Computational Screening of Indian Phytochemicals as Candidate Efflux-Pump Inhibitors of Salmonella AcrB

Pranoy Mathur[1]

Received August 16, 2026
Accepted October 11, 2026
Electronic access November 15, 2026

The AcrB efflux pump in *Salmonella Typhimurium* is a major mechanism in the development of drug resistance. The purpose of the study is to screen drug-like Indian phytochemicals from the IMPPAT 2.0 database against the AcrB protein, benchmarked against the known inhibitors of PAβN and NMP. Molecular docking (Vina), machine learning (RDKit descriptors, random forest), and molecular dynamics for the top hit were utilized. Of the 863 compounds docked against AcrB's T-state, 254 compounds produced stronger scores than PAβN. The top hit was Glabroisoflavanone B (IMPHY001793), with a score of -11.040 kcal/mol. The protocol was validated by redocking minocycline to 1.05 Å; decoy enrichment, seed robustness and conformer-ensemble docking characterized the model's limits. The ML classifier used was able to separate compounds in the two extreme docking-score classes with computed scaffold-split AUC of 0.944, and external test AUC of 0.949. The classification was mainly driven by RingCount (+0.0250), LogP (+0.0236), RotBonds (+0.0156). Across the five top-scoring compounds, contact analysis (4 Å cutoff) showed that the residues F178, F610, V612 and F628 were contacted by all five, and F617 by three of the five. ADMET profiling of the 17 primary hits ($z \leq -2$) showed that a majority of them had favorable drug-like properties. A 100 ns molecular dynamics simulation was analyzed to ascertain the top-ranked compound's docked pose's dynamic stability. The ligand maintained contact with the pocket without dissociation despite the larger-scale protein motion. A prioritized list of candidate AcrB inhibitors was produced, assessable through further experimental studies.

Keywords: Bioinformatics, Computational Pharmacology, Antimicrobial resistance, Efflux pump inhibitors, Molecular docking

Introduction

Antimicrobial resistance (AMR) is a growing threat, with over 4.71 million deaths associated with the phenomenon in 2021 alone[1]. Antibiotics are drugs that target bacteria and have been historically successful in increasing life expectancy, reducing financial burdens on healthcare, and minimizing risks of disability. However, their overuse and misuse paired with bacteria's ability to rapidly evolve and transfer resistance genes are threatening to unravel decades of progress. It is further projected to be associated with 8.22 million deaths per year by 2050[1], making it crucial to advance methods to prevent and combat its proliferation[2].

One crucial mechanism in achieving drug resistance and multi-drug resistance (MDR) in bacteria is through protein transporters called efflux pumps, responsible for expelling antibiotics out of the cell[3]. They can be the quickest and most efficient mechanism among the numerous methods possessed by bacteria to attain resistance, though it depends on the specific substrate.

Of the five major families of efflux pumps, the resistance nodulation division family (RND) is responsible for multi-drug resistance in gram-negative bacteria including *Salmonella* and *Escherichia coli* (*E. coli*)[3]. They exist as a tripartite system that consists of the transporter protein, accessory protein and outer-membrane protein. The AcrAB-TolC pump, present in *Salmonella* and targeted in this study, is an example of this division. AcrB is the inner-membrane transporter protein that is responsible for the expulsion of antibiotics and in recognizing the substrate. Additionally, it is polyspecific making it highly successful in pumping out a broad range of antibiotics and increasing chances of MDR[4].

In 2017 alone, invasive non-typhoidal *Salmonella* (iNTS) was observed in approximately 535,000 cases and caused 77,500 deaths globally[5]. The large case-fatality rate of approximately 14.5% shows *Salmonella*'s clinical severity, only worsened in India through increasing antibiotic-resistance infections. Additionally, AcrB exports host-derived antimicrobial compounds such as bile salts and host-defense molecules, contributing to bacterial virulence by allowing increased colonization of the intestine[6,7]. As *S. Typhimurium* is a leading NTS serovar, this study targets the AcrB transporter of *S. Typhimurium*.

[1] *Inventure Academy; Bengaluru, India*

© The National High School Journal of Science 2026

As AcrB can export antibiotics that are structurally diverse, potential inhibitors would enable the continued use of existing antibiotics rather than developing new ones. The use of efflux pump inhibitors (EPIs) would mean that the small molecules could bind to the protein and prevent the transport of substrates. The two best-characterized EPIs in literature are phenylalanine-arginine β-naphthylamide (PAβN)[8] and 1-(1-naphthylmethyl)-piperazine (NMP)[9]: both bind to the distal binding pocket (DBP) of AcrB and are able to increase the bacteria's susceptibility to antibiotics[10]. Though they can successfully bind to AcrB and inhibit its action, neither has transitioned to clinical trials as limitations including toxicity and unfavorable pharmacokinetic properties restrict their clinical applications[11]. Therefore, PAβN and NMP act as benchmarks against which candidate compounds are assessed.

