



ARTIFICIAL INTELLIGENCE-ASSISTED DRUG DESIGN COMBINED WITH MOLECULAR DOCKING FOR PRIORITIZATION OF POTENTIAL LEAD MOLECULES ACROSS FIVE THERAPEUTIC TARGETS

Harshveer Singh Jaitawat¹, Dr. C.S. Sharma¹, Dr. H.S. Udawat², Dr. P.S. Naruka²

¹Department of pharmaceutical chemistry, B.N. College of Pharmacy, Faculty of Pharmacy, B.N. University, Udaipur, Rajasthan, India.

²Department of pharmaceuticals, B.N. College of Pharmacy, Faculty of Pharmacy, B.N. University, Udaipur, Rajasthan, India.

Article DOI: <https://doi.org/10.36713/epra28874>

DOI No: 10.36713/epra28874

ABSTRACT

Background: Artificial intelligence-assisted drug discovery can reduce the chemical search space by integrating molecular generation, drug-likeness assessment and structure-based docking. However, computational rankings require careful validation and should be interpreted as prioritization rather than evidence of therapeutic efficacy.

Objectives: To integrate AI-assisted molecular design, physicochemical/ADMET screening, molecular docking and ligand-protein interaction analysis for identification of potential leads across selected cancer, metabolic and neurodegenerative targets.

Methods: A documented computational workflow was applied to five ligand-bound human targets—EGFR (1M17), KRAS G12C (5V71), DPP-4 (1X70), PTP1B (1C83) and AChE (4EY7). The supplied thesis dataset described progressive filtering of a 1350-member virtual library to 14 final candidates, followed by drug-likeness/ADMET assessment, AutoDock Vina docking and redocking validation. Candidate prioritization integrated docking score, target-relevant interactions and computational pharmacokinetic/toxicity characteristics.

Results: All five reference systems passed the predefined redocking criterion of $<2.0 \text{ \AA}$, with RMSD values of 0.65–1.14 $\text{\AA}$. The prioritized leads were CAN-AI-01 (EGFR, -9.71 kcal/mol), KRA-AI-01 (KRAS G12C, -9.12 kcal/mol), DIA-AI-01 (DPP-4, -9.35 kcal/mol), PTP-AI-01 (PTP1B, -8.84 kcal/mol) and NEU-AI-01 (AChE, -12.42 kcal/mol). All five showed zero reported Lipinski violations, high predicted gastrointestinal absorption, negative Ames predictions and no predicted hERG inhibition or hepatotoxicity in the supplied screening dataset.

Conclusions: The integrated workflow supports computational narrowing of chemical space and target-specific lead prioritization. The identified molecules should be considered computational leads requiring biochemical, cellular, pharmacokinetic and toxicity validation before any therapeutic claim is made.

KEYWORDS: Artificial intelligence, Drug design, Molecular docking, AutoDock Vina, Virtual screening, ADMET, Lead prioritization, Structure-based drug discovery

INTRODUCTION

In order to speed up target selection, virtual screening, molecule creation, and pharmacokinetic or toxicity property prediction, artificial intelligence (AI) is being included more and more into early-stage drug research. By evaluating chemical structures against predetermined biological targets prior to experimental synthesis, computer-aided drug design offers an organized substitute for merely empirical screening. This manuscript's thesis presented molecular docking and AI-assisted molecular design as complementing methods for reducing chemical space and finding molecules with likely target engagement. [29, 31, 50, 56]

In addition to estimating a ligand's preferred orientation within a protein binding site, structure-based molecular docking generates a computational affinity score that can be utilized for comparison ranking. Because AutoDock Vina combines effective optimization with a score feature intended for quick virtual screening, it is frequently utilized for this purpose. [17] This method is extended by contemporary AI-assisted workflows, which combine docking with drug-likeness and ADMET filters and generate or prioritize compounds inside chemically limited areas. The viability of de novo molecular design has been shown using deep generative models, reinforcement learning, and graph-based techniques; nonetheless, reproducibility and prospective validation continue to be significant obstacles. [34, 38, 39, 69, 70, 72]



