Proceedings of ICBioMed 2025 Symposium: AI for Healthcare: Advanced Medical Data Analytics and Smart Rehabilitation
DOI: 10.54254/2753-8818/2025.AU27582

Scaffold Optimization Strategies in Medicinal Chemistry: Chemical Approaches and Artificial Intelligence Applications

Youci Ye

University of Glasgow,Glasgow, United Kingdom
yeyouci629@gmail.com

Abstract. Chemical skeleton optimization is a core strategy in medicinal chemistry. By adjusting the chemical skeleton to improve biological activity, selectivity and ADMET properties, it can enhance the drugability of drug candidate molecules. In chemical methods, such as skeleton transitions, heterocyclic substitution, and skeleton editing have been applied to enhance efficacy, selectivity, and pharmacokinetic properties. These strategies allow researchers to explore new chemical space, overcome limitations of existing leads, and develop molecules with better safety profiles. In recent years, AI technology has completely transformed scaffold design by deeply integrating skeleton transition algorithms with reinforcement learning frameworks to achieve data-driven molecular representation, generative modeling, and structure-activity prediction. This article will describe the optimization strategies of chemical skeletons, the application of AI in skeleton transitions, case analyses, future trends and challenges. The synergy between chemical expertise and artificial intelligence is expected to accelerate drug discovery and increase the success rate of new therapy development.

Keywords: Skeleton optimization, scaffold hopping, chemical scaffolds

1. Introduction

The role of the skeleton in drug design is mainly reflected in optimizing molecular structure, enhancing efficacy and pharmacokinetic properties. The scaffold is the core structure of a molecule. It supports functional groups through stable carbon chains or other chemical bonds and determines the overall three-dimensional conformation of the compound, providing a stable three-dimensional spatial arrangement for the molecule [1]. Scaffold optimization is a core strategy in drug design, which refers to the modification or replacement of the core skeleton structure of small molecule compounds to enhance their pharmacological activity, selectivity, and metabolic stability, while retaining their basic pharmacodynamic characteristics [2].

Scaffold hopping is a strategy in drug discovery to obtain compounds with novel structures but similar activities by replacing the core skeleton of molecules, while preserves the key interactions with the target. This can expand the chemical space and break through patent limitations [3].

With the advancement of computational chemistry and artificial intelligence, scaffold design has evolved from a purely empirical approach to a hybrid workflow that combines human expertise with machine learning models. Researchers can now explore the chemical space more effectively and

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2025.AU27582

systematically [4,5]. This paper reviews both classical chemical scaffold optimization strategies and recent AI-powered innovations, supplemented with case studies and future perspectives.

2. Fundamentals of chemical scaffolds

2.1. Definition and classification

Chemical scaffolds usually refer to the cyclic or rigid structural parts remaining in a molecule after all variable functional groups have been removed. They represent the basic connectivity of compounds and are core elements in the construction of compound libraries and drug design [1]. The classification of chemical skeletons is of vital importance to drug design and the study of structure-activity relationships. Schuffenhauer and Varin [6] proposed a rule-based hierarchical classification method, dividing chemical skeletons into five progressive levels: Level 1: Bemis-Murcko skeleton, retaining rings and connection bonds; Level 2: Ring connectivity skeleton, ignoring atomic types; Level 3: Topological skeleton, only considering the connection mode of rings; Level 4: Pure carbon skeleton, removing heteroatoms; Level 5: Simplify the graph to represent the loop system with nodes and edges. This classification system can systematically compare the core structures of different molecules, which is conducive to the identification of dominant skeletons, the clustering of compound libraries and the optimization of lead compounds, providing an important tool for medicinal chemistry and chemoinformatics research.

2.2. Scaffold hopping for SAR and pharmacophore verification

Pharmacophiles refer to the spatial arrangement of chemical characteristics responsible for biological activity, and they are also a set of structural and electronic features in a molecule that are necessary for its biological activity, especially for its interaction with specific biological targets [7]. Skeleton hopping is a powerful strategy for expanding and optimizing structure-activity relationship (SAR) research. It is not merely about seeking new structures, but also a key tool for understanding the mechanism of drug-target interactions. Its application can directly verify key pharmacophore hypotheses and distinguish between necessary interactions and redundant structures. This strategy effectively resolves the "molecular redundancy" dilemma in homolog optimization by introducing a structurally novel skeleton, greatly expanding the chemical space for SAR exploration. When the optimization of lead compounds encounters activity or property bottlenecks, skeleton hopping a new path to break through the original structural limitations, achieving a balance between maintaining activity and avoiding defects. Therefore, it has become a bridge connecting structural innovation and SAR research, promoting a paradigm shift in drug discovery from "fine-tuning" to "innovation" [7].