Natural products are an underused area when screening for EPI candidates. Their chemical variety due to evolution gives them promise for this purpose[12]. Indian medicinal plants have been historically used for therapeutic purposes stemming from their abundance of bioactive compounds. The Indian Medicinal Plants, Phytochemistry And Therapeutics (IMPPAT 2.0) database provides a curated and structured list of phytochemicals derived from Indian medicinal plants along with their usable structures, making *in silico* screening easy[13].

In a study by El-Demerdash et al., 2024[14], only a few specific phytochemical compounds were tested. Oyedara et al. docked 71 phytochemicals from Mexican medicinal plants against *Salmonella* AcrB[15], and Mehta et al. identified an efflux-pump inhibitor from Himalayan medicinal plants, but docked it against the regulators RamA and RamR rather than AcrB itself[16]. This study, on the other hand, screens the whole drug-like IMPPAT library, benchmarked against PAβN and NMP, and includes an interpretable machine learning (ML) on the *S. Typhimurium* AcrB DBP. This study aims to evaluate whether drug-like Indian phytochemicals can produce docking affinities competitive with PAβN and NMP, and to observe whether an ML model can be built to identify the features separating strong and weak docking scores. It also aims to assess, via molecular dynamics simulations, whether the top-ranked candidate by docking affinity has a stable binding pose.

Methods

Selection of target protein

The protein selected was AcrB from *Salmonella enterica* serovar Typhimurium. Within *S. Typhimurium*, the Resistance Nodulation cell Division (RND) family of efflux pumps plays the most significant role in the pumping out of toxins, including antibiotics. The AcrB protein is the inner-membrane transporter in the AcrAB-TolC efflux pump complex from the RND family. Within the complex, it is responsible for recognizing and expelling substrates towards TolC, and can do so for a diverse range of antibiotic classes. Accordingly, it is a drug-relevant target. The UniProt number of the protein is Q8ZRA7, strain LT2, which consists of 1049 amino acids and is coded by the gene *acrB*[17]. An experimental *Salmonella* AcrB structure (PDB 6Z12)[18] exists but its 4.60 Å resolution is insufficient for pocket-level docking. The AcrB protein functions as an asymmetric trimer that cycles through three distinct states: access (L), binding (T), and extrusion (O)[19]. The DBP is associated with the binding (T) state, and given that a symmetric template cannot be assumed to have the correct pocket geometry, a T-state template was used. A homology model was built in SWISS-MODEL[20] using 4DX5 (chain B, T protomer, *E. coli* AcrB, 1.9 Å) as the template[19]. Importantly, the template and target sequence share 94.96% sequence identity. As the receptor is a homology model and not a crystal structure, the quality of the model produced was tested. SWISS-MODEL's global model quality confidence estimates, GMQE and QMEANDisCo both produced 0.85[21], indicating a high-confidence model was produced. The Ramachandran statistics (97.77% of residues in favored regions, 0.29% outliers), the clashscore of 0.95 (99$^{\text{th}}$ percentile), and the MolProbity score (1.45)[22], also suggest that the model's backbone geometry is strong. Though 6.42% rotamers were poor against a $<$0.3% target, and only 90.37% favored against $>$98% goal, indicating that side-chain rotamers are weak, this result is expected for a homology model. The concern regarding side-chain accuracy is addressed in the docking section that details the redocking validation conducted. The DBP lies entirely within a single protomer and is not formed at a protomer-protomer interface[19], so a single chain (residues 1–1032 of the 1049-residue protein) was modeled. As seen in Table 1, the pocket was conserved, with 19 out of 21 pocket positions being identical, including the conservation of all six Phe; only I277 was changed to V and M573 to L, both substitutions being conservative.

Receptor and Binding-Site Preparation

Using Open Babel 3.1.0[23], polar hydrogens were added and AutoDock atom types assigned; non-polar hydrogens are not included in the rigid-receptor PDBQT. The protein was subsequently saved in PDBQT format. Since Vina's scoring function does not factor in partial charges, their omission makes no change to the scoring. The presence of the six residues that make up the DBP (F136, F178, F610, F615, F617, F628)[19,24] was verified in PyMOL[25]. The presence of the G-loop (flexible loop, residues 615–620, sequence FGFAGR, beside the DBP) was also confirmed[19]. The docking grid was prepared with a size of 22 $\times$ 22 $\times$ 22 Å, center x = 35.75, y = $-$60.85, z = $-$41.81, which was taken from the crystallographic minocycline centroid[19].

