

Integrating Computer-Aided and AI-Driven Drug Design into Medicinal Chemistry Education: A Computational Course on EGFR L858R/V948R-Targeted Oral Anti-Lung Cancer Agents

Qingkun Li^{1†}, Ayixian Wusiman^{1†}, Nan Wang^{1†}, Zhenming Liu^{1,2*}, Shengwen Yang^{1*}

¹Key Laboratory of Xinjiang Endemic Phytomedicine Resources, Ministry of Education; School of Pharmacy, Shihezi University, Shihezi 832002, Xinjiang, China

²State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, Beijing 100191, China

[†]*These authors contributed equally to the work.*

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: The traditional medicinal chemistry education suffers from the high barrier of equipment and the teaching of modern computational methods. To solve this, we designed an AI-assisted computational drug design course, “Design of Oral Anti-Lung Cancer Drugs Targeting EGFR L858R/V948R,” for pharmacy undergraduates. The course focuses mainly on a resistant mutation in NSCLC and employs a structured workflow with several software tools such as Schrödinger, PyMOL, MolSnapper, Pipeline Pilot and GROMACS. Through 16 class hours, students can complete the closed-loop learning experience from target selection to candidate compound confirmation. In summary, this course is helpful for technical barriers, encourages students to form an interdisciplinary thinking consciousness, and provides a transferable model for the modernization of hands-on medicinal chemistry education.

Keywords: Medicinal chemistry education; CADD; AIDD; EGFR L858R/V948R; Virtual screening; Molecular generation; Molecular dynamics simulations

Online publication: April 26, 2026

1. Introduction

Medicinal chemistry is an interdisciplinary science at the interface of chemical biology, pharmacology, and medicine, playing a crucial role in drug discovery^[1]. Consequently, its educational content and methods must evolve with the field. However, two problems continue to plague undergraduate instruction: (1) limited students are available to costly, sophisticated laboratory instrumentation^[2] and (2) teaching methods and content that are often lagging behind the field, rely on passive instruction rather than fostering critical research skills or incorporating modern computational approaches^[3].

To tackle these challenges, we have created a computational and AI-assisted drug design laboratory course titled

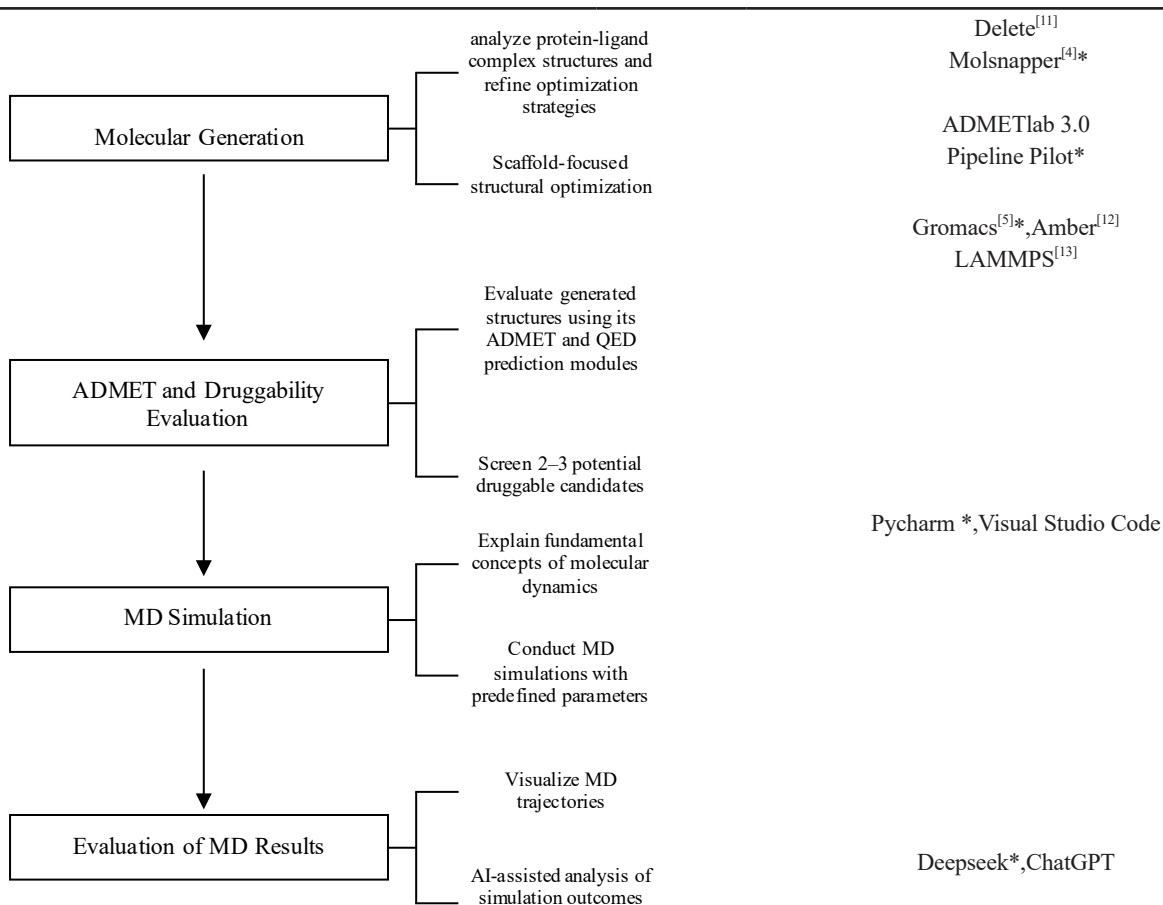
“Design of Oral Anti-Lung Cancer Drugs Targeting EGFR L858R/V948R”. Focused on a key resistant mutation in EGFR(L858R/V948R), the course integrates multidisciplinary knowledge of computational and medicinal chemistry, and covers the whole process of drug development, including virtual screening, SAR analysis and molecular optimization. In the course, students will greatly boost their interest in modern medicinal chemistry in the practice of traditional drug design. traditional drug design practice.