3. Skeleton transformation strategy

3.1. Chemical modification and functional group replacement

The "Rule of Five" (Ro5) revolutionarily reshaped the strategy for optimizing lead compounds, shifting the core objective of chemical modification from merely pursuing activity enhancement to a comprehensive optimization that balances potency and drug-like properties. This paradigm profoundly criticizes the traditional practice of enhancing activity by increasing molecular weight and lipid solubility (cLogP), which leads to "molecular obesity", and instead advocates Property-Based Optimization. One of the most direct strategies is to regulate physical and chemical properties by modifying the connected functional groups without changing the core framework. For instance,

DOI: 10.54254/2753-8818/2025.AU27582

bioelectronic and other complexes can be used to reduce cLogP, polar groups can be introduced to increase solubility, or fine structural adjustments can be made to enhance metabolic stability and reduce toxicity. Ro5 acts as a crucial decision filter in this process, guiding medicinal chemists to prioritize modification schemes that can enhance Ligand Efficiency, thereby systematically avoiding development risks and ultimately obtaining balanced molecules with both excellent activity and good drug-like properties [8].

3.2. Scaffold hopping

Skeleton hopping replaces the molecular core with different structures while retaining key pharmacodynamic elements, and it is a widely used drug modification strategy in both academic and industrial fields as shown in figure 1. This strategy focuses on modifying the nuclear structure of compounds to optimize the physicochemical properties, pharmacodynamic properties and pharmacokinetic properties of molecules has been widely applied in the design of drug molecules in both academic and industrial fields at present [9]. The methods of scaffold hopping can be classified into four main categories. The first is similarity-based methods, such as two-dimensional molecular similarity; Secondly, there are substructure-based methods, such as fragment-based methods; Then there is the method based on docking; Finally, there are rule-based methods, such as bioelectronic isosomal substitution [10].

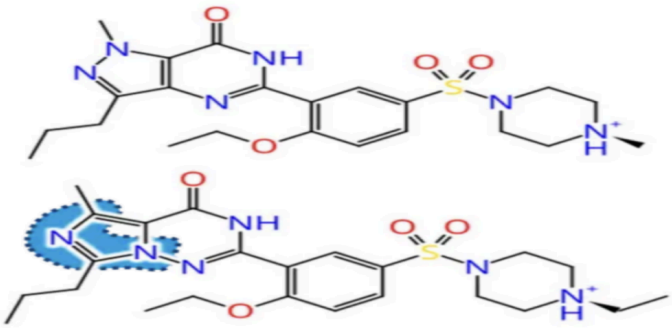


Figure 1. Scaffold hopping, from bioSolveIT

3.3. Heterocycle Replacement

Heterocycle Replacement is the simplest form of skeleton transition, which replaces the circular mother nucleus while the connected functional groups and pharmacophore remain basically unchanged. This strategy can obtain new chemical entities while maintaining the key ligand-target interactions optimize the physicochemical properties and PK properties of the molecules as shown in figure 2.

DOI: 10.54254/2753-8818/2025.AU27582

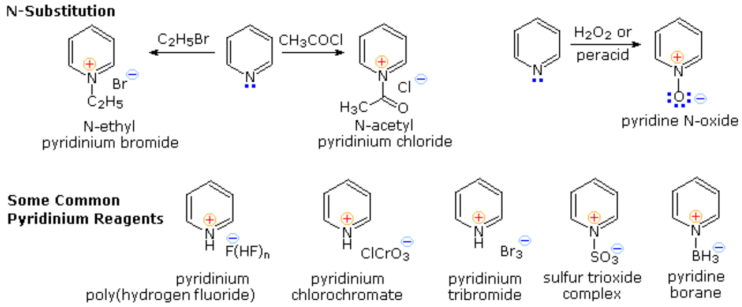


Figure 2. Heterocycle replacement, from MSU chemistry

In the optimization process of lead compounds, replacing the pyridine ring with its bioelectronic isotactic benzene ring is an important strategy for regulating the molecular logP value, solubility and target selectivity. In recent years, Conboy and Greaney have developed an efficient skeleton editing strategy, enabling the direct conversion of pyridine rings to benzene rings via 'N to C switch'. This method provides a powerful tool for the rapid construction of such analogue libraries and the acceleration of SAR research, making it more convenient to systematically evaluate the impact of pyridine or benzene ring substitution on various properties such as solubility and selectivity [11].