© The National High School Journal of Science 2026

Table 1 Mapping of *E. coli* AcrB binding-pocket residues onto the *Salmonella* Typhimurium model.

Position	E. coli (4DX5)	S. Typhimurium model	Identical
46	SER	SER	✔
89	GLN	GLN	✔
128	SER	SER	✔
136	PHE	PHE	✔
139	VAL	VAL	✔
176	GLN	GLN	✔
178	PHE	PHE	✔
273	GLU	GLU	✔
277	ILE	VAL	–
279	ALA	ALA	✔
563	PHE	PHE	✔
566	ASP	ASP	✔
570	GLY	GLY	✔
573	MET	LEU	–
610	PHE	PHE	✔
612	VAL	VAL	✔
615	PHE	PHE	✔
617	PHE	PHE	✔
620	ARG	ARG	✔
626	ILE	ILE	✔
628	PHE	PHE	✔

✔ identical; – substituted. 19 of 21 positions are identical. Both substitutions (I277V, M573L) are conservative, and all six phenylalanines of the distal pocket are conserved.

Compound Library and Ligand Preparation

The IMPPAT 2.0 database, a collection of Indian medicinal plant phytochemicals, was used to source the ligands. 17,967 phytochemicals in the database were reduced to 865 using the following criteria: Lipinski RO5 Violations = 0[26], Veber = Good[27], Egan = Good[28], Ghose = 0[29], QEDw > 0.8[30]. The QEDw value was chosen to obtain a drug-like subset as opposed to a representative sample of the IMPPAT database. It also provided a more computationally manageable set of ligands to dock exhaustively. The 2D SDFs were used for descriptor calculation, detailed in the Machine Learning section below. All 865 ligands and the two benchmarks (PAβN and NMP) were prepared using Open Babel 3.1.0[23], following the same pipeline: hydrogens were added with pH-based correction at pH 7.4, 3D coordinates were generated with MMFF94 over 100 optimization steps[31], they were written to PDBQT with AutoDock atom types, and no partial charges were assigned. The SMILES used to build the ligands were procured from the IMPPAT database for the 865-compound library and from the PubChem CIDs for PAβN (CID 443301) and NMP (CID 701891)[32]. These two were deemed appropriate benchmarks as they are well-established efflux pump inhibitors that bind to the AcrB DBP. For PAβN, it was confirmed that the CID gave the structure of the free base and not the salt.

Molecular Docking

AutoDock Vina 1.2.7 was used for molecular docking[33,34]. Exhaustiveness was set to 8 and up to 9 binding modes were requested per ligand. For each ligand, Vina's top-ranked (lowest energy) pose was selected. The benchmarks PAβN and NMP were docked to ensure that the known inhibitors produced sensible Vina scores. PAβN produced a score of -8.771 kcal/mol, and NMP's was -7.773 kcal/mol. Crucially, the structures were also visualized in PyMOL to confirm that they were sitting in the DBP[25]. All 865 compounds were docked under the same benchmark-validated protocol. The outcome was 863 successful dockings and 2 failures. The 2 failures (IMPHY005215 and IMPHY017899) were hydrochloride salts and the only two multi-component SMILES; Open Babel wrote each one as two disconnected fragments in a single PDBQT file which Vina cannot dock. They were excluded from any subsequent analysis. The exclusion of 2 ligands was not based on the binding scores making it unlikely that affinity-based selection bias was induced. To filter for the top hits, a z-score threshold of -2 (-9.969 kcal/mol) was computed over 863 docking scores with a mean of -8.322 kcal/mol and standard deviation of 0.824 kcal/mol. This produced 17 compounds, forming the primary hit set. As 254 out of the 863 compounds scored better than PAβN, the more stringent benchmark, the z-score method was implemented to identify top hits.

Protocol Validation and Benchmarking

For redocking validation, the co-crystallized ligand minocycline was extracted from the 4DX5 template[19] and redocked (using a protocol identical to the described above) into the finished model. The RMSD was heavy-atom, computed in place, with no superposition, and the best binding pose returned was reported. A success threshold of <2.0 Å was set prior to redocking. Subsequently, four more AcrB ligands with crystallized structures were tested: doxorubicin (4DX7)[19], D13-9001 (3W9H)[35], MBX2319 (5ENO), MBX2931 (5ENP)[36]. All of them were built using PDB ideal-geometry's chemical definitions via Meeko[37] as opposed to using SMILES via Open Babel (like the library of ligands) as the aim was to accurately reproduce a deposited crystallographic pose and not to filter through a SMILES database. Each crystal pose was then transferred into the model coordinate frame by Kabsch superposition on common Cα atoms, the same heavy-atom RMSD being employed. The fraction of atoms enclosed by the 22 Å grid box is reported in addition to the RMSD. Preparation of 7 known AcrB actives and 496 property-matched decoys from DUD-E[38] (503 total compounds) using the same protocol as the initial compound library was carried out. ROC-AUC and enrichment factors were computed at 1%, 5% and 10%. The top 50 compounds were re-docked using three independent random seeds (101, 202, 303), and agreement was assessed through Spearman rank correlation between seeds, Jaccard overlap of the 17-compound primary hit set, and the per-

