INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES
[ISSN: 0975-4725; CODEN(USA): IJPS00]
Journal Homepage: https://www.ijpsjournal.com

OPEN ACCESS

Research Article

A Validation-First Workflow for Structure-Based Drug Design

Varade Omkar*, Navale Komal, Cholke Pratiksha

Pravara Rural Education societys Institute of Pharmacy loni.

ARTICLE INFO

Published: 2 Nov 2026
Keywords:
structure-based drug design, computer-aided drug design, molecular docking, virtual screening, ADMET, molecular dynamics, artificial intelligence, reproducibility
DOI:
10.5281/zenodo.23876741

ABSTRACT

Structure-based drug design (SBDD) uses three-dimensional information about a biological target to guide the identification and optimisation of small molecules. Molecular docking, virtual screening, molecular dynamics, free-energy calculations, and artificial-intelligence methods can help prioritise compounds for experimental testing. However, a plausible docking pose or favourable score is not proof of binding or therapeutic activity. Outcomes can be affected by protein structure quality, protonation, ligand preparation, receptor flexibility, scoring-function limitations, and the chemical space represented in a screening library. This narrative review presents a validation-first workflow: define a biologically justified target; assess the target structure and binding site; prepare proteins and ligands transparently; validate docking using reference ligands and appropriate controls; combine structural interpretation with complementary ranking; assess absorption, distribution, metabolism, excretion and toxicity (ADMET) risks; and confirm prioritised candidates experimentally. Recent AI-enabled methods are considered as tools for hypothesis generation rather than substitutes for medicinal chemistry judgement or laboratory evidence. The review also discusses reproducibility, uncertainty, and transparent reporting. The central recommendation is to connect every computational stage to a clear decision gate and to distinguish predictions from observations. This framework is intended to support student and early-stage research projects; it is not a report of new experimental data or a claim that any compound has been discovered.

INTRODUCTION

Drug discovery combines biological reasoning, chemical design, experimental testing, and progressively more demanding evidence. Many compounds fail because of insufficient potency, selectivity, exposure, safety, or developability. Computer-aided drug design (CADD) supports early decisions by representing molecules and biological targets computationally. SBDD is particularly useful when a credible three-dimensional structure of the target or binding site is available [1,2].

***Corresponding Author:** Varade Omkar
Address: Pravara Rural Education societys Institute of Pharmacy loni
Email : omkarvarade2005@gmail.com
Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

SBDD is not one algorithm. It can involve target assessment, structural biology, binding-site analysis, virtual screening, docking, fragment-based design, ligand optimisation, molecular dynamics, and free-energy estimation. Structural information can suggest how a ligand might occupy a pocket and which chemical changes could improve interactions. Nevertheless, a predicted pose remains a hypothesis. The key question is not simply which molecule has the best score, but which candidate is most defensible to test given the available evidence and uncertainty.

AI methods have expanded access to predicted protein structures and candidate generation. These methods may broaden chemical exploration, but their output can be affected by training-data bias, uncertain pocket geometry, protein flexibility, and limited prospective validation [3–5]. Computational complexity does not guarantee accuracy. Progress depends on suitable controls, reliable data, physical reasoning, and experimental confirmation.

This review therefore focuses on a practical, validation-first workflow for pharmacy and medicinal-chemistry projects. It connects common computational techniques to the decisions they support, identifies checks needed before results are interpreted, and outlines the information required for reproducibility.

2. Review approach and scope

This is a narrative review and workflow synthesis, not a systematic review or meta-analysis. Sources include methodological papers, peer-reviewed reviews, and established structural-biology resources on docking, structure prediction, AI-based design, and reproducibility. No original docking study, biological assay, or quantitative meta-analysis was performed. A formal systematic review would require documented database-specific search strings, eligibility criteria, duplicate screening, and a reproducible study-selection process.

3. Target selection and structure quality

A project should begin with a biological question. Target selection should consider disease relevance, evidence that modulation may change a meaningful phenotype, binding-site tractability, selectivity risks, and the availability of a suitable assay. A high-quality structure cannot compensate for a weak biological rationale.

The Protein Data Bank (PDB) provides experimentally determined macromolecular structures and associated metadata [6]. Before selecting a structure, examine resolution, missing residues, alternate conformations, mutations, bound ligands, metal ions, cofactors, crystallographic waters, and biological assembly. If several structures exist, comparing them can reduce dependence on one conformation.

3. Target selection and structure quality (continued)

Predicted structures can help when experimental structures are unavailable. However, global fold confidence does not guarantee that a binding pocket, flexible loop, or side-chain orientation is accurate. AlphaFold Database models and their confidence information can support hypothesis generation, but high-impact conclusions should be checked against experimental structures, homologous complexes, mutagenesis, or other independent evidence when feasible [7,8].