3.4. Skeleton editing

Skeleton editing is a process that alters the atomic composition and connection sequence of a molecule by making specific point changes to its core skeleton, such as selective insertion, deletion or exchange of atoms. This method enables direct reprogramming of the "genes" of molecules, demonstrating the huge opportunities existing in this field [12].

The mutual transformation between aromatic rings and heterocyclic rings is a very important type of skeleton editing reaction. This strategy abandons the traditional synthetic path and achieves direct conversion between aromatic hydrocarbons and heterocyclic hydrocarbons by performing atomic-level precise operations on the molecular skeleton, such as inserting nitrogen atoms into aromatic rings to generate heterocyclic rings like reducing pyridine to a benzene ring through denitrification reactions. These strategies include Atomic Insertion, Atomic Deletion, Ring Contraction or Expansion [12] as shown in figure 3.

DOI: 10.54254/2753-8818/2025.AU27582

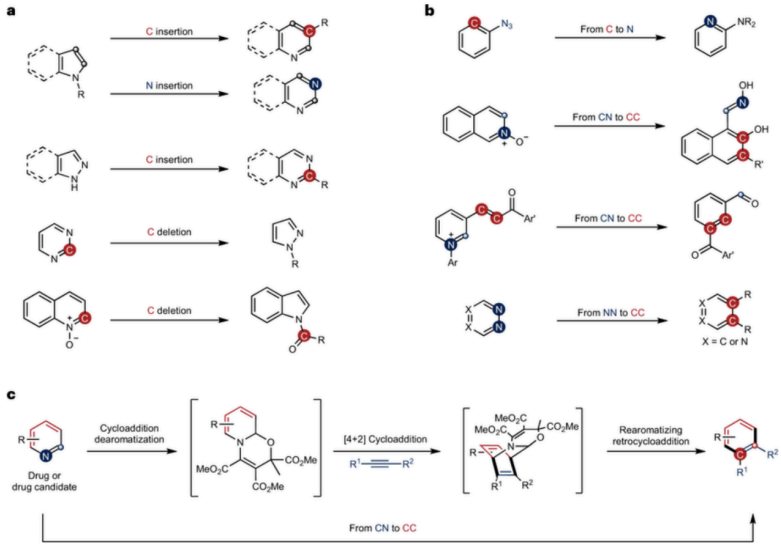


Figure 3. Skeleton editing from researchgate

3.5. Influence of chemical skeleton changes on solubility and binding mode

In the influence of chemical frameworks on solubility, introducing polar functional groups can enhance the solubility of molecules in polar solvents such as water, while increasing hydrophobic groups will reduce solubility [13] and the new framework can promote or hinder the formation of favorable intermolecular forces, such as hydrogen bonds with the solvent, which is crucial for dissolution. In the influence on the binding mode, the altered chemical structure can cause conformational changes at the ligand or albumin binding sites, thereby altering their binding mode and the specific binding mode adopted. When the molecular structure is modified, it may migrate to different binding domains on the protein [14].

3.6. Case analysis: the skeleton optimization process of marketed drugs

Structural simplification is an efficient and powerful strategy in lead compound optimization. This strategy aims to reduce molecular complexity by removing redundant functional groups, reducing chiral centers and simplifying the ring system structure. Eventually, it often leads to the acquisition of candidate molecules with better activity, higher selectivity and better drug-like properties [15] as shown in figure 4.

The essence of this process lies in the meticulous optimization and reconstruction of the molecular skeleton. Taking the simplification from Pravastatin to the development of Rosuvastatin as an example, the structure of the original lead drug Pravastatin is relatively complex, containing multiple chiral centers and union ring structures. By simplifying the ring structure and replacing its decanhydride naphthalene ring structure with a pyrimidine pyrrole double heterocyclic ring, or by optimizing the functional groups to retain the key β -hydroxy acid introducing sulfonyl and methoxy groups to enhance the binding with MG-CoA reductase, Rosuvastatin is finally obtained. This method brings many advantages, such as significantly enhanced activity, stronger binding of the simplified structure to the target, and higher therapeutic effect; Improved pharmacokinetic properties, better metabolic stability, and a longer half-life; The synthesis process is simplified, the molecular symmetry is higher, and there are fewer chiral centers, which greatly reduces the cost and difficulty of industrial production [15]. Through rational framework simplification, not only better

DOI: 10.54254/2753-8818/2025.AU27582

drugs are obtained, but also the fundamental electrification of the original drug in production and use is solved.

