Validating an Antibiotic Screening of the *E. coli* Proteome

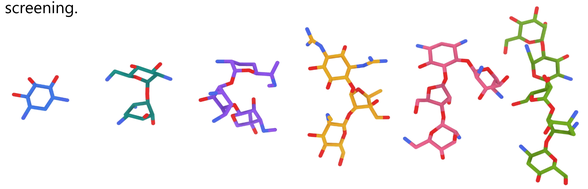
ROBINSON OHM, AAYATTI M. GUPTA, ERIKA A. TAYLOR
Department of Chemistry, Wesleyan University | QAC Summer 2026

INTRODUCTION

Aminoglycosides (AGs), a class of highly effective antibiotics, are historically understood to inhibit essential bacterial functions by binding to the cell ribosome. Our lab unexpectedly discovered an alternative binding target, suggesting additional drug-protein interactions beyond the ribosome.

Therefore, we seek to identify any similar phenomena by surveying the remainder of the proteome. This large-scale, unbiased surveillance is most efficiently conducted in a virtual environment. Molecular docking software first ranks drug-protein binding hits to prioritize candidates for further laboratory investigation with the ultimate goal of improving patient care by developing and broadening AG usage.

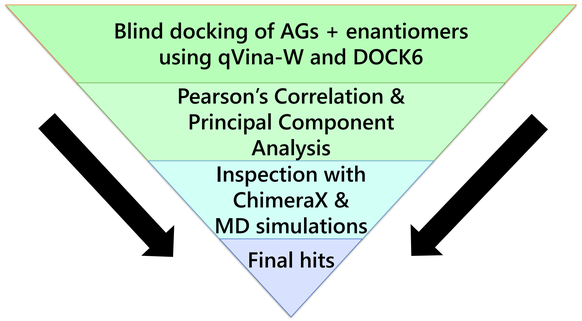
We chose six AGs to computationally dock upon structurally characterized proteins throughout the *E. coli* proteome. This sub-project refines our docking protocol with rigorous validation conditions to maximize the accuracy of the screening.



Left to right: 2-deoxystreptamine, neamine, gentamicin, streptomycin, neomycin B, and lividomycin exhibit diverse physical and chemical properties to collectively represent most AGs.

Validating a docking protocol for accuracy can save time and resources by omitting misleading results. Biological accuracy is met when the protocol reproduces known binding interactions and excludes false molecules from the binding site. These positive and negative controls must be satisfied before the protocol can reliably discover unknown docking hits.

INITIAL SCREENING

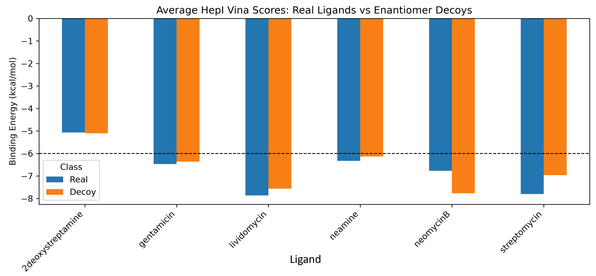


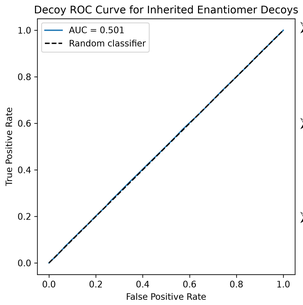
- The *E. coli* proteome is represented by 4,155 protein characterizations.
  - 6,676 models are available, but only one is used to represent each protein.
  - Models were obtained from the Protein DataBase and AlphaFold.
- 6 AG ligands and 6 corresponding enantiomer decoys (negative controls) were docked to each receptor.
  - Enantiomers have similar chemical properties but inverted molecular structure.
  - Negative controls ensure the protocol accounts for both structural and chemical properties of AGs.
- Docking software (QuickVina-W) used a "blind" method:
  - Each protein's binding pocket was defined not according to experimental observations, but rather computationally predicted.
  - Exhaustiveness was set to 20; CPUs were set to 8.
- Pearson's correlation was planned for analyzing impactful variables.
- Docking hits were also to be analyzed in Principal Component Analysis (PCA), where masses of data can be organized into target candidates.
- PCA hits were to be investigated in ChimeraX and MD simulations to ensure regional accuracy and physical stability of docked results, eventually yielding final hit candidates.