Binding-site definition is a consequential decision. Sites may be identified from a co-crystallised ligand, known active-site residues, conserved motifs, cavity-detection methods, or published evidence. The chosen region should be biologically plausible and appropriately sized. Allosteric, membrane-associated, metal-dependent, and conformationally dynamic targets may require multiple structures or specialised treatment rather than a generic rigid-pocket approach.

4. Protein and ligand preparation

Preparation converts structural files into representations accepted by docking or simulation software. Typical steps include handling alternate locations, repairing appropriate missing atoms, adding hydrogens, assigning protonation states and charges, and deciding whether to retain waters, cofactors, or metals. These choices can alter electrostatics and steric fit. Incorrect histidine protonation or omission of a catalytic metal, for example, can change predicted interactions. Automated preparation is useful but should be inspected.

Ligand preparation should account for stereochemistry, ionisation, tautomerism, conformer generation, and

geometry. Incorrect stereochemistry or an unsuitable protonation state can make the result misleading. Check for duplicate structures, salts, reactive groups, and possible assay-interfering features where relevant. Record software names, versions, parameters, and manual changes so the workflow can be reproduced.

5. Molecular docking and virtual screening

Docking searches for plausible ligand orientations and conformations within a target site. Most methods combine pose sampling with a scoring function that approximates steric complementarity, hydrogen bonding, hydrophobic interactions, electrostatics, and ligand strain. Docking is efficient for ranking large libraries, but scores are not direct measurements of binding free energy and may not be comparable across unrelated targets or protocols [1,2,9].

A staged screen is preferable. First, check chemical quality and remove duplicates. Second, define the receptor and docking region. Third, inspect top poses for plausible geometry and clashes. Fourth, use complementary evidence such as known actives, inactive controls, pharmacophore compatibility, alternative scoring, or consensus ranking. Ranking by the most negative score alone can favour large compounds or artefacts and generate false positives.

A benchmark should match the intended use. Known actives can test pose recovery or enrichment, while suitable inactive compounds or decoys can test discrimination. Retrospective benchmarks may be biased by close analogues appearing in both training and test sets or by unrealistic decoys. Prospective testing of previously untested compounds provides stronger evidence of practical value than redocking alone.

6. Docking validation and pose interpretation

Validation should be planned before screening. When a co-crystallised ligand is available, redocking can test whether the protocol approximately reproduces the observed pose. Pose similarity may be measured using root-mean-square deviation (RMSD), but interpretation depends on atom mapping, ligand symmetry, and the intended application. Successful redocking does not prove that the scoring function will rank new molecules accurately; it tests only part of the workflow.

Further checks include cross-docking into alternative structures, screening known actives and inactives, testing sensitivity to receptor conformation, and comparing results with established structure–activity relationships. A credible pose should have sensible geometry: hydrogen bonds should be appropriately oriented, hydrophobic groups should occupy compatible regions, and buried charges should have a plausible environment. Severe clashes or dependence on poorly supported residues should reduce confidence.

Two-dimensional interaction diagrams are useful summaries, not proof. A displayed hydrogen bond may be weak or poorly oriented, and a predicted salt bridge may depend on protonation. Visualisation should support, not replace, chemical reasoning and independent evidence.

7. Molecular dynamics and binding-energy estimates

Molecular dynamics (MD) simulates atomic movement under a selected force field and can help investigate pose stability, flexible loops, water networks, and alternative binding modes. Yet a stable-looking trajectory does not establish biological activity. Results depend on the initial pose, force field, protonation, solvent model, sampling, and simulation duration. Replicates and convergence checks can improve confidence but require additional resources.

End-point methods such as MM/GBSA or MM/PBSA are sometimes used to compare related ligands. Alchemical free-energy methods can estimate relative affinity changes in carefully designed series, but they require appropriate sampling, validated protocols, and uncertainty analysis. Such methods are best used to answer focused questions rather than as universal replacements for experimental measurements [3,5].

Method choice should match the decision. Docking may be sufficient to nominate diverse candidates in an early screen. MD or free-energy calculations may add value during lead optimisation when the system is well characterised. A more computationally expensive method is not automatically more accurate.

8. ADMET and developability assessment

Potency is only one component of drug quality. ADMET assessment asks whether a candidate may be absorbed, reach the relevant tissue, remain stable, avoid problematic metabolism, and present acceptable toxicity risks. Early computational triage can estimate physicochemical properties, solubility, permeability, lipophilicity, metabolic liabilities, transporter effects, and selected toxicity endpoints. Predictions are model outputs, not clinical facts, and reliability varies by chemical class and training domain.