© The National High School Journal of Science 2026

compound score spread. Five conformers were extracted from the 100 ns MD trajectory at roughly 20, 40, 60, 80 and 100 ns. For the protein only, each conformer was aligned to the reference model used by Cα superposition. The coordinates were then transferred onto the model_01 PDBQT template so that the atom typing is identical across all five conformers, and the geometry is the only variable factor. Docking of 150 compounds for each conformer was conducted: the top 50 and 100 sampled across the full score range in the library, as the top 50 on their own would only cover 1.47 kcal/mol against a conformer-to-conformer standard deviation of 0.41 producing an uninformative rank correlation.

Machine Learning

For each compound, the 2D structure (SDF format) was retrieved from IMPPAT and 15 physicochemical descriptors were computed using RDKit 2026.03[39]. The 15 descriptors accounting for size, polarity, shape, complexity, and drug-likeness were: MW, LogP, TPSA, HBD, HBA, RotBonds, AromaticRings, RingCount, HeavyAtoms, FractionCSP3, MolarRefractivity, HeteroAtoms, SaturatedRings, AliphaticRings, QED. A dataset was created with all 863 compounds and their 15 descriptors, paired with their docking affinity. The ML was performed locally using scikit-learn 1.9.0[40]. Murcko scaffolds were computed from the IMPPAT SMILES using RDKit's MurckoScaffoldSmiles[41], producing 355 distinct scaffolds across the 863 compounds. Twenty percent of the scaffolds (71 out of 355, comprising 176 out of 863) were left out completely as an external test set before any model was fitted, and evaluated once at the end. All cross-validation on the remainder was grouped by scaffold such that no scaffold was shown in both a training and a test fold, thereby ensuring that close analogues do not span the split. A conventional random stratified split was also run for the purpose of comparison, quantifying how much apparent performance is increased when the scaffold structure is ignored.

Both a regression and a binary classification approach were attempted, though the regression performed poorly. For the regression, the model RandomForestRegressor with n_estimators=400 and random_state=0 was used[42]. For its evaluation, GroupKFold cross-validation, grouped by Murcko scaffold, was set to n_splits=5. For binary classification, compounds were split by their docking affinity according to their percentile threshold: more favorable docking-scores were those with affinity $\leq$ 33rd percentile (-8.675) and less favorable docking-scores with affinity $\geq$ 67th percentile (-7.959). To ensure the model was given unambiguous training data, the middle third was excluded, leaving 285 compounds each in the upper and lower tertiles (570 total). A RandomForestClassifier(n_estimators=400, random_state=0, n_jobs=-1, max_depth=None) was trained on the two docking-score classes. Evaluation was done with GroupKFold (n_splits=5), scaffold grouped via cross_val_predict; AUC was computed once on pooled out-of-fold probabilities. A permutation test (1000 permutations, scaffold-grouped, exceedance count reported, null models using 120 trees, via GroupKFold) was run to assess the statistical significance[43]. Feature importance was assessed by permutation importance as opposed to built-in Gini importance to avoid distorting the ranking stemming from the presence of correlated descriptors. For each of the five scaffold-grouped folds, the model was fitted on the four training folds, and importance was only measured on the untouched held-out fold, as the fall in AUC when each feature is permuted. Repeats were carried out 20 times per fold and the mean of the values was found across the five folds.

Interaction and ADMET Analysis

The top 5 hits were analyzed for their binding-site contacts, where contact between ligand and binding-site residues was defined as heavy atoms within 4 Å of one another. The specific DBP residues that had contact were checked for each hit. SwissADME was applied to the 17 hits of the primary hit set[44] to assess the following parameters: GI absorption, Lipinski violations, bioavailability, synthetic accessibility, PAINS/Brenk structural alerts[45,46], BBB permeation, Consensus LogP.

