assisted computational chemistry. One of the problems is the absence of high-quality and diverse datasets, as the performance of models highly relies on the quality of the input data. Also, many advanced machine learning methods lack the interpretability, which makes the analysis of the predictions challenging. Other challenges involve the problem of the computational cost, poor model generalization, the combination of different kinds of chemical and biological data, and experimental validation.

Future research should address the issue of creating interpretable and transparent artificial intelligence methods, as well as the creation of a universal database of chemicals and biomolecules. Further developments in the area of foundation models, autonomous laboratories driven by AI, and physics-informed machine learning will bring new opportunities for accurate molecule prediction, reaction optimization, and green chemistry. Cooperation between chemists, molecular biologists, computer scientists, and data scientists will play an important role in applying computational results to practical solutions related to drug development, precision medicine, and advanced materials.

9. CONCLUSION

Machine learning-enabled computational chemistry is a groundbreaking way of enhancing the speed of molecular design and achieving sustainable chemical innovation. With the help of the integration of computational chemistry and artificial intelligence, one can effectively predict molecular properties, optimize chemical structures, screen molecules virtually, and develop new substances with enhanced properties. In addition, such an integration has considerably reduced the effort, costs, and complexity involved in traditional molecular discovery and made the process more effective. Apart from molecular design, the merging of computational chemistry with machine learning has broadened the scope of computational chemistry beyond molecular design and into molecular biology, genetics, pharmacogenomics, protein structure analysis, and precision medicine. The integration of biomolecular databases, enhanced molecular representation, graph-based machine learning, and generative artificial intelligence has improved drug targets identification, protein-ligand interaction prediction, and structure-based drug discovery. Besides, artificial intelligence-driven computation approaches have helped to advance sustainable chemistry via facilitating green molecular design, catalyst discovery, reactions optimization, and eco-friendly material and pharmaceutical production. In spite of significant advancements made in the field, there still remain challenges such as poor data quality, computational complexity, lack of interpretability, standardization, and validation of results which limit broad applications of AI in computational chemistry. These challenges can be solved through the use of explainable AI, standardized databases, multidisciplinary approaches, and autonomous molecular design platforms. With time and development, artificial intelligence will become an even more important component of computational chemistry which is anticipated to advance many areas of molecular sciences, therapy development, and sustainable chemistry in the future.

REFERENCES

1. Chen, G., Shen, Z., Iyer, A., Ghumman, U. F., Tang, S., Bi, J., ... & Li, Y. (2020). Machine-learning-assisted de novo design of organic molecules and polymers: opportunities and challenges. Polymers, 12(1), 163.
2. Coley, C. W., Green, W. H., & Jensen, K. F. (2018). Machine learning in computer-aided synthesis planning. Accounts of chemical research, 51(5), 1281-1289.
3. Cova, T. F., & Pais, A. A. (2019). Deep learning for deep chemistry: optimizing the prediction of chemical patterns. Frontiers in chemistry, 7, 809.
4. Du, S., Wang, X., Wang, R., Lu, L., Luo, Y., You, G., & Wu, S. (2022). Machine-learning-assisted molecular design of phenylnaphthylamine-type antioxidants. Physical Chemistry Chemical Physics, 24(21), 13399-13410.
5. Gao, L., Lin, J., Wang, L., & Du, L. (2024). Machine learning-assisted design of advanced polymeric materials. Accounts of Materials Research, 5(5), 571-584.
6. Glick, B. R., & Patten, C. L. (2022). Molecular biotechnology: principles and applications of recombinant DNA. John Wiley & Sons.
7. Gu, G. H., Noh, J., Kim, I., & Jung, Y. (2019). Machine learning for renewable energy materials. Journal of Materials Chemistry A, 7(29), 17096-17117.
8. Hong, Y., Ha, J., Sim, J., Lim, C. J., Oh, K. S., Chandrasekaran, R., ... & Lee, J. (2024). Accurate prediction of protein–ligand interactions by combining physical energy functions and graph-neural networks. Journal of Cheminformatics, 16(1), 121.
9. Keith, J. A., Vassilev-Galindo, V., Cheng, B., Chmiela, S., Gastegger, M., Muller, K. R., & Tkatchenko, A. (2021). Combining machine learning and computational chemistry for predictive insights into chemical systems. Chemical reviews, 121(16), 9816-9872.
10. Koc, E., Pantazis, E., Soibelman, L., & Gerber, D. J. (2020). Emerging trends and research directions. Construction 4.0, 460-476.
11. Leonard, K. C., Hasan, F., Sneddon, H. F., & You, F. (2021). Can artificial intelligence and machine learning be used to accelerate sustainable chemistry and engineering?. ACS Sustainable Chemistry & Engineering, 9(18), 6126-6129.