2. Course overview

For the purpose of systematically guiding pharmacy undergraduates to learn the logic of structure-based drug design, the study has developed the laboratory course “Design of Oral Anti-Lung Cancer Drugs Targeting EGFR L858R/V948R”. The course is organized into 16 class periods in 8 coherent modules focused on practical application rather than theoretical details. Thus, students can run a full drug design pipeline through learning the use of tools such as MolSnapper^[4], Pipeline Pilot, and Gromacs^[5], acquiring a working knowledge of CADD/AIDD principles, including virtual screening, molecular generation, ADMET evaluation, and molecular dynamics validation.

Table 1. Teaching workflow, CADD, and AIDD. Tools marked with an asterisk (*) represent those emphasized in teaching case studies.

Teaching segments and contents	Tools for Computational Drug Design	pedagogy
Theoretical Knowledge Instruction The background of CADD and AIDD The essential workflow of drug design Key concepts of drug design		
Data Acquisition Retrieve natural product and approved drug ligand libraries Obtain the PDB structure of EGFR L858R/V948R		MCE Database * RCSB Protein Data Bank AlphaFold3^[6], UniProtKB
Virtual Screening The use of Maestro module of Schrödinger software Screen 2–3 candidate drug molecules		AutoDock Vina^[7] Schrödinger maestro *
Lead compound identification Interpret key research papers on 8A2D Analyze binding modes between selected molecules and		Discovery Studio^[8] UCSF ChimeraX^[9] PLIP^[10], PyMOL *