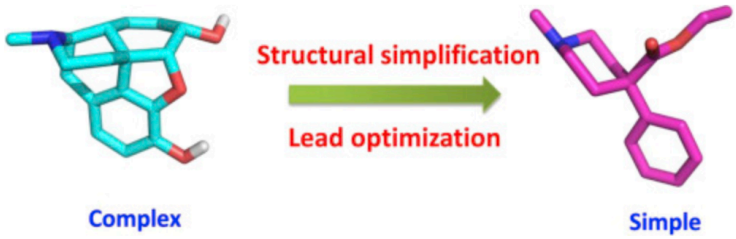


Figure 4. Structural simplification from sciencedirect

4. Application of AI in skeleton optimization

4.1. Application of AI in drug design

AI is used in drug design to identify disease targets, design novel drug molecules with desired characteristics, predict drug efficacy and toxicity, and optimize clinical trials. Compared with traditional methods, this technology significantly accelerates the drug development process and reduces costs. Technologies such as machine learning and deep learning can analyze massive biological and chemical datasets, thereby virtually screening compound libraries, predicting pharmacokinetic and pharmacodynamic properties, and guiding the development of more effective and personalized therapies [16].

4.2. Strategy for AI to generate new skeleton structures

Molecular graph neural networks (GNNs) have recently attracted significant attention in drug discovery because of their ability to capture complex structure–property relationships. Beyond property prediction, several studies have demonstrated that trained GNN models can also be reversed for molecular generation. In this approach, a GNN predictor is first established using molecular graph representations to learn correlations with target properties such as solubility or electronic energy gaps. Then, by applying gradient-based optimization strategies, the predictor can be backpropagated to iteratively modify molecular graphs, guiding the design of new compounds with desired properties. Compared with conventional generative models that rely on distribution learning, this strategy takes advantage of the differentiable nature of GNNs and integrates chemical knowledge more explicitly into the generation process. Such methods have shown promise in the optimization of lead compounds and the design of functional molecular materials, providing an alternative paradigm for structure-based drug design [17].

SMILES, or Simplified Molecular Linear Input Specification, is a chemical language that explicitly describes the structure of molecules using simple ASCII strings. Its advantages are simplicity and efficiency. The string form is very compact, making it easy to store, transmit and process. Human-readable, after training, chemists can intuitively "understand" SMILES strings compared to more complex notations. Computationally friendly; It is highly suitable as input for machine learning models, especially NLP models, because the models can handle the "vocabulary" and "grammar" of atoms and chemical bonds just like natural language. Widely adopted, it has become one of the most mainstream molecular representation formats in chemical information software, databases and online platforms.

DOI: 10.54254/2753-8818/2025.AU27582

However, there are also drawbacks and challenges, such as ineffectiveness. Machine learning models may generate strings that are grammatically invalid and do not conform to the SMILES rules or chemically invalid with incorrect atomic valence states. Ambiguity: The same molecular structure may have multiple valid SMILES representations. For instance, both CCO and OCC represent ethanol, which can cause ambiguity in machine learning models and require data standardization. Structural information loss: The linear format of strings makes it difficult to directly reflect the topological connections and spatial three-dimensional structure information of molecules. In the application of AI drug discovery, as the main representation method for AI molecule generation, models such as RNN, Transformer, and GAN can automatically generate new molecules with specific attributes after learning SMILES strings. As input to the QSPR model, the model learns features from SMILES strings to predict the biological activity, solubility and toxicity of molecules [18]. SMILES, with its unparalleled simplicity and friendliness to computers, has become a key bridge connecting chemistry and artificial intelligence. Despite challenges such as ineffectiveness and polysemy, it remains one of the most mainstream and important molecular representation methods in the field of AI-driven drug discovery to this day.