Molecular Dynamics

The top-ranked hit, IMPHY001793, was subjected to a molecular dynamics simulation to assess the stability of its docked pose. The protein-ligand complex was prepared using the Making-it-Rain OpenMM workflow[47,48]. The ligand was built from the canonical SMILES via RDKit[39] and was superimposed onto the docked pose (parameterized at net charge 0). The ligand was parameterized with GAFF2[49], protein with ff14SB[50], solvated in a cuboidal TIP3P water box[51] (12 Å buffer, 55,362 water molecules added), neutralized with 0.15 M NaCl (201 Na^+ / 189 Cl^-). The system's electrostatics were treated with PME (1.0 nm cutoff)[52]. The system was then energy-minimized, and was equilibrated in the NPT ensemble at 298 K, 1 bar (5 ns equilibration). A 100 ns production MD simulation was run in 20 strides of 5 ns (NPT, 298 K, 1 bar, 2 fs timestep as bonds to hydrogen were constrained). Temperature was controlled by a LangevinIntegrator (friction 1.0 ps^{-1}) and pressure by a MonteCarloBarostat. Frames were saved every 10 ps, giving 10,000 total frames. A monomer was simulated in a water box with no lipid bilayer, so the simulation was exploratory and is not used to claim stability in the protein's native environment. The trajectory was analyzed using the following tools: pytraj was used for protein Cα RMSD, ligand-pocket distance to DBP residues,

© The National High School Journal of Science 2026

and a negative control residue set where distances were measured to residues remote from the binding pocket[53]; ProLIF was used for per-residue contact frequencies[54]; MM-GBSA was found by computing with igb = 2, 0.15 M salt, over ~11 evenly spaced frames from a 500-frame subsample, with no configurational entropy term[55].

Ethical Considerations

This study did not involve any animals or humans, and was conducted entirely *in silico*.

Results

PAβN and NMP are known inhibitors of the target protein and were used as benchmarks to rank the Vina scores of the candidates. PAβN produced a score of -8.771 kcal/mol and NMP had one of -7.773 kcal/mol. From the 863 compounds docked, the best score was -11.040 kcal/mol, the weakest, -5.119 kcal/mol, the mean -8.322 kcal/mol, and the standard deviation 0.824 kcal/mol. 656 compounds had stronger Vina scores than NMP though it has a relatively lower bar compared to PAβN. 254 compounds produced stronger scores than PAβN (Figure 1). The top hit was IMPHY001793 (Glabroisoflavanone B), a prenylated isoflavanone from the root of *Glycyrrhiza glabra*, at -11.040. 17 primary hits were produced from the $z \leq -2$ threshold.

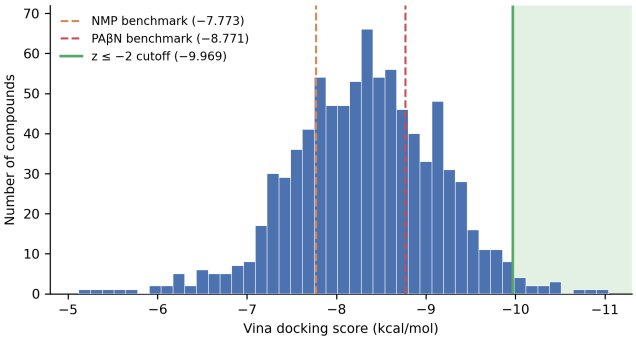


Figure. 1 Distribution of docking scores for the 863 screened phytochemicals. Dashed lines mark the NMP (-7.773 kcal/mol) and PAβN (-8.771 kcal/mol) benchmarks; the green line and shaded region mark the $z \leq -2$ cutoff (-9.969 kcal/mol) defining the 17 primary hits.

Minocycline that was redocked into the model had a binding affinity of -8.95 kcal/mol, the best-pose heavy-atom RMSD being 1.05 Å (was measuring in place without any superposition). The <2.0 Å threshold was set prior to redocking, meaning the result passed the initial criterion. The binding score, RMSD, and atoms in box for four inhibitors whose binding positions were sourced from X-ray crystallography were computed: doxorubicin had a score of -9.09 kcal/mol, RMSD of 2.80 Å and 39/39 atoms in box; MBX2319 had a score of -8.87 kcal/mol, RMSD of 3.41 Å and 22/29 atoms in box; D13-9001 had a score of -8.81 kcal/mol, RMSD of 3.44 Å and 41/49 atoms in box; MBX2931 had a score of -7.98 kcal/mol, RMSD of 3.81 Å and 30/37 atoms in box. The two fully enclosed ligands (doxorubicin, 39/39, and minocycline, 33/33) recovered the best. Box truncation and ligand flexibility (17 torsions for D13-9001, 16 for MBX2931, 11 for doxorubicin) are both plausible explanations for the remainder, and cannot be separated across five compounds.

