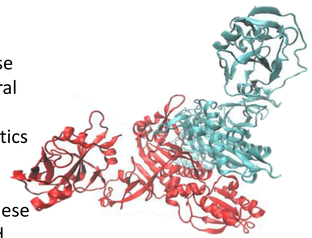
Application of Fragment-based Generative Models to Optimize Enzyme Inhibitors

Nakul Balaji and Breanna Hayden

Mentor: Dr. Sudeep Bhattacharyay. Collaborator: Dr. Sanchita Hati | Department of Chemistry and Biochemistry, UW-Eau Claire

Introduction & Objectives

Prolyl-tRNA synthetase (ProRS) is an enzyme essential to life found in all living organisms that catalyzes the attachment of the amino acid proline to its corresponding tRNA during protein synthesis[1]. Because ProRS is required for cell survival and exhibits structural differences across species, it represents a promising target for the development of species-specific antibiotics that can selectively inhibit microbial pathogens while minimizing side effects in humans. We investigated a dataset of ligands similar to ATP, Halofuginone (a Chinese derivative of herbal medicine that inhibits ProRS), and NCP26 (a drug that inhibits ProRS) in *Staphylococcus aureus* ProRS[2].

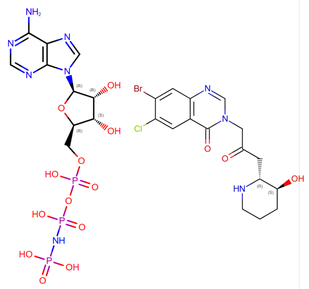


Prolyl-tRNA Synthetase(ProRS)[5]

To optimize these species-specific binding affinities, we used:

- Modof, a molecular optimization framework that generates chemically valid ligand modifications[3].
- A physics-based docking model to calculate molecular interactions and test if the compounds binding improved[4].

Our goal is to optimize ligand structures to improve binding affinity toward target ProRS enzymes while preserving chemical validity.



ANP (ATP-like)[6]

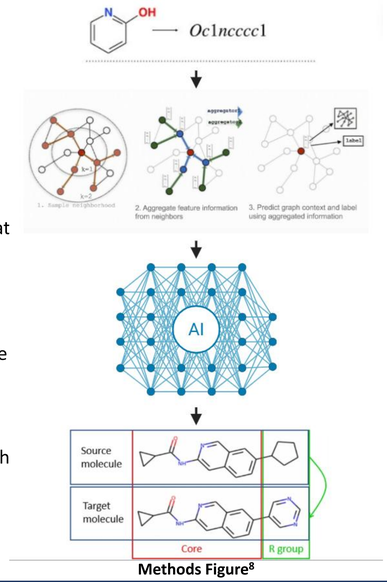
HFG (Halofuginone)[7]

Background

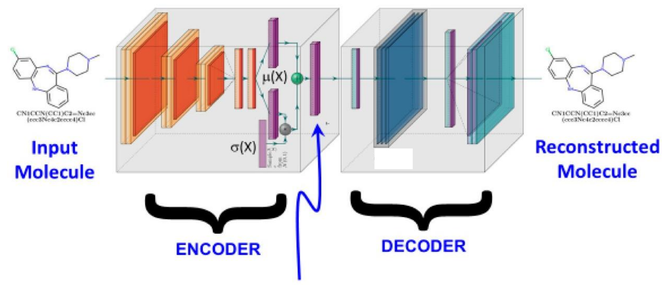
- ProRS is an attractive antibacterial target because it is essential for protein synthesis yet exhibits structural differences across species, enabling selective inhibitor design.
- Traditional drug discovery requires screening millions of compounds, making the identification of species-specific ProRS inhibitors both time-consuming and computationally expensive.
- Generative machine learning models offer an alternative approach by optimizing existing drug molecules to improve predicted binding affinity rather than relying solely on large-scale virtual screening.
- Combining AI-guided molecular optimization with physics-based docking enables rapid evaluation of potential inhibitors and prioritization of molecules with improved predicted ProRS binding.

Methods

- Molecules are represented as graphs, where atoms are treated as nodes and chemical bonds as edges, enabling structural information to be learned directly.
- A graph-based encoder generates a latent representation of each ligand that captures its chemical and topological features.
- The generative model proposes molecular edits (adding, removing, or modifying atoms and bonds) to produce chemically valid analogs while preserving the parent structure.
- An optimization objective is to balance improved predicted binding affinity with structural similarity and chemical validity, selecting candidate molecules that are both realistic and likely to exhibit stronger target binding.



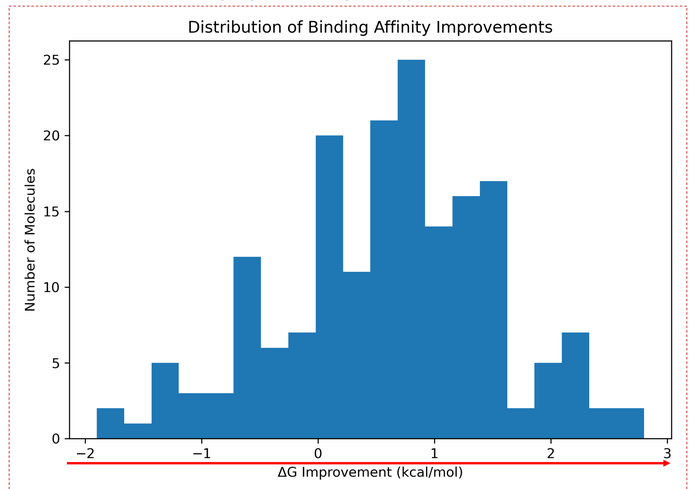
Methods Figure[8]



Deep Learning Pipeline[9]

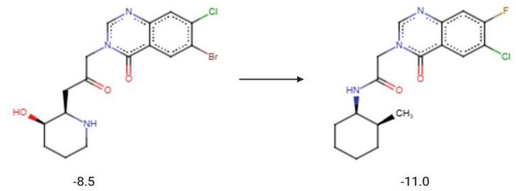
Current Results

- Molecular docking was performed on *Staphylococcus aureus* ProRS (Sa ProRS) with ligands similar to ATP, NCP26, and halofuginone, producing binding affinity values using physics-based simulations.
- Original and optimized binding affinity values were compared to quantify improvement and identify structural modifications associated with stronger binding, with Halofuginone exhibiting higher binding affinity values across the board.