The current study used an integrated computational framework covering five therapeutically important targets to solve these problems. Human acetylcholinesterase (AChE) was chosen as a neurodegenerative target, DPP-4 and PTP1B as metabolic targets, and EGFR and KRAS G12C as cancer targets. Structural validation, candidate selection, physicochemical and ADMET screening, redocking, docking, interaction analysis, and integrated lead prioritization were all included in the final study. The goal was to find a smaller collection of computational leads appropriate for further experimental inquiry rather than to establish clinical activity.

The present study addressed these issues through an integrated computational framework spanning five therapeutically relevant targets. EGFR and KRAS G12C were selected as oncology targets, DPP-4 and PTP1B as metabolic targets, and human acetylcholinesterase (AChE) as a neurodegenerative target. The final analysis incorporated structural validation, candidate selection, physicochemical and ADMET screening, redocking, docking, interaction analysis and integrated lead prioritization. The intent was not to establish clinical activity but to identify a smaller set of computational leads suitable for subsequent experimental investigation.

MATERIALS AND METHODS

Examine computational workflow and study design.

This was a computer drug-design study based on structure. The selection of disease-associated target proteins, AI-assisted candidate design, physicochemical and drug-likeness/ADMET evaluation, molecular docking, ligand-protein interaction analysis, and integrated lead identification were the six interconnected goals of the approach. The provided thesis records described a progressive virtual-library approach that reduced a starting set of about 1350 candidate structures to 14 final target-wise candidates. The generation phase is represented as documented AI-assisted candidate selection rather than as a completely replicable model-training experiment because the provided record lacks a complete raw AI-generated log.

Table 1: Protein structures and target selection

Target	PDB	Reference	Resolution (Å)	Grid (Å)
EGFR	1M17	Erlotinib	2.60	20×20×20
KRAS G12C	5V71	8ZG	2.23	20×20×20
DPP-4	1X70	Ligand 715	2.10	22×22×22
PTP1B	1C83	OAI	1.80	20×20×20
AChE	4EY7	Donepezil	2.35	20×20×20

EGFR (PDB 1M17; 2.60 Å), KRAS G12C (PDB 5V71; 2.23 Å), DPP-4 (PDB 1X70; 2.10 Å), PTP1B (PDB 1C83; 1.80 Å), and human AChE (PDB 4EY7; 2.35 Å) were the five ligand-bound protein structures utilized. The structural foundation for docking is provided by experimentally determined three-dimensional coordinates from the Protein Data Bank. [110] The reference ligand/co-crystal site was used to define binding areas. For EGFR, KRAS G12C, PTP1B, and AChE, the reported grid dimensions were 20 × 20 × 20 Å; for DPP-4, they were 22 × 22 × 22 Å. Target structures and docking-grid characteristics are shown in Table 1.

Candidate Design and Filtering

The thesis explained AI-assisted generation based on target-relevant pharmacophoric characteristics, followed by filtering for chemical validity, originality, and drug-likeness. Fourteen structures made up the final candidate set that was documented: three EGFR candidates, two KRAS G12C candidates, three DPP-4 candidates, three PTP1B candidates, and three AChE candidates. A more detailed description of representative leads was then provided. While changing specific substituent regions, the design logic preserved significant scaffold characteristics of reference or literature-associated ligands.