The initial regression model achieved a scaffold-split R^2 of 0.568 (RMSE 0.564) and external R^2 of 0.408 (RMSE 0.520). The ML classifier used instead was able to separate the two docking-score classes, indicated by a scaffold-split AUC of 0.944, random-split AUC of 0.963 (leakage between the two being +0.019), and external AUC of 0.949. A permutation test suggested that the performance was unlikely to be attributable to chance (0 out of 1000 permutations reached the observed AUC; $P < 0.001$; null mean 0.497, SD 0.035, max 0.612), with shuffled-label models performing at the random baseline (~0.50). The classification was mainly driven by RingCount (+0.0250), LogP (+0.0236), RotBonds (+0.0156) (Figure 2). MW (-0.0011) and HeavyAtoms (+0.0006) did not appear to contribute significantly to classification.

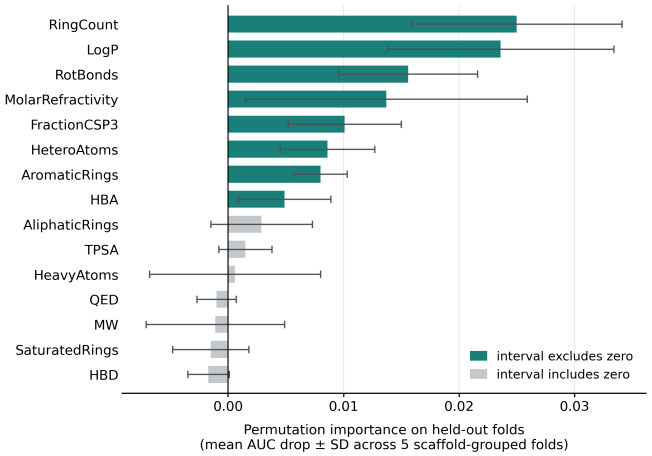


Figure. 2 Held-out permutation importance of the 15 descriptors used by the classifier. Bars show the mean AUC drop across five scaffold-grouped folds, error bars the standard deviation; teal marks intervals excluding zero, grey those including it.

The seven known AcrB actives and 496 property-matched DUD-E decoys that were docked under an identical protocol produced an ROC-AUC of 0.567. The enrichment factors computed were an EF1% of 0.00, EF5% of 0.00, and EF10% of 1.44. Out of the actives, only minocycline is in the top 10%

© The National High School Journal of Science 2026

at rank 41/503, while the rest fall between 115 and 333; PAβN is at 282 and NMP at 333. From the 496 decoys, 52 scored better than -10 kcal/mol. For AcrB, a multi-drug transporter whose innate biological function is to bind to structurally unrelated substrates, these results are not unexpected as complementarity to the pocket carries very little discriminating information. Accordingly, the screen produces a list of prioritized candidates as opposed to acting as a prediction of inhibitory activity.

The redocking performed at three independent random seeds (101, 202, 303) for the top 50 compounds produced a Spearman $\rho = 0.958$ between the seeds. The top 17 Jaccard produced a score between 0.789–0.889. These results suggest that the ranking is stable against the algorithm's own stochastic nature (Table 2).

The 150 compounds spanning the entire docking score range that were docked into five conformers taken from the MD trajectory produced a mean of $\rho = 0.810$ against the rigid receptor (range 0.717–0.846). The inter-conformer agreement produced a mean of $\rho = 0.797$, implying that the rigid model behaves like one more conformation and not like a systematic outlier. Within the top 50, the mean was $\rho = 0.053$, due to the top 50 spanning a score range of 1.47 kcal/mol, as opposed to the conformer-to-conformer variation of 0.41 kcal/mol. The top 50 had a signal-to-noise ratio of 3.1 while the full range had one of 11.9. From the 17 hits, 6 of them were retained in a majority of conformers. The ensemble-mean scoring produced $\rho = 0.882$, the highest out of any measure, and consistent with averaging out conformational noise. The findings show that the receptor's flexibility preserves the separation of favorable from unfavorable compounds, but is unable to accurately order closely scored candidates.

Across the five top-scoring compounds, contact analysis (4 Å cutoff) showed that the residues F178, F610, V612, and F628 were contacted by all five, the G-loop (615–620) was contacted by four of the five, and F617 was contacted by three of five (Table 3). 15 residues were contacted in total, 13 of them being within the DBP and 2 outside (G179, Y327). The predicted binding pose of the top-scoring compound is shown in Figure 3.