8. ADMET and developability assessment (continued)

A balanced selection process avoids maximising docking score while ignoring exposure and safety. Excessive lipophilicity can increase nonspecific binding or metabolic burden; poor solubility can limit exposure; and high polarity may restrict passive permeability. These are trends, not universal rules. Predicted CYP inhibition, hERG liability, hepatotoxicity, or other alerts should prompt contextual review and follow-up rather than automatic rejection without considering the intended use.

ADMET tools should be used within their applicability domains. A model trained on limited chemical diversity may be unreliable for new scaffolds, unusual elements, or rare mechanisms. Reports should name the model and version, define the endpoint, describe any confidence estimate, and state the threshold used. Where feasible, compare multiple methods and experimentally confirm decisions with important safety or development implications.

9. Artificial intelligence in structure-aware design

AI approaches increasingly support protein-structure prediction, interaction modelling, virtual screening, and de novo molecular design. Generative models can propose structures conditioned on a binding pocket, and learned scoring functions may complement conventional energy-based methods [4,5,10]. These approaches can broaden chemical exploration when combined with medicinal-chemistry constraints such as synthetic accessibility, novelty, selectivity, and suitable physicochemical properties.

Plausible-looking molecules are not necessarily synthesizable, stable, selective, or active. Models can exploit biases in training data or optimise surrogate metrics that do not translate into experimental success. Predictions may be less reliable for novel chemistry, unusual binding sites, or flexible targets. AI should therefore be embedded in an iterative loop: generate candidates, check chemical validity and developability, assess structural plausibility, synthesise selected compounds, measure activity, and use reliable results to guide the next design cycle.

For reproducibility, report model names and versions, conditioning information, constraints, sampling settings, filters, and selection criteria. A generated structure is not automatically biologically novel or patentable; chemical similarity and prior-art searches remain necessary. AI-assisted writing or analysis should also be disclosed when required by the institution or journal.

10. Proposed validation-first workflow

Each computational stage should have a decision gate. Candidates should advance only when they meet the stage's purpose and relevant uncertainties have been documented.

Stage	Task	Decision checkpoint
1.Target	Biological rationale and assay plan	Target relevance and measurable endpoint are supported
2. Structure	Assess experimental/predicted model and pocket	Structure is fit for purpose; uncertainty recorded
3. Preparation	Protein, waters, cofactors, protonation, ligands	Choices documented and inspected
4. Validation	Redocking, controls, sensitivity checks	Protocol passes relevant checks; failures reported
5. Ranking	Docking, pose review, orthogonal evidence	Chemically plausible candidates prioritised

Stage	Task	Decision checkpoint
6. ADMET	Properties, liabilities, model-domain checks	Risks balanced against intended use
7. Experiment	Binding/functional assay and counter-screens	Hypothesis tested with suitable controls
8. Reporting	Record results, SAR, files, negative data	Methods allow independent interpretation

11. Reproducibility and common failure modes

Reproducibility is central to computational science. Small differences in receptor preparation, protonation, grid placement, ligand conformers, or software defaults can alter rankings. At minimum, report the target and structure identifier; structure source and access date; chain and residue selection; handling of missing atoms, waters, cofactors and metals; protonation assumptions; ligand source and stereochemistry; software versions; scoring and search settings; random seeds where relevant; validation protocol; and selection criteria.

Reporting only top-ranked compounds hides the size and quality of the library, preparation failures, selection thresholds, and the influence of manual inspection. A stronger report gives the number of compounds entering and leaving each filter and includes representative failures. Negative results can reveal that a structure is unsuitable, a protocol cannot distinguish known actives, or a property filter removes most candidates.

Common errors include treating a docking score as measured affinity, treating predicted ADMET as proof of safety, using a predicted structure without evaluating local confidence, changing thresholds after seeing results, and presenting one computational output as causal evidence. Reduce these risks with controls, sensitivity checks, pre-defined decision rules where feasible, and explicit separation of observations from predictions.

12. Experimental confirmation and translation

Computational predictions need an experimental plan. Depending on the target, follow-up may include biochemical inhibition, binding assays, cellular functional assays, target-engagement studies, selectivity panels, and counterscreens for assay interference. Use positive and negative controls, replicates, appropriate concentration ranges, and transparent analysis. In-vitro binding may not translate to cells because of permeability, efflux, instability, protein binding, or pathway compensation. A cellular effect may also arise from an off-target mechanism; complementary assays help distinguish these possibilities.

Medicinal chemistry uses experimental results to update structure–activity relationships (SAR). The most informative next compound is not always the one predicted to be strongest. A matched molecular pair may test a specific interaction; another analogue may improve solubility or selectivity; a control may probe mechanism. This experiment-driven cycle prevents computational optimisation from becoming detached from biological evidence.