GENERAL VALIDATION

To assess the baseline docking protocol for methodological precision, we used the AG decoy results to implicitly gauge the specificity with which real AGs were docked to proteins. A high decoy binding rate indicates potential for a more selective protocol.

Under the initial protocol, the real AG molecules and the false AG enantiomer decoys bind with a similar docking score.



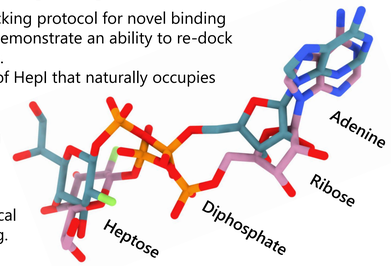


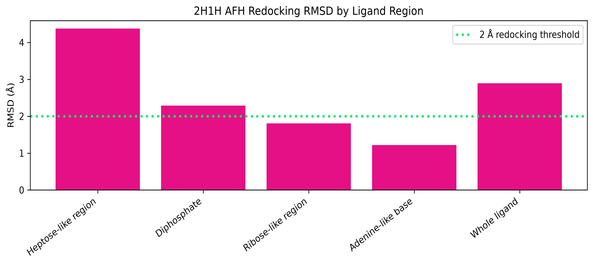
- The Receiver Operating Characteristic (ROC) curve displays the rate at which enantiomer decoys bind with more success than their AG counterparts, creating false positives.
- Our area under the curve (AUC) of 0.501 and near-complete overlap with the reference line indicates that the protocol provides almost no molecular selectivity.
- Advancement down the original result analysis pipeline would be premature before refinement.

FOCUSED VALIDATION

We assessed biological accuracy by examining Heptosyltransferase I (HepI), a bacterial membrane-reinforcement enzyme. Various characterizations of HepI were used to explore possible drug binding locations across the same protein.

- Before we rely on the docking protocol for novel binding discoveries, it must first demonstrate an ability to re-dock molecules to known sites.
- ADPH is a native ligand of HepI that naturally occupies the donor binding site.
- Displayed to the right is a redocked ADPH analog upon an experimentally-defined native pose.
- This superimposition shows the partial biological accuracy of the redocking.

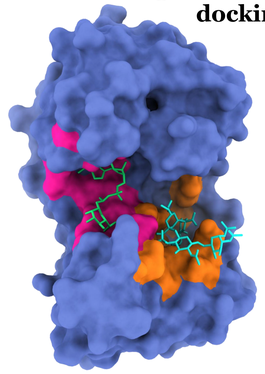




- Root Mean Square Deviation (RMSD) measures average distance between atoms.
- The heptose-analogous region raises overall RMSD to 2.89 Å, exceeding the 2.00 Å benchmark and further warranting a protocol refinement.

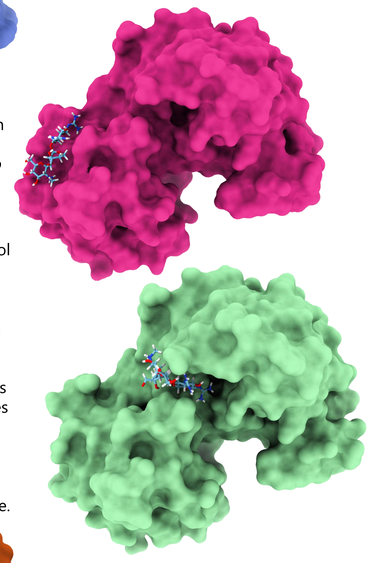
A redock of HepI to all 12 ligands under the initial protocol yielded a small binding energy range of ±0.8 kcal/mol, supporting reproducibility.

Multiple characterizations of HepI were recruited to investigate diverse complications in the docking process.



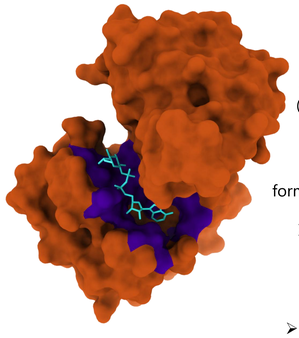
Characterization 1: Both native ligands experimentally bound to HepI

- HepI transfers heptose sugar from ADPH (green) to Kdo_2–Lipid A (cyan) for bacterial membrane assembly.
- The two native ligands occupy distinct binding pockets (magenta & orange).
- Each binding pocket should be defined according to external verification and screened independently.