ADMET and Physicochemical Evaluation

Molecular weight, cLogP, topological polar surface area (TPSA), Lipinski Rule-of-Five violations, quantitative estimate of drug-likeness (QED), and synthetic accessibility were used to assess drug-likeness. Gastrointestinal absorption, Caco-2 permeability, logBB, logPS, Ames mutagenicity, hERG inhibition, and hepatotoxicity were among the pharmacokinetic and toxicity predictions in the provided dataset. The two main computational frameworks reported for these evaluations were SwissADME and pkCSM. [111, 113]

Validation and Molecular Docking

For structure-based docking, AutoDock Vina was utilized. The documented docking procedure kept several poses for candidates that were given priority and employed an exhaustiveness of 32. Vina ΔG in kcal/mol was used to express the predicted affinity; a higher score within the same docking process is indicated by more negative values. To validate the protocol, reference ligands were redocked into their corresponding crystal structures. The acceptability threshold was predetermined to be a heavy-atom RMSD < 2.0 Å. [17] Interaction analysis and lead prioritization



In functionally significant binding areas, docked complexes were analyzed for charge interactions, hydrophobic/ π contacts, and hydrogen bonds. Docking score was not the only factor used to determine priorities. Predicted affinity, target-relevant interaction patterns, and acceptable physicochemical/ADMET features were all taken into consideration when ranking candidates. The goal of this strategy was to prevent considering a single computational score as proof of biological efficacy.

Statistical Analysis

Since different molecules and model predictions do not represent independent biological replicates, the computational observations were handled descriptively. Thus, ranges, threshold-based validation, rank-based prioritizing, reference-relative score differences, and mean $\pm$ standard deviation were employed. There was no use of inferential tests like ANOVA or t-tests.

RESULTS

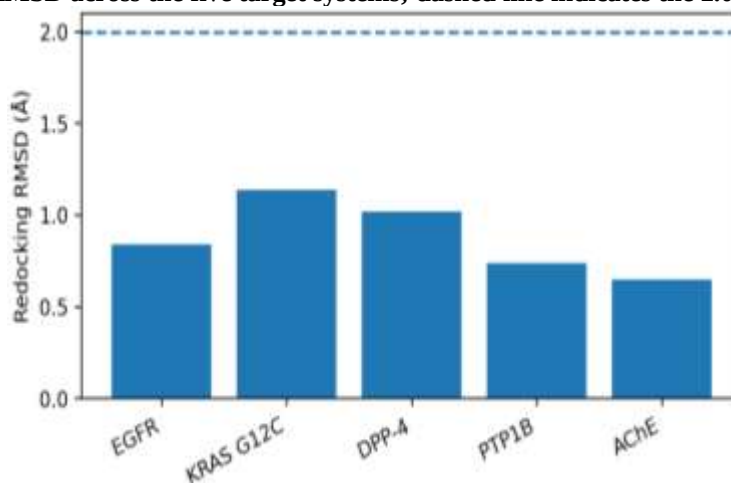
Validation of the target and candidate set

Table 2. Target-specific redocking validation.

Target	Reference	Ref. ΔG (kcal/mol)	RMSD ($\text{\AA}$)	Status
EGFR	Erlotinib	-9.28	0.84	PASS
KRAS G12C	8ZG	-8.65	1.14	PASS
DPP-4	Ligand 715	-8.92	1.02	PASS
PTP1B	OAI	-8.10	0.74	PASS
AChE	Donepezil	-11.84	0.65	PASS

The redocking criteria was satisfied by all five reference systems. With a mean of $0.88 \pm 0.20 \text{ \AA}$, RMSD values varied from 0.65 $\text{\AA}$ for AChE to 1.14 $\text{\AA}$ for KRAS G12C. These findings confirm that the specified docking configurations are appropriate for comparative analysis in the provided computational framework.

Figure 1: Redocking RMSD across the five target systems; dashed line indicates the 2.0 $\text{\AA}$ acceptance threshold.