13. Limitations and future directions

This workflow is a narrative synthesis, not a formally validated standard operating procedure. Adaptation is required for different target classes, therapeutic areas, modalities, and project budgets. Docking and ADMET models may be biased by limited datasets; MD may not sample relevant states; free-energy methods can be costly; and AI predictions may not generalise to novel chemistry. Experimental assays also have uncertainty and may disagree across laboratories.

Future progress is likely to come from integration rather than one universally superior algorithm. Useful directions include better structural coverage, explicit treatment of protein ensembles and water, uncertainty-aware ranking, prospective benchmarking, synthesis-aware design, and reporting that captures failed as well as successful predictions. Structure-aware AI may accelerate hypothesis generation, but prospective experimental evidence remains essential [3–5,10].

14. Practical recommendations for student CADD projects

A student project can be scientifically useful without claiming to discover a drug. Choose a target with a known structure and reference ligands, state a narrow research question, and use a manageable library.

Document the docking protocol, validate it with a reference ligand where possible, inspect poses, and apply ADMET predictions as preliminary triage. Unless biological experiments have been performed, describe results as computational prioritisation rather than confirmed activity.

A results table should distinguish observations from predictions. For each candidate, report the compound identifier, docking score and software-defined units, predicted interactions, physicochemical properties, ADMET alerts, and a brief uncertainty note. Avoid calling the best-scoring molecule the "best drug" or claiming that a predicted interaction proves inhibition. Prefer wording such as "prioritised for experimental evaluation."

State limitations explicitly: the selected structure may not represent all target conformations; docking scores approximate interactions; prediction tools have applicability limits; and biological activity is not established without an appropriate assay. Clear limitations show that conclusions are calibrated to evidence.

CONCLUSION

SBDD is most useful when treated as a sequence of testable hypotheses rather than a scoring contest. Structural quality, preparation choices, docking validation, complementary ranking, ADMET triage, and experimental confirmation provide different types of evidence. AI and advanced simulations can broaden the search and refine decisions, but they do not eliminate uncertainty or replace laboratory testing. A validation-first workflow makes uncertainty visible and connects computational effort to decisions that can be checked. Transparent methods, suitable controls, careful interpretation, and accurate citation are essential for scientifically useful results.

Abbreviations

ADMET, absorption, distribution, metabolism, excretion and toxicity; AI, artificial intelligence; CADD, computer-aided drug design; MD, molecular dynamics; PDB, Protein Data Bank; QSAR, quantitative structure–activity relationship; SBDD, structure-based drug design; SAR, structure–activity relationship.

REFERENCES

1. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery*. 2004;3:935–949. doi:10.1038/nrd1549.
2. Ferreira LG, dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. *Molecules*. 2015;20 (7):13384–13421. doi:10.3390/molecules200713384.
3. Lam JH, Katritch V. Navigating structure-based drug discovery with emerging innovations in physics- and knowledge-based approaches. *npj Drug Discovery*. 2025;2:29. doi:10.1038/s44386-025-00031-4.
4. Zhong Z, Durrant JD. Generative AI in structure-based drug discovery. *BMC Biology*. 2026;24. doi:10.1186/s12915-026-02575-x.
5. Agboola OE, Agboola SS, Adedeji OS, et al. A critical appraisal of structure-based virtual screening, binding free energy methods, and AI-drivenmolecular design across the evolving computational landscape of drug discovery. *Journal of Computer-Aided Molecular Design*. 2026;40 (1). doi:10.1007/s10822-026-00955-7.
6. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Research*. 2000;28 (1):235–242. doi:10.1093/nar/28.1.235.
7. Varadi M, Anyango S, Deshpande M, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence spacewith high-accuracy models. *Nucleic Acids Research*. 2022;50 (D1):D439–D444. doi:10.1093/nar/gkab1061.
8. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583–589. doi:10.1038/s41586-021-03819-2.
9. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.*Journal of Computational Chemistry*. 2010;31 (2):455–461. doi:10.1002/jcc.21334.

10. Li Y, Zhan R, Rao J. Structure-informed machine learning for drug discovery: a task-centric perspective. *Briefings in Bioinformatics*. 2026;27 (1):bbag081. doi:10.1093/bib/bbag081.
11. Baell JB, Holloway GA. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for theirexclusion in bioassays. *Journal of Medicinal Chemistry*. 2010;53 (7):2719–2740. doi:10.1021/jm901137j.

HOW TO CITE: Varade Omkar, Navale Komal, Cholke Pratiksha, A Validation-First Workflow for Structure-Based Drug Design, Int. J. of Pharm. Sci., 2026, Vol 4, Issue 11, 74-79. https://doi.org/10.5281/zenodo.23876741