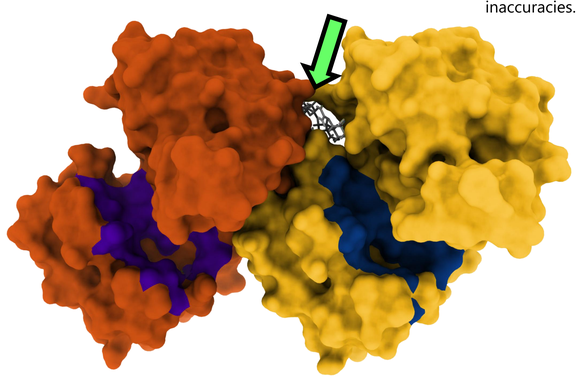
Characterization 2: Streptomycin bound to two crystallographic characterizations of HepI: an *apo* (magenta) receptor & a *holo* receptor with the ADPH ligand removed (green)

- The baseline docking protocol uses binding scores to estimate the binding pocket location, meaning any high-scoring region of the protein can be considered a pocket.
- Despite different predicted pockets, both representations returned near-identical scores (-8.2 kcal/mol and -8.1 kcal/mol).
- Therefore, docking scores alone cannot identify the most biologically relevant site.



Characterization 3: HepI variant with experimentally-defined binding pocket (left), & streptomycin incorrectly bound between two packed crystals (below)

- Many characterizations are obtained via protein crystallization, forming a structure that is computationally usable but physically affected.
- This overlap of structure caused the incorrect docking of streptomycin into a non-biological cleft between packed protein complexes outside of the experimentally-defined site (purple).
- The protocol should separate crystal-packed proteins to prevent such inaccuracies.



REFINED STRATEGY

Identify control cases for validation.
HepI- primary case | OSBS- monomeric control | PNP- multimeric control

Select assembly of complexes.
Use software such as PDBePISA and ProtCID to identify monomers and multimers | *Remove crystal-packing* before locating the docking site.

Prepare receptors and ligands.
Consistently apply waters, cofactors, hydrogens, partial charges, and file conversion.

Define binding sites
Instead of blind docking, reference native ligands, known pockets, pocket-finding software (SiteMap). | When a protein has *multiple sites* (see HepI), define and dock each site separately.

Use reproducible, specific docking configurations.
Record the receptor and ligand state, retained substrates, grid center and dimensions, exhaustiveness, output, and log for every run.

Redock native ligands.
Compare predictions with experimental poses using overall/regional RMSD and visual inspection.

Investigate distinct biological cases.
Confirm that the protocol works for *multiple docking regions* of HepI, a simple monomeric OSBS, and a complex multimeric PNP.

Improve decoy controls.
Replace the original enantiomers with property-matched DeepCoy molecules. | Compare ROC/AUC and inspect visually.

Interpret location as well as score.
Determine if pose occupies a known site, allosteric candidate, or crystal-packing complication. | Save receptor, active ligand, and decoy poses together for easy comparison.

Apply the validated protocol.
Screen AGs according to the new protocol and re-evaluate using general and focused validation.

Continue the overall workflow.
Return to DOCK6, hit ranking, PCA, visualization, MD, and eventually, *in vitro* experimentation.

KEY TAKEAWAYS

- Original enantiomer decoy controls were not distinguished from the real AGs.
- Blind docking based off of the docking score alone could not establish a biologically correct drug-binding location.
- Improved controls, site-specific redocking, and biologically relevant assembly are required before interpreting proteome-scale hits.

REFERENCES

1. Carter, A., Clemons, W., Brodersen, D. et al. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. Nature 407, 340–348 (2000). https://doi.org/10.1038/35030019
2. Milicaj, J., Hassan, B. A., Cote, J. M., Ramirez-Mondragon, C. A., Jaunbocus, N., Rafalowski, A., Patel, K. R., Castro, C. D., Muthyala, R., Sham, Y. Y., & Taylor, E. A. (2022). Discovery of first-in-class nanomolar inhibitors of heptosyltransferase I reveals a new aminoglycoside target and potential alternative mechanism of action. Scientific reports, 12(1), 7302. https://doi.org/10.1038/s41598-022-10776-x
3. Bender, B.J., Gahbauer, S., Luttens, A. et al. A practical guide to large-scale docking. Nat Protoc 16, 4799–4832 (2021). https://doi.org/10.1038/s41596-021-00597-z

ACKNOWLEDGMENTS

I would like to thank Professor Emannuel Kaparakis and the Quantitative Analysis Center for funding and managing this summer research apprenticeship, through which my project was completed.