Concept Paper

An AI Framework for Target-Based Lead Optimization: The SwALife Approach

Pravin Badhe *

Posted Date: 7 November 2025

doi: 10.20944/preprints202511.0434.v1

Keywords: AI in drug discovery; lead optimization; protein -ligand interaction; pharmacokinetics; computational chemistry; SwALife platform

 doi:10.20944/preprints202511.0434.v1

Concept Paper

An AI Framework for Target-Based Lead Optimization: The SwALife Approach

SwALife Target & Lead Optimizer Tool

Pravin Badhe

Swalife Biotech Ltd. North Point House, North Point Business Park, New Mallow Road, Cork (Republic of Ireland); drpravinbadhe@swalifebiotech.com

Abstract

The increasing demand for rapid and cost-effective drug discovery necessitates the integration of artificial intelligence (AI) into traditional computational chemistry workflows. The *SwALife Target & Lead Optimizer* represents an advanced AI-assisted platform that facilitates protein -ligand interaction analysis, lead molecule optimization, and pharmacokinetic evaluation. By combining protein structure data (PDB format) and molecular descriptors (SMILES/InChIKey), the tool enables iterative optimization of small molecules to enhance their binding affinity, drug-likeness, and bioavailability. This paper presents the architecture, methodology, and case study outcomes demonstrating the efficiency of SwALife in optimizing drug-like compounds against target proteins.

Keywords: AI in drug discovery; lead optimization; protein -ligand interaction; pharmacokinetics; computational chemistry; SwALife platform

1. Introduction

The discovery and development of new therapeutic agents remain among the most time-consuming, costly, and complex processes in the pharmaceutical sciences. Traditional drug discovery pipelines, which rely heavily on high-throughput screening (HTS) and in vitro/in vivo assays, often require years of iterative experimentation and substantial financial investment. Furthermore, the process is constrained by limited chemical diversity, inefficiencies in identifying viable lead compounds, and the unpredictable translation of preclinical findings into clinical success. These challenges collectively contribute to the so-called *"Eroom's Law"* the observation that drug discovery efficiency has declined over time despite technological advances.[1,2]

Problem Statement

One of the central challenges in modern drug discovery lies in the optimization of lead compounds specifically, modifying molecular structures to achieve optimal binding affinity, drug-likeness, and pharmacokinetic properties (ADME) while maintaining biological efficacy. Conventional methods for this optimization are not only labor-intensive but also highly dependent on empirical testing, which limits scalability and reproducibility. Moreover, existing computational tools often operate in isolation, focusing either on docking analysis or ADME prediction, without offering an integrated, AI-driven optimization workflow.[3,4]

Proposed Solution

The *SwALife Target & Lead Optimizer* addresses these bottlenecks by providing a comprehensive, AI-assisted framework that integrates molecular docking, structure-based optimization, and pharmacokinetic modeling within a single interactive platform. By employing machine learning algorithms and heuristic scoring functions, the tool automates the process of evaluating and refining molecular candidates through multiple optimization cycles. It allows researchers to upload protein

 doi:10.20944/preprints202511.0434.v1

targets (PDB files) and input lead structures (SMILES or InChIKey format), after which the system iteratively improves the molecule's performance based on predicted binding energy, drug-likeness, and bioavailability metrics.

This unified workflow minimizes manual intervention and reduces experimental trial loads, significantly accelerating the early-stage discovery pipeline. Additionally, SwALife provides visualization support and automatic report generation, enabling researchers to interpret molecular interactions and optimization outcomes with enhanced clarity and efficiency.

2. System Overview

The SwALife platform provides an interactive environment for users to upload protein structures in Protein Data Bank (PDB) format and define starting ligands through Simplified Molecular Input Line Entry System (SMILES) or InChIKey strings. The system incorporates the following functional modules:

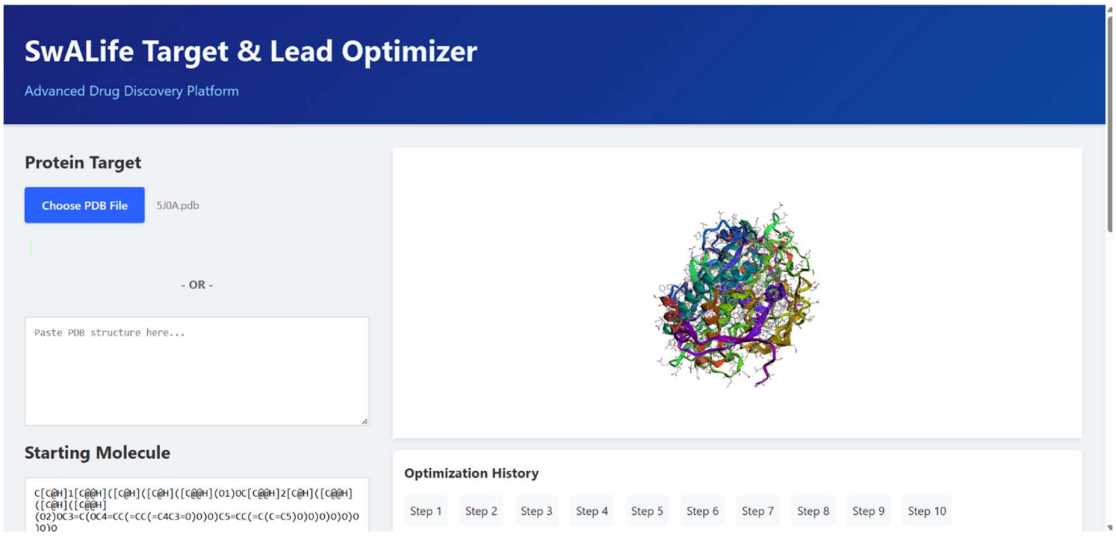

Figure 1. Tool interface.