ADMET profiling of the 17 compounds of the primary hit set showed that a majority of them had favorable drug-like properties (Table 4). All of them had high GI absorption, none had Lipinski violations (this was explicitly pre-filtered during initial compound selection from IMPPAT), bioavailability was ~0.55 for thirteen and ~0.85 for four. The GI absorption and bioavailability scores, however, were also strongly conditioned by the Egan and Veber filters applied during initial compound selection. No structural alerts were found in 13/17 compounds. The four with alerts were IMPHY000398 (1 PAINS), IMPHY008805 (1 PAINS), IMPHY013011 (1 Brenk), IMPHY013057 (1 Brenk). Additionally, though these parameters

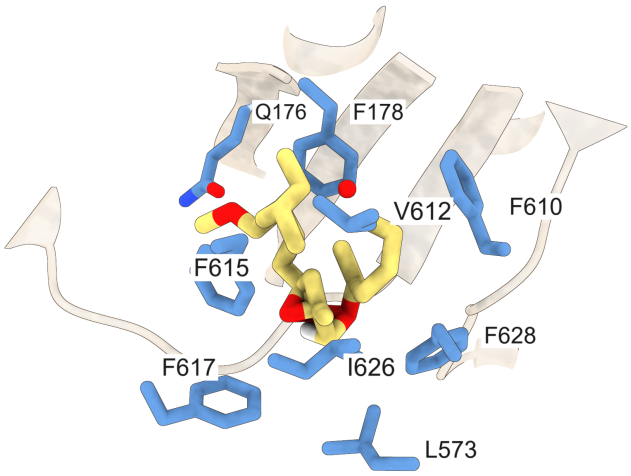


Figure. 3 Predicted binding pose of Glabroisoflavanone B in the AcrB distal binding pocket. The compound is shown as yellow sticks (oxygen, red), the backbone as a tan cartoon, and the nine residues within 4.0 Å as blue sticks, labeled (Table 3).

were not prefiltered, 15 compounds were positive for Blood-Brain Barrier (BBB) permeation, all had a Synthetic Accessibility score from 3.26 to 5.61, and had a consensus LogP within a drug-like range from 1.18 to 4.53.

A 100 ns molecular dynamics simulation was analyzed to ascertain the dynamic stability of the top-ranked compound's docked pose (Figure 4). The ligand remained at 3.91 ± 0.63 Å from the DBP residues, with no dissociation being observed. The residue set (remote from the pocket) held at 17.28 ± 0.86 Å, confirming that the ligand did not drift anywhere else on the protein and stayed in-place. The key phenylalanines (178, 610, 615, 617, 628) were observed at 4.36 ± 0.67 Å. The contact between the DBP and ligand loosens progressively, with DBP going from 3.10 to 4.54 Å, and the key phenylalanines from 3.82 to 5.07 Å when comparing the first and last 20 ns. Moreover, the Protein Cα RMSD rose to 4–4.5 Å and plateaued from around 40 ns onwards. The Cα rise occurs in the same window, going from 3.14 Å at 20–30 ns to 4.28 Å at 40–50 ns. At 30.5 ns, the ligand's largest drift took place, moving 7.53 Å from the five key phenylalanines, and remaining over 6 Å away between 30.4 and 31.8 ns. The ligand then returns to contact distance and the protein plateaus. All five key phenylalanines maintained contact with the ligand through hydrophobic contact, with π-stacking detected at F178 and F615. During the simulation, the DBP only moved 1.2–2.8 Å compared to the global Cα RMSD that reached 4–4.5 Å. The calculated MM-GBSA binding energy was -27.43 ± 3.23 kcal/mol, influenced strongly by van der Waals forces (-36.70); the interaction energy estimate was computed without a configurational entropy term and so is incomparable to

© The National High School Journal of Science 2026

Table 2 Summary of protocol validation and robustness assessments.

Assessment	Metric	Value	Notes
Redocking validation	Minocycline, heavy-atom RMSD (in place, no superposition)	1.05 Å	<2.0 Å threshold set a priori
Pose recovery	Doxorubicin / MBX2319 / D13-9001 / MBX2931	2.80 / 3.41 / 3.44 / 3.81 Å	39/39, 22/29, 41/49, 30/37 crystal atoms inside box
Enrichment	ROC-AUC	0.567	7 actives vs 496 property-matched DUD-E decoys
	EF1% / EF5% / EF10%	0.00 / 0.00 / 1.44	52/496 decoys scored better than −10 kcal/mol
Seed robustness	Spearman ρ between seeds	0.958	seeds 101, 202, 303; top 50 compounds
	Top-17 Jaccard	0.789–0.889	
Conformer ensemble	Spearman ρ, full score range	0.810	150 compounds × 5 MD conformers
	Spearman ρ, top 50 only	0.053	S/N 3.1 vs 11.9 over the full range
	Inter-conformer ρ	0.797	comparable to rigid-receptor agreement
	Primary hits retained	6/17	in a majority of conformers
Machine learning	Scaffold-split AUC	0.944	GroupKFold on 355 Murcko scaffolds
	External test AUC	0.949	71 scaffolds / 176 compounds, evaluated once
	Leakage (random-scaffold)	+0.019	
	Permutation test	$p < 0.001$	0/1000 permutations reached the observed AUC

an experimental ΔG value.