4.3. Virtual screening and activity prediction model

Virtual screening and activity prediction are the two most mature and core application fields of AI in drug discovery. Virtual screening is a traditional method. Screening based on molecular docking and screening based on drug clusters will be briefly mentioned. In terms of AI methods, machine learning-based learning screening can discuss how to utilize classification models or regression models to learn from known active or inactive compound data, thereby rapidly screening out potential active molecules. Deep models, on the other hand, emphasize the advantages of learning from structures like CNNs and GNNs that automatically extract complex features from molecular structures such as SMILES strings for screening [19]. Activity prediction is the most classic direction of AI application. The QSAR model will review how to use AI models to establish quantitative relationships between chemical structure and activity or properties [19].

4.4. Case studie: PLK1 inhibitor design

PLK1 is a core member of the Polo-like kinase family of highly conserved serine or threonine kinases and a key regulatory factor in the process of cell mitosis. The development of PLK1 inhibitors is a model of the combination of computer-aided drug design (CADD) and synthetic chemistry. Molecular docking is carried out by using the crystal structure of PLK1, and molecules that can form key interactions are rationally designed. Through methods such as 3D-QSAR and pharmacopod models, the structure-activity relationship of known active molecules is analyzed to guide the optimization of new compounds. Onvansertib is an oral, highly selective PLK1 inhibitor that has demonstrated promising efficacy and better tolerability in Phase II clinical trials for colorectal cancer, prostate cancer, and AML. The greatest challenge of PLK1 inhibitors is to strike a balance between effectively killing cancer cells and their toxicity to normal cells, especially the rapidly renewing bone marrow cells, and long-term use may lead to the development of drug resistance. In the future, PLK1 inhibitors will be combined with other chemotherapy drugs, targeted therapies or immune checkpoint inhibitors to enhance efficacy, reduce dosage and toxicity, and avoid inhibiting other members of the PLK family, which has already reduced the mop effect [20].

DOI: 10.54254/2753-8818/2025.AU27582

5. Discussion

Artificial intelligence models are highly dependent on the quality of training data. Many public datasets have measurement variability, analytical inconsistency and chemical spatial bias, especially in fields such as healthcare and drug development. These reasons stem from data collection bias, which leads to a lack of diversity in the algorithm and the data itself. In the future, researchers need to standardize data practices, collect large amounts of data, and ensure interoperability across datasets, which is crucial for building fair and reliable artificial intelligence algorithms. They should adopt statistical debias methods and conduct comprehensive and impartial model evaluations.

The combination of AI and skeleton optimization is driving drug discovery from an "art" to an "exact science". Future development will be multi-dimensional, with more intelligent models, more automated processes, and more usable outputs. The human-machine collaboration model will significantly enhance the success rate and efficiency of new drug research and ultimately bring more innovative therapies to patients more quickly.

6. Conclusion

Skeleton optimization remains an important strategy in medicinal chemistry, providing a way to enhance drug efficacy, selectivity and pharmacokinetics. Scaffold hopping, heterocyclic replacement and skeleton editing provide a powerful framework for structural innovation. Through molecular representation learning and generative modeling, artificial intelligence integration is transforming scaffold design, making the exploration of chemical Spaces faster, more diverse, and more targeted. The synergy between computational methods and experimental methods means that the design and validation of new therapies in the future will become more efficient.

References

[1] Brown N, Jacoby E. On scaffolds and hopping in medicinal chemistry. Mini Rev Med Chem. 2006 Nov; 6(11): 1217-29. doi: 10.2174/138955706778742768. PMID: 17100633.

[2] Schneider G. Automating drug discovery. Nat Rev Drug Discov. 2018 Feb; 17(2): 97-113. doi: 10.1038/nrd.2017.232. Epub 2017 Dec 15. PMID: 29242609.

[3] Wang, S., Zhang, R., Li, X. et al. Recent advances in molecular representation methods and their applications in scaffold hopping. npj Drug Discov. 2, 14 (2025). https: //doi.org/10.1038/s44386-025-00017-2.

[4] Han R, Yoon H, Kim G, Lee H, Lee Y. Revolutionizing Medicinal Chemistry: The Application of Artificial Intelligence (AI) in Early Drug Discovery. Pharmaceuticals (Basel). 2023 Sep 6; 16(9): 1259. doi: 10.3390/ph16091259. PMID: 37765069; PMCID: PMC10537003.