Histogram of HFG Binding Affinities[10]

The histogram exhibits an approximately bell-shaped distribution with a slight rightward skew. Most predicted affinity changes fall between zero and one, indicating that the majority of molecular modifications have only modest effects on binding affinity.



Molecular Optimization of Halofuginone-like Ligand[11]

Fluorine (F): Highly electronegative and may interact favorably with an electron-deficient/electron-withdrawing region of the ProRS binding pocket. Replacing fluorine with a methyl ($-CH_3$) group decreases polarity and increases hydrophobicity, potentially improving interactions with nearby hydrophobic residues.

Chlorine (Cl): The substitution of fluorine with chlorine may be due to chlorine's larger atomic size, greater polarizability, or altered electronic properties, leading to more favorable interactions within the binding pocket.

Amine (NH): The amine substitution may be driven by favorable π–π stacking, conjugation effects, or changes in hydrogen-bonding interactions that improve ligand binding or stability within the active site.

Molecular Optimization Metrics

Metric	Value
Total molecules evaluated (molecules)	738
Positive improvements (molecules)	216 (29.3%)
Best improvement (kcal/mol)	0.73 ± 0.69
Best similarity	0.70 ± 0.11

Optimization Metrics for HFG Binding Affinities[12]

- Screened 738 molecules, with nearly 30% exhibiting measurable optimization.
- Optimized candidates maintained high structural similarity (~70%) while achieving moderate binding improvements.
- Results suggest the model preferentially generates conservative structural modifications rather than large backbone changes.

Statistical Metrics

	Original	Optimized
Mean (kcal/mol)	-7.37	-8.17
Variance (kcal/mol)	0.383	1.08
Observations (molecules)	193	193
Pearson Correlation	0.377	
P(T<=t) one-tail	6.48×10^{-23}	

Statistical Metrics for HFG Binding Affinities

- Modof generated chemically valid analogs with significantly improved predicted binding affinity, reducing the mean docking score from −7.37 to −8.17 kcal/mol (one-tail paired *t*-test, $p < 0.05$).
- Optimized molecules generally outperformed their original counterparts across 193 ligand pairs, demonstrating that the model effectively learns structural modifications that enhance ligand binding to ProRS.

Conclusion & Next Steps

- Modof generated chemically valid analogs of ligands similar to ATP, halofuginone, and NCP26, and physics-based redocking identified multiple optimized compounds with improved predicted binding affinity toward *S. aureus* ProRS.
- These results demonstrate that AI-guided molecular optimization can effectively propose structural modifications to ligands that strengthen predicted ProRS binding while preserving the parent structure, providing improved methods for future antibiotic development.
- Future work will validate the top-performing analogs using further simulations or experimental methods and extend the optimization workflow to additional ProRS species to enable selective binding.

References & Acknowledgements

[1]*Hu et al.* Editing Domain Motions Preorganize the Synthetic Active Site of Prolyl-tRNA Synthetase. *ACS Catal. 2020 Sep 4;10(17):10229-10242. doi: 10.1021/acscatal.0c02381.*

[2]*Cheng et al.* Structure-Guided Design of Halofuginone Derivatives as ATP-Aided Inhibitors Against Bacterial Prolyl-tRNA Synthetase. *J Med Chem. 2022 Dec 8;65(23):15840-15855. doi: 10.1021/acs.jmedchem.2c01496.*

[3]*Chen et al.* A deep generative model for molecule optimization via one fragment modification. *Nature Machine Intelligence 3:2522-5839*

[4]*Pradeep et al.* AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility. *PLOS Computational Biology 11(12): e1004586.*

[5] *Liu et al.* Interplay between bacterial deubiquitinase and ubiquitin E3 ligase regulates ubiquitin dynamics on Legionella phagosomes *eLife 9:e58114*

[6] *Dranow et al.* Crystal Structure of Prolyl-tRNA Synthetase from Onchocerca volvulus with bound Halofuginone and nucleotide

[7]*Dranow et al.* Crystal Structure of Prolyl-tRNA Synthetase from Onchocerca volvulus with bound Halofuginone and nucleotide

[8]*Biorender: Scientific Figure and Illustration Software*, www.biorender.com/.

[9]*Samanta et al.* VAE-Sim: A Novel Molecular Similarity Measure Based on a Variational Autoencoder. *Molecules. 2020 Jul 29;25(15):3446. doi: 10.3390/molecules25153446. PMID: 32751155*

[10,11,12]*This work*

[13]*Monsia et al.* Efficient and Explainable Virtual Screening of Molecules through Fingerprint-Generating Networks Integrated with Artificial Neural Networks. *ACS Omega. 2025 Jan 28;10(5):4896-4911. doi: 10.1021/acsomega.4c10289.*

This research was funded by the National Science Foundation (NSF) Research Experience for Undergraduates (REU) grant OAC- 2447779. The computational resources of this study were provided by the Blugold Center for High-Performance Computing under NSF grant CNS-1920220