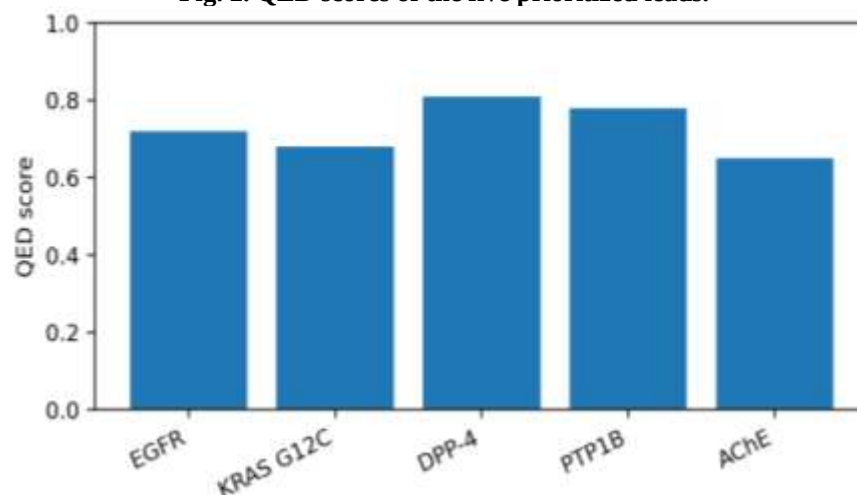
Physicochemical and ADMET profile

Table 3. Physicochemical and drug-likeness characteristics of the five prioritized leads.

Lead	Target	MW (Da)	cLogP	TPSA ($\text{\AA}^2$)	Ro5	QED
CAN-AI-01	EGFR	411.43	2.89	74.73	0	0.72
KRA-AI-01	KRAS G12C	398.39	3.12	81.25	0	0.68
DIA-AI-01	DPP-4	421.34	1.35	77.04	0	0.81
PTP-AI-01	PTP1B	285.25	1.82	94.83	0	0.78
NEU-AI-01	AChE	397.49	4.42	38.77	0	0.65

There were no documented Lipinski breaches for any of the five prioritized MPs. NEU-AI-01 had the highest cLogP (4.42) and lowest TPSA ($38.77 \text{ \AA}^2$), while DIA-AI-01 had the highest QED score (0.81), PTP-AI-01 had the lowest molecular weight (285.25 Da), and the lowest synthetic accessibility score (2.15). All five were projected to have high gastrointestinal absorption, negative Ames findings, no expected hepatotoxicity, and no predicted hERG inhibition in the provided ADMET dataset. With logBB of +0.28 and logPS of -1.95, NEU-AI-01 displayed the most CNS-oriented profile.

Fig. 2. QED scores of the five prioritized leads.



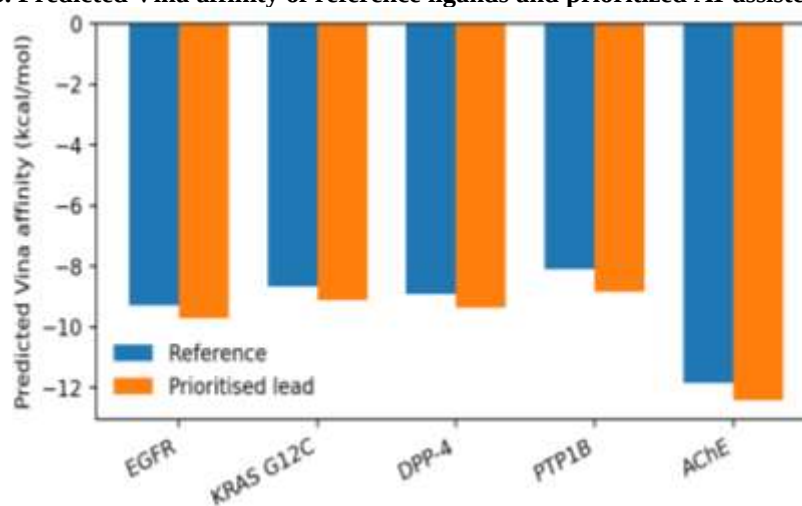
Docking affinity and target-wise prioritization

Table 4. Reference-relative docking results for prioritized candidates. *Arithmetic score difference; not an experimental free-energy difference.