3. Results and discussion

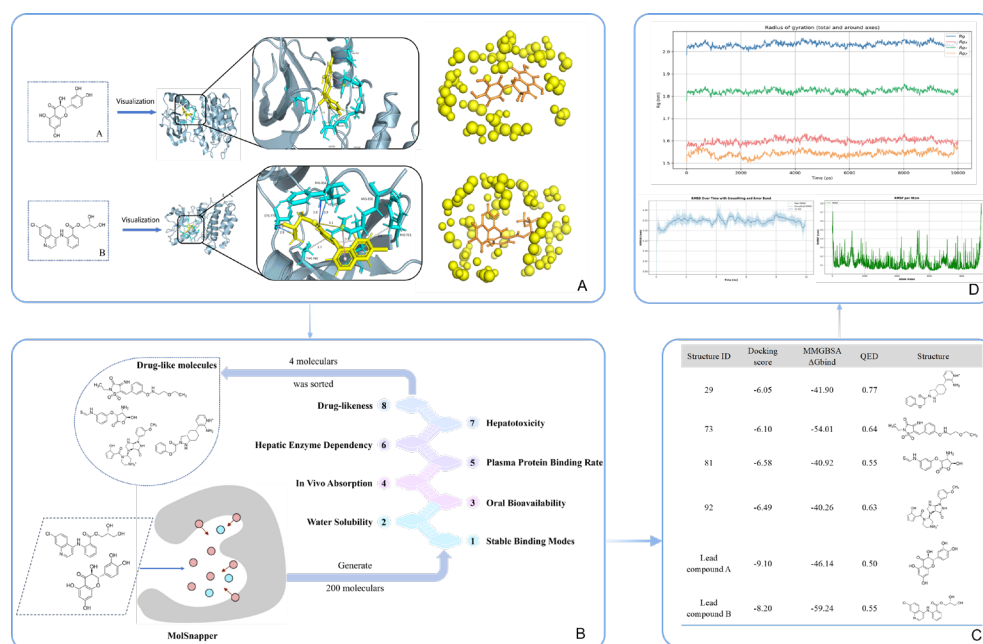


Figure 1. The teaching demonstration of the integrated computational drug design workflow targeting EGFR L858R/V948R. (A) Molecular Docking and Pharmacophore Models. (B) Molecular Generation Workflow. (C) Workflow of Evaluation Framework for Drug-like Candidates (D) Molecular Dynamics Simulation. (RMSD, RMSF, and Rg analysis).

- (1) Theoretical Knowledge Instruction. At the beginning of the course, two class periods of theoretical instruction based on three key textbooks ^[14,15] introduce CADD/AIDD fundamentals and the drug design pipeline to provide students with essential knowledge for subsequent practical work.
- (2) Virtual Screening. Firstly, students download the 3D protein structure of EGFR L858R/V948R (PDB ID: 8A2D) from online databases such as RCSB PDB, AlphaFold3 ^[6], and UniProt. They prepare a compound library from the MedChemExpress (MCE) collection, using a two-pronged approach of screening both the FDA-approved drug library and the Natural Product Library. All virtual screening was performed in Schrödinger Maestro. Protein preparation and binding site definition were performed using the Protein Preparation Workflow and Receptor Grid Generation, respectively. The Virtual Screening Workflow subsequently performs a stepwise filtering approach (HTVS, SP, XP) to dock the compound library and to identify 119 final ligands. The main evaluation is performed with GlideScore / XP Score and MMGBSA dG Bind. Lower scores (e.g., -11.928 and -73.46) indicate stronger predicted binding. For proteins with unknown ligands, the SiteMap tool can be used as an alternative to identify binding sites and generate grids.
- (3) Lead Compound Identification. After obtaining a lead compound, the course introduces AI-driven structural optimization using the MolSnapper^[4] model, a structure-based molecular generation (SBMG) ^[16] diffusion model that incorporates pharmacophore constraints. The pharmacophore models are generated from the protein interactions of the virtual screening hits Ligand-A and Ligand-B by using the “develop pharmacophore model” module in Schrödinger Maestro. Next, students perform ligand-based generation with MolSnapper, adjusting parameters in the sample_MolDiff. yml file to control pharmacophore matching and steric avoidance. This exercise generates 100 molecules per input ligand and results in 200 novel candidates. Preliminary evaluation shows that these molecules have an average QED > 0.40, a synthetic accessibility (SA) score > 0.60, a pharmacophore matching > 0.62, and an interaction evaluation > 4.40, indicating improved drug-likeness, synthetic feasibility, and predicted binding.
- (4) ADMET and Druggability Evaluation. To refine the 200 AI-generated compounds, a multi-dimensional evaluation was conducted. First, XP precision docking in Schrödinger’s Virtual Screening Workflow against the preprocessed EGFR target filtered candidates to 56 with docking scores < -6.0, suggestive of good affinity. Then, the Pipeline Pilot platform was used for the drug-likeness (QED) and ADMET prediction using a customized workflow, focusing on important properties for oral drugs such as bioavailability and hepatotoxicity. The QED values of the parent leads A and B (0.50476 and 0.55204, respectively) were used as a benchmark for comparison. Only four compounds satisfy all the criteria of improved QED and acceptable ADMET ranges. Notably, candidates 1 and 2 obtained QED > 0.65. This process serves as a demonstration of a modern drug design workflow, teaching students how to synergistically optimize for affinity, drug-likeness, and pharmacokinetics to discover promising candidates for further study.
- (5) Evaluation of MD Results. Following static filtering, molecular dynamics (MD) simulations in GROMACS ^[5] were introduced to assess the binding stability of the four candidate complexes dynamically. After preprocessing the trajectories (removing periodic boundary effects and aligning the structure), the key parameters, RMSD, RMSF, and Rg, were calculated. The second protein-ligand complex showed the most favorable dynamic profile with RMSD equilibrating stably between 0.10-0.25 nm with a narrow error band indicating little conformational drift. RMSF analysis confirmed the rigidity of key active-site residues (≤ 0.3 nm) and increased flexibility only in a distal loop region. Furthermore, the total radius of gyration (Rg) remained constant (1.7–1.8 nm), and all the spatial components showed isotropic low-amplitude fluctuations. This multi-parameter analysis quantitatively confirmed the structural compactness of the complex and stability of the binding sites to finalize the integrated CADD/AIDD evaluation workflow.