12. Li, Z., Jiang, M., Wang, S., & Zhang, S. (2022). Deep learning methods for molecular representation and property prediction. Drug Discovery Today, 27(12), 103373.
13. Linkuviene, V., Talibov, V. O., Danielson, U. H., & Matulis, D. (2018). Introduction of intrinsic kinetics of protein–ligand interactions and their implications for drug design. Journal of Medicinal Chemistry, 61(6), 2292-2302.
14. Ma, X. (2022, December). Development of computational chemistry and application of computational methods. In Journal of Physics: Conference Series (Vol. 2386, No. 1, p. 012005). IOP Publishing.
15. Meuwly, M. (2021). Machine learning for chemical reactions. Chemical Reviews, 121(16), 10218-10239.
16. Naqvi, A. A., Mohammad, T., Hasan, G. M., & Hassan, M. I. (2018). Advancements in docking and molecular dynamics simulations towards ligand-receptor interactions and structure-function relationships. Current topics in medicinal chemistry, 18(20), 1755-1768.
17. Osman, A. I., Nasr, M., Eltaweil, A. S., Hosny, M., Farghali, M., Al-Fatesh, A. S., ... & Abd El-Monaem, E. M. (2024). Advances in hydrogen storage materials: harnessing innovative technology, from machine learning to computational chemistry, for energy storage solutions. International journal of hydrogen energy, 67, 1270-1294.
18. Osman, A. I., Zhang, Y., Lai, Z. Y., Rashwan, A. K., Farghali, M., Ahmed, A. A., ... & Yap, P. S. (2023). Machine learning and computational chemistry to improve biochar fertilizers: a review. Environmental Chemistry Letters, 21(6), 3159-3244.
19. Patwardhan, D., & Sharma, N. (2022). Application of Molecular Genetics. In Genetics Fundamentals Notes (pp. 761-802). Singapore: Springer Nature Singapore.
20. Prabhakaran, P., Bhardwaj, S., Chopra, B., Dhingra, A. K., & Kant, M. (2024). Chemistry and Technology Innovation to Advance Green and Sustainable Chemistry. Sustainable Green Catalytic Processes, 273-300.
21. Raghunathan, S., & Priyakumar, U. D. (2022). Molecular representations for machine learning applications in chemistry. International Journal of Quantum Chemistry, 122(7), e26870.
22. Rapley, R., & Whitehouse, D. (Eds.). (2015). Molecular biology and biotechnology. Royal Society of Chemistry.
23. Salmaso, V., & Moro, S. (2018). Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. Frontiers in pharmacology, 9, 923.
24. Silverman, A. D., Karim, A. S., & Jewett, M. C. (2020). Cell-free gene expression: an expanded repertoire of applications. Nature reviews genetics, 21(3), 151-170.
25. Singh, S., Baker, Q. B., & Singh, D. B. (2022). Molecular docking and molecular dynamics simulation. In Bioinformatics (pp. 291-304). Academic Press.
26. Sit, M. K., Das, S., & Samanta, K. (2024). Machine Learning-Assisted Mixed Quantum-Classical Dynamics without Explicit Nonadiabatic Coupling: Application to the Photodissociation of Peroxynitric Acid. The Journal of Physical Chemistry A, 128(38), 8244-8253.
27. Song, M., Hu, Z., Wang, M., Jia, S., Cao, F., Duan, L., ... & Wang, R. (2025). Machine Learning-Assisted Design of Molecular Structure of Diphenylamine Antioxidants. ACS omega, 10(30), 33063-33078.
28. Stepišnik, T., Škrlj, B., Wicker, J., & Kocev, D. (2021). A comprehensive comparison of molecular feature representations for use in predictive modeling. Computers in Biology and Medicine, 130, 104197.
29. Sun, W., Zheng, Y., Yang, K., Zhang, Q., Shah, A. A., Wu, Z., ... & Sun, K. (2019). Machine learning–assisted molecular design and efficiency prediction for high-performance organic photovoltaic materials. Science advances, 5(11), eaay4275.
30. Uddin, M., Manickam, S., Ullah, H., Obaidat, M., & Dandoush, A. (2023). Unveiling the metaverse: Exploring emerging trends, multifaceted perspectives, and future challenges. IEEE access, 11, 87087-87103.
31. Wang, X., Shi, K., Peng, A., & Snurr, R. Q. (2024). Computational Chemistry and Machine Learning-Assisted Screening of Supported Amorphous Metal Oxide Nanoclusters for Methane Activation. ACS Catalysis, 14(24), 18708-18721.
32. Wigh, D. S., Goodman, J. M., & Lapkin, A. A. (2022). A review of molecular representation in the age of machine learning. Wiley Interdisciplinary Reviews: Computational Molecular Science, 12(5), e1603.
33. Yue, T., He, J., & Li, Y. (2025). Machine-Learning-Assisted Molecular Design of Innovative Polymers. Accounts of Materials Research, 6(8), 1033-1045.
34. Zhang, H., Zhao, H., Zhang, X., Su, Q., Du, H., Shen, C., ... & Hou, T. (2023). Delete: Deep lead optimization enveloped in protein pocket through unified deleting strategies and a structure-aware network. arXiv preprint arXiv:2308.02172.
35. Zhang, Q., Zheng, Y. J., Sun, W., Ou, Z., Odunmbaku, O., Li, M., ... & Sun, K. (2022). High-efficiency non-fullerene acceptors developed by machine learning and quantum chemistry. Advanced Science, 9(6), 2104742.
36. Zhao, Q., Shan, Y., Zhou, H., Zhang, G., & Liu, W. (2023). Machine learning-assisted performance prediction and molecular design of all-small-molecule organic solar cells based on the Y6 acceptor. Solar Energy, 265, 112115.
Zhu, H., Li, C., & Gao, C. (2020). Applications of CRISPR–Cas in agriculture and plant biotechnology. Nature Reviews Molecular Cell Biology, 21(11), 661-677.