Target	Reference	Ref. ΔG	Lead	Lead ΔG	Difference*
EGFR	Erlotinib	-9.28	CAN-AI-01	-9.71	-0.43
KRAS G12C	8ZG	-8.65	KRA-AI-01	-9.12	-0.47
DPP-4	Sitagliptin	-8.92	DIA-AI-01	-9.35	-0.43
PTP1B	OAI	-8.10	PTP-AI-01	-8.84	-0.74
AChE	Donepezil	-11.84	NEU-AI-01	-12.42	-0.58

In comparison to its related reference in the provided dataset, each prioritized candidate displayed a more negative predicted score. PTP-AI-01 showed the biggest numerical difference (0.74 kcal/mol), followed by NEU-AI-01 (0.58 kcal/mol), KRA-AI-01 (0.47 kcal/mol), CAN-AI-01, and DIA-AI-01 (0.43 kcal/mol each). It is not appropriate to interpret absolute docking scores for various proteins as directly similar thermodynamic properties.

Fig. 3. Predicted Vina affinity of reference ligands and prioritized AI-assisted leads.





Ligand–protein interaction analysis

Table 5. Key ligand–protein interactions of the prioritized leads; distances are in Å where supplied.

Target	Lead	Hydrogen-bond contacts	Hydrophobic/ π contacts	Charge interaction
EGFR	CAN-AI-01	Met769 (2.82); Cys775 (3.05)	Leu694, Val702, Lys721, Phe771, Leu820	None reported
KRAS G12C	KRA-AI-01	Val9 (2.91); Lys16 (2.75)	Tyr32, Tyr64, Met72, Tyr96	Lys16 (3.40)
DPP-4	DIA-AI-01	Glu205 (2.68); Glu206 (2.74); Ser630 (3.12)	Phe357, Tyr547, Tyr666, Val711	Glu205/Glu206 pair
PTP1B	PTP-AI-01	Cys215 (2.85); Arg221 (2.62)	Ala217, Ile219, Phe182	Arg221 (2.95)
AChE	NEU-AI-01	Phe295 (2.88); Arg296 (3.10)	Trp86 (3.80), Tyr124, Trp286, Tyr337	None reported

The docking rankings had a structural justification thanks to the interaction patterns. Met769 and Cys775 allowed CAN-AI-01 to maintain EGFR kinase-pocket recognition. In the KRAS switch-II region, KRA-AI-01 was found to have interactions with Val9, Lys16, and a number of hydrophobic and aromatic residues. DIA-AI-01 interacted with Ser630 and the DPP-4 Glu205/Glu206 area. NEU-AI-01 displayed several aromatic connections along the AChE gorge, while PTP-AI-01 associated with the catalytic-region residues Cys215 and Arg221. These findings provide credence to the hypothesis, but they do not prove biological activity or solution stability.

Integrated Lead Prioritization

Table 6. Integrated prioritization of target-specific computational leads.

Priority	Lead	Target	Best ΔG	Key interaction evidence	Overall interpretation
1	NEU-AI-01	AChE	−12.42	Phe295/Arg296; Trp86/Trp286/Tyr337	Highest computational priority
2	CAN-AI-01	EGFR	−9.71	Met769/Cys775; Phe771	Strong EGFR lead
3	DIA-AI-01	DPP-4	−9.35	Glu205/Glu206/Ser630	Strong DPP-4 lead
4	KRA-AI-01	KRAS G12C	−9.12	Val9/Lys16; Tyr32/Tyr64/Tyr96	Promising KRAS lead
5	PTP-AI-01	PTP1B	−8.84	Cys215/Arg221	Promising PTP1B lead