4. Conclusion

In order to close the gap between traditional medicinal chemistry education and the current industrial practice, we designed

a teaching framework based on the protein “8a2d”, integrating the tools Schrödinger, PyMOL, MolSnapper, Pipeline Pilot and GROMACS in a closed-loop workflow: from data acquisition and virtual screening to complex visualization, molecular generation, property evaluation and dynamic validation. “A low threshold practice with cutting-edge technology integration” enables students to run a complete pipeline from target to potential drug molecule. The hands-on, visual approach demystifies the advanced CADD/AIDD tools and stimulates interest through an interactive design-feedback cycle, providing a replicable model for practical drug design instruction.

Funding

Ministry of Education Industry-Academia Cooperation Collaborative Education Project (Project No.: 2508153613); Shihezi University Education and Teaching Reform Research Project (Project No.: JGY-2025-55)

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Gioiello A, Piccinno A, Lozza A, et al., 2020, The Medicinal Chemistry in the Era of Machines and Automation: Recent Advances in Continuous Flow Technology. *J. Med. Chem.*, 63: 6624–6647.
- [2] You Q, Yao W, Xi X, et al., 2019, Problems and Countermeasures for Innovative Pharmaceutical Talents Training. *Strategic Study of CAE*, 21: 79–83.
- [3] Wilson Fetrow M, Svihla V, Burnside B, et al., 2023, Course Based Undergraduate Research Experiences in a Chemical Engineering Laboratory Promote Consequential Agency. *J. Chem. Educ.*, 100: 3752–3763.
- [4] Ziv Y, Imrie F, Marsden B, et al., 2025, MolSnapper: Conditioning Diffusion for Structure Based Drug Design. *J. Chem. Inf. Model.*, 65: 4263–4273.
- [5] Hess B, Kutzner C, van der Spoel D, et al., 2008, GROMACS 4: Algorithms for Highly Efficient, Load Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.*, 4: 435–447.
- [6] Abramson J, Adler J, Dunger J, et al., 2024, Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature*, 630: 493–500.
- [7] Eberhardt J, Santos Martins D, Tillack A, et al., 2021, AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.*, 61: 3891–3898.
- [8] Dassault Systèmes BIOVIA, 2019, Discovery Studio Modeling Environment, release 2019, Dassault Systèmes, France.
- [9] Meng E, Goddard T, Pettersen E, et al., 2023, UCSF ChimeraX: Tools for Structure Building and Analysis. *Protein Sci.*, 32: e4792.
- [10] Schake P, Bolz S, Linnemann K, et al., 2025, PLIP 2025: Introducing Protein–Protein Interactions to the Protein–Ligand Interaction Profiler. *Nucleic Acids Res.*, 53: W463–W465.
- [11] Chen S, Zhang O, Jiang C, et al., 2025, Deep Lead Optimization Enveloped in Protein Pocket and Its Application in Designing Potent and Selective Ligands Targeting LTK Protein. *Nat. Mach. Intell.*, 7: 1–15.
- [12] Case DA, Aktulga HM, Belfon K, et al., 2023, AmberTools. *J. Chem. Inf. Model.*, 63: 6183–6191.
- [13] Thompson AP, Aktulga HM, Berger R, et al., 2022, LAMMPS – A Flexible Simulation Tool for Particle Based Materials Modeling at the Atomic, Meso, and Continuum Scales. *Comput. Phys. Commun.*, 271: 108171.
- [14] Li H, Zheng M, 2021, Artificial Intelligence and Drug Design. Science Press, Beijing.
- [15] Hu J, et al., 2020, Computer Aided Drug Design: Theory and Applications. Chemical Industry Press, Shanghai.

- [16] Schneuing A, Harris C, Du Y, et al., 2024, Structure Based Drug Design with Equivariant Diffusion Models. Nat. Comput. Sci., 4: 899–909.

Publisher's note

Whioce Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.