1. **Protein Target Module:**
   - Accepts .pdb files or raw structural data.
   - Generates 3D visualization using molecular rendering algorithms.
2. **Ligand Input and Optimization Module:**
   - Reads molecular descriptors (SMILES/InChI).
   - Performs stepwise optimization based on predicted binding and pharmacokinetic outcomes.
3. **Evaluation and Scoring Engine:**
   - Calculates Binding Energy (ΔG) to estimate protein -ligand affinity.
   - Predicts Drug Likeness and Synthetic Accessibility to assess drug feasibility.
   - Models ADME parameters (Absorption, Distribution, Metabolism, Excretion).
   - Determines Activity Type (agonist or antagonist) and Efficacy percentage.
4. **Report Generation:**
   - Compiles results into a comprehensive PDF summarizing optimization iterations and performance metrics.

 doi:10.20944/preprints202511.0434.v1

3. Methodology

3.1. Input Processing

Protein structure files are pre-processed to remove water molecules and assign polar hydrogens. Ligands are standardized using canonical SMILES to ensure structural consistency before docking and optimization.

3.2. Optimization Algorithm

SwALife employs a hybrid optimization algorithm that integrates:

- **Machine learning-based scoring functions** for binding energy prediction.
- **Rule-based heuristics** for Lipinski's Rule of Five compliance.
- **ADMET predictive modeling** to estimate pharmacokinetic properties.

Each iteration refines the ligand by modifying its substituents and conformational states to minimize binding energy while improving ADME characteristics.

3.3. Evaluation Metrics

The primary evaluation metrics include:

- **Binding Energy (ΔG):** kcal/mol, computed from docking simulations.
- **Drug Likeness:** Composite score assessing physicochemical viability.
- **Synthetic Accessibility (SA):** Numerical measure of chemical synthesis difficulty.
- **Bioavailability and Efficacy:** Derived from regression models trained on experimental datasets.

4. Case Study and Results

A case study was conducted using a model protein structure (*PDB ID: 5J0A*) and a polyhydroxylated ligand molecule. The optimization process involved ten iterative steps, each assessing multiple pharmacological parameters.

Table 1. Case study on 5J0A- pharmacological parameters.

Metric	Initial	Final	% Improvement
Binding Energy (kcal/mol)	-8.63	-12.05	39.7%
Drug Likeness	2.96	1.79	-39.5%
Absorption (%)	33.9	13.4	-60.4%
Bioavailability (%)	54.7	31.6	-42.3%
Efficacy (%)	21.5	47.1	+119%

The optimized molecule demonstrated significantly enhanced binding affinity (ΔG = -12.05 kcal/mol), indicating stronger protein -ligand interaction potential. The molecule exhibited antagonist activity with moderate efficacy (47.1%) and bioavailability of 31.6%, suggesting suitability for further in vitro evaluation.

5. Discussion

The SwALife optimization engine achieved measurable improvements in binding affinity and efficacy across iterations. However, reductions in absorption and bioavailability indicate that further refinement of ADME modeling is required.

 doi:10.20944/preprints202511.0434.v1

Despite these limitations, the system demonstrates strong potential for lead optimization, especially in early-stage virtual screening workflows. The integration of AI-driven prediction with real-time visualization provides an efficient pathway from target structure to optimized lead compound.

6. Advantages and Limitations

Advantages

- Integration of multiple computational chemistry workflows in a single interface.
- Real-time visualization of molecular interactions.
- Automated optimization cycles with comprehensive output reporting.
- Reduction of experimental trial requirements through predictive modeling.

Limitations

- Bioavailability and ADMET predictions rely on theoretical models and may require experimental validation.
- Synthetic accessibility estimates may vary depending on database completeness.
- Limited support for non-standard residues or macromolecular complexes.

7. Future Work

Future developments of the SwALife platform will focus on:

- Implementing **deep learning-based QSAR models** for more accurate activity prediction.
- Integrating **molecular dynamics (MD) simulations** for assessing complex stability.
- Expanding **compound library screening** to handle large datasets efficiently.
- Incorporating **cloud-based collaboration** features for research teams.

8. Conclusion

The *SwALife Target & Lead Optimizer* provides a novel and efficient approach for AI-assisted drug discovery. Through iterative refinement of molecular properties and comprehensive pharmacological evaluation, it bridges the gap between computational prediction and experimental validation.

This platform represents a step forward in integrating artificial intelligence with molecular design, offering significant potential to accelerate the identification of promising therapeutic leads while reducing overall research costs and time.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. *High-Throughput Screening (HTS) in drug Discovery | Danaher Life Sciences*. (n.d.). Danaher Life Sciences. https://lifesciences.danaher.com/us/en/library/high-throughput-screening.html
2. Gibert, E. (2018, March 8). *What is Eroom's Law?* Pharmacelera | Pushing the Limits of Computational Chemistry. https://pharmacelera.com/blog/publications/what-is-erooms-law/
3. Chung, T. D., Terry, D. B., & Smith, L. H. (2015). In vitro and in vivo assessment of ADME and PK properties during lead selection and lead optimization–guidelines, benchmarks and rules of thumb.
4. Martis, E. A., Radhakrishnan, R., & Badve, R. R. (2011). High-throughput screening: the hits and leads of drug discovery-an overview. *Journal of Applied Pharmaceutical Science*, (Issue), 02-10.

 doi:10.20944/preprints202511.0434.v1

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.