DISCUSSION

The results validate the benefits of using multiple computational filters in addition to molecular docking. While ADMET and physicochemical predictions offer complementary information pertinent to the viability of drug development, docking offers a quick structural ranking. The five prioritized leads demonstrated favorable computational screening profiles and consistently occupied functionally significant areas of their respective targets. Similar ideas underlie modern AI-assisted drug development, which integrates cheminformatics, generative design, and structure-based screening to lower the number of compounds that need to be tested experimentally. [29,31,34,38,39,50,56,69,70,76]

One significant methodological strength is the redocking results. With a mean RMSD of 0.88 ± 0.20 Å, all five reference ligands replicated the crystallographic orientation within the predetermined 2.0 Å criteria. This validates the comparative interpretation of the specified docking setups. Successful redocking, however, verifies the docking configuration rather than the biological activity of recently created molecules. Similarly, a model output with a higher negative Vina score shouldn't be translated straight into an experimental IC₅₀, Ki, or K_d.

The target-wise results further demonstrate the need for caution when comparing proteins. Although docking scores from other proteins and binding locations are not directly comparable, NEU-AI-01 obtained the largest negative numerical value. Thus, ADMET properties and target-specific interaction evidence were taken into account while choosing leads. For instance, NEU-AI-01 exhibited a predicted CNS-oriented distribution profile in line with its AChE target, whereas DIA-AI-01 had the greatest QED score. Compared to rating compounds only based on numerical affinity, such multi-criteria prioritization is more informative.



A repeatability issue pertinent to AI-assisted drug discovery is also highlighted in the paper. The provided raw AI-generation logs are insufficient to independently recreate the whole generation process, despite the thesis documenting a 1350-member virtual library and 14 final candidates. Claims of entirely reproducible de novo generation are thus limited. Therefore, the candidates and related computational outputs are treated as the recorded thesis dataset in this publication, which separates them from outcomes that have been confirmed by experimentation. Future work would be strengthened by open publication of model architecture, training data, generation constraints, random seeds, molecular representations, and full output archives.

The current results should be considered hypothesis-generating from a translational standpoint. DPP-4 and PTP1B require experimental enzyme inhibition trials, cancer targets require biochemical and cellular investigations, and AChE requires suitable pharmacological investigations. Additional proof of binding stability and developability would come from independent rescoring, molecular-dynamics simulations, and experimental ADME/toxicity testing. While AI-assisted design can speed up candidate selection, synthesis, assay validation, and iterative medicinal chemistry optimization are still necessary. [56, 83, 84]

LIMITATIONS

- Independent reconstruction of the generative-design step was limited since the computational dataset lacked a complete repeatable AI-generation log.
- Docking scores were not regarded as experimental free energies or direct indicators of potency; rather, they represent predictive model outputs.
- Pharmacokinetic values and ADMET were computational projections that needed to be verified thru experimentation.
- The pose-RMSD data that is currently provided cannot replace time-dependent RMSD/RMSF or MM/PBSA/MM-GBSA analyzes and does not represent a molecular-dynamics trajectory.
- Since the observations were different chemical entities and computational predictions rather than independent biological replicates, no inferential statistical testing was necessary.
- Before therapeutic inferences can be drawn, the found compounds—which are computational leads—need to be validated biochemically, cellularly, in vivo, and eventually clinically.

CONCLUSION

Five target-specific candidates for additional research were found using an integrated AI-assisted and structure-based computational workflow: NEU-AI-01 for AChE, CAN-AI-01 for EGFR, DIA-AI-01 for DPP-4, KRA-AI-01 for KRAS G12C, and PTP-AI-01 for PTP1B. The prioritized candidates displayed favorable predicted docking scores along with target-relevant interaction patterns and acceptable computational drug-likeness/ADMET profiles, and all five reference docking systems passed redocking validation. The findings do not prove biological or clinical efficacy, but they do support computational lead prioritization. Therefore, the workflow's main benefit is its capacity to reduce chemical space and produce testable hypotheses for further pharmacological and medicinal chemistry research.