[5] Xia, Y., Wang, Y., Wang, Z., & Zhang, W. (2024). A comprehensive review of molecular optimization in artificial intelligence–based drug discovery. Quantitative Biology, 12(1), 15–29. https: //doi.org/10.1002/qub2.30.

[6] Schuffenhauer A, Varin T. Rule-Based Classification of Chemical Structures by Scaffold. Mol Inform. 2011 Aug; 30(8): 646-64. doi: 10.1002/minf.201100078. Epub 2011 Aug 16. PMID: 27467257.

[7] Acharya A, Yadav M, Nagpure M, Kumaresan S, Guchhait SK. Molecular medicinal insights into scaffold hopping-based drug discovery success. Drug Discov Today. 2024 Jan; 29(1): 103845. doi: 10.1016/j.drudis.2023.103845. Epub 2023 Nov 25. PMID: 38013043.

[8] Lipinski CA. Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol. 2004 Dec; 1(4): 337-41. doi: 10.1016/j.ddtec.2004.11.007. PMID: 24981612.

[9] Huo, T., Zhao, X., Cheng, Z., Wei, J., Zhu, M., Dou, X., & Jiao, N. (Year). Late-stage modification of bioactive compounds: Improving druggability through efficient molecular editing. Journal Name, Volume(Issue), pages. https: //doi.org/10.1016/j.apsb.2023.11.021.

[10] Sun, H., Tawa, G., & Wallqvist, A. (2012). Classification of scaffold hopping approaches. Drug Discovery Today, 17(7–8), 310–324. https: //doi.org/10.1016/j.drudis.2011.10.024.

DOI: 10.54254/2753-8818/2025.AU27582

[11] Conboy, A., & Greaney, M. F. (2024). Synthesis of benzenes from pyridines via N to C switch. Chem. Advance online publication. https: //doi.org/10.1016/j.chempr.2024.05.004.
[12] Joynson, B. W., & Ball, L. T. (2023). Skeletal Editing: Interconversion of Arenes and Heteroarenes. Helvetica Chimica Acta, 106(3), Article e202200182. https: //doi.org/10.1002/hlca.202200182.
[13] Mabesoone, M. F. J., Palmans, A. R. A., & Meijer, E. W. (2020). Solute–solvent interactions in modern physical organic chemistry: Supramolecular polymers as a muse. Journal of the American Chemical Society, 142(47), 19781–19798. https: //doi.org/10.1021/jacs.0c09293.
[14] Ashraf S, Qaiser H, Tariq S, Khalid A, Makeen HA, Alhazmi HA, Ul-Haq Z. Unraveling the versatility of human serum albumin - A comprehensive review of its biological significance and therapeutic potential. Curr Res Struct Biol. 2023 Nov 27; 6: 100114. doi: 10.1016/j.crstbi.2023.100114. PMID: 38111902; PMCID: PMC10726258.
[15] Wang, S., Dong, G., & Sheng, C. (2019). Structural simplification: An efficient strategy in lead optimization. European Journal of Medicinal Chemistry, 179, 623–636. https: //doi.org/10.1016/j.apsb.2019.05.004.
[16] Singh, S., Kaur, N., & Gehlot, A. (2023). Application of artificial intelligence in drug design: A review. Computational Biology and Chemistry, 107, 107018. https: //doi.org/10.1016/j.compbiomed.2024.108810.
[17] Therrien, F., Sargent, E.H. & Voznyy, O. Using GNN property predictors as molecule generators. Nat Commun 16, 4301 (2025). https: //doi.org/10.1038/s41467-025-59439-1.
[18] van Deursen, R., Ertl, P., Tetko, I.V. et al. GEN: highly efficient SMILES explorer using autodidactic generative examination networks. J Cheminform 12, 22 (2020). https: //doi.org/10.1186/s13321-020-00425-8.
[19] Bhat, A.R. & Ahmed, S., 2025. Artificial intelligence (AI) in drug design and discovery: A comprehensive review. Biomedicine & Pharmacotherapy, 141, p.111836. https: //doi.org/10.1016/j.insi.2025.100049.
[20] Oh Y, Jung H, Kim H, Baek J, Jun J, Cho H, Im D, Hah JM. Design and Synthesis of a Novel PLK1 Inhibitor Scaffold Using a Hybridized 3D-QSAR Model. Int J Mol Sci. 2021 Apr 8; 22(8): 3865. doi: 10.3390/ijms22083865. PMID: 33917995; PMCID: PMC8068361.