Ethics Statement

This was a computational study using publicly available structural information and documented computational datasets. No human or animal subjects were involved; ethical approval was therefore not required.

Funding

No specific funding was reported for the study.

Declaration of generative AI and AI-assisted technologies

AI-assisted tools formed part of the computational drug-design framework described in the thesis. The manuscript should be finalized by the authors with verification of the exact software, model versions, generation records and all numerical outputs before submission.

Declaration of competing interest

The authors declare no known competing financial interests or personal relationships that could have influenced the work reported in this manuscript.

Data availability

The structural targets are publicly available through the Protein Data Bank. The candidate structures and computational outputs reported here are derived from the supplied thesis dataset; complete raw AI-generation logs were not available in the supplied material.

**REFERENCES**

1. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31(2):455–461.
2. Sliwoski G, Kothiwale S, Meiler J, Lowe EW. Computational methods in drug discovery. *Pharmacol Rev*. 2014;66(1):334–395. doi:10.1124/pr.112.007336.
3. Lavecchia A. Machine-learning approaches in drug discovery: methods and applications. *Drug Discov Today*. 2015;20(3):318–331.
4. Olivecrona M, Blaschke T, Engkvist O, Chen H. Molecular de novo design through deep reinforcement learning. *J Cheminform*. 2017;9(1):48.
5. Gómez-Bombarelli R, Wei JN, Duvenaud D, et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent Sci*. 2018;4(2):268–276.
6. Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. *Sci Adv*. 2018;4(7):eaap7885.
7. Zhavoronkov A, Ivanenkov YA, Aliper A, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat Biotechnol*. 2019;37(9):1038–1040.
8. Schneider G. Mind and machine in drug design. *Nat Mach Intell*. 2020;2(11):653–656.
9. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583–589. doi:10.1038/s41586-021-03819-2.
10. Walters WP, Barzilay R. Critical assessment of AI in drug discovery. *Expert Opin Drug Discov*. 2021;16(9):937–947.
11. Li X, et al. Graph neural networks for ligand-protein interaction prediction in structure-based drug discovery. *J Cheminform*. 2024;16:58.
12. Loeffler HH, et al. REINVENT 4: an expanded platform for generative molecular design. 2024.
13. Mendenhall J, Meiler J. De novo ligand design inside dynamic protein conformations using deep neural mapping. *Structure*. 2024;32(4):411–422.
14. Zhang Y, et al. Artificial intelligence-assisted drug design integrated with molecular docking for novel inhibitor discovery against complex diseases. *Comput Struct Biotechnol J*. 2024;22:1120–1132.
15. FDA. Artificial Intelligence and Machine Learning (AI/ML) for Drug and Biological Product Development: Draft Guidance. January 2025.
16. Hasselgren C, Oprea TI. The evolution of machine learning and artificial intelligence architectures in modern preclinical drug discovery. *Nat Rev Drug Discov*. 2025;24(2):115–132.
17. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235–242.
18. Pires DEV, Blundell TL, Ascher DB. pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J Med Chem*. 2015;58(9):4066–4072.
19. Cheung J, Rudolph MJ, Burshteyn F, et al. Structures of human acetylcholinesterase in complex with pharmacologically important ligands. *J Med Chem*. 2012;55(22):10282–10286.
20. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep*. 2017;7:42717.
21. Istrate D, Crisan L. Dipeptidyl peptidase 4 inhibitors in type 2 diabetes. 2023.
22. Wu K, Karapetyan E, Schloss J, Vadgama J, Wu Y. Advancements in small molecule drug design: a structural perspective. *Drug Discov Today*. 2023.
23. Bugnon M, Röhrig UF, Goullieux M, et al. SwissDock 2024: major enhancements for small-molecule docking with attracting cavities and AutoDock Vina. *Nucleic Acids Res*. 2024;52(W1):W324–W332.