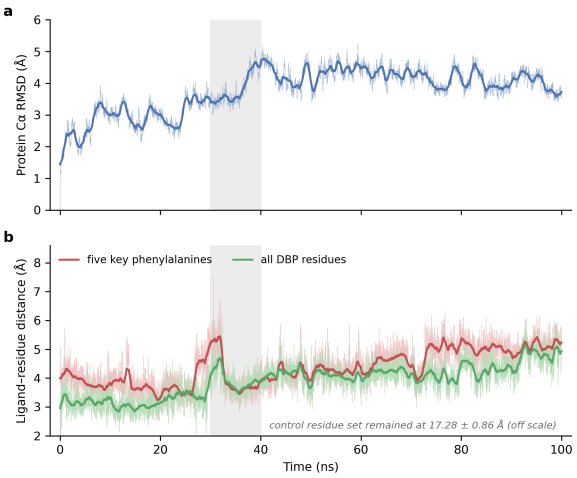


Figure. 4 Molecular dynamics of the Glabroisoflavanone B–AcrB complex over 100 ns. (a) Protein Cα RMSD. (b) Mean ligand distance to all DBP residues (green) and to the five key phenylalanines (red); the shaded band marks 30–40 ns, over which the ligand reseats.

Discussion

The 254 phytochemicals that scored more strongly than PAβN have docking affinities competitive with a known efflux-pump inhibitor, making them candidates that justify experimental validation.

The ML classifier was able to separate the two docking-score classes as seen by the scaffold-split AUC of 0.944 and external set AUC of 0.949 (scaffolds not seen by any cross-validation loop). The random-split comparison produced an AUC of 0.963, meaning leakage from ignoring the scaffold structure was +0.019, which is minimal. The hits differ from the library in a consistent manner, with the average LogP being $+1.09\sigma$, RotBonds being -0.78σ (more rigid), and HBD being -0.49σ; this indicates that, on average, the hits were lipophilic, more rigid, and had fewer hydrogen-bond donors. The lipophilic profile and few hydrogen-bond donors seen among top hits matches the six phenylalanines that form a hydrophobic cage in the binding pocket, which favors lipophilic aromatic ligands and offers few hydrogen-bonding partners. Additionally, the conformational rigidity of the ligand lowers the entropic cost of binding into a defined site, favoring more rigid ligands.

The top three descriptors driving the ML model's classification were RingCount (+0.0250), LogP (+0.0236), RotBonds (+0.0156), and the descriptors with the least contribution to classification were MW (−0.0011) and HeavyAtoms (+0.0006). The results suggest that hits are larger than the average molecule in the library, indicated by an MW of $+0.53\sigma$ and HeavyAtoms of $+0.59\sigma$, despite the removal of size bearing little importance to the model. This is explained by the fact that the ring and refractivity terms already carry the size information, wherein RingCount and MolarRefractivity are seen

© The National High School Journal of Science 2026

Table 3 Binding-site contacts (4 Å heavy-atom cutoff) of the five top-scoring hits.

Residue	In DBP	001793	000398	005217	004477	005683	N hits
PHE136	✔	–	✔	–	✔	✔	3
VAL139	✔	–	✔	✔	–	✔	3
GLN176	✔	✔	–	–	✔	–	2
PHE178	✔	✔	✔	✔	✔	✔	5
GLY179	–	–	–	✔	–	–	1
VAL277	✔	–	✔	✔	–	✔	3
ALA279	✔	–	✔	–	–	✔	2
TYR327	–	–	✔	–	–	–	1
LEU573	✔	✔	✔	–	–	✔	3
PHE610	✔	✔	✔	✔	✔	✔	5
VAL612	✔	✔	✔	✔	✔	✔	5
PHE615	✔	✔	–	–	✔	–	2
PHE617	✔	✔	✔	–	–	✔	3
ILE626	✔	✔	✔	–	–	–	2
PHE628	✔	✔	✔	✔	✔	✔	5

✔ indicates contact (or membership of the distal binding pocket); – indicates no contact. Contact is defined as any receptor residue with at least one heavy atom within 4 Å of any heavy atom of the top-ranked docked pose.

Table 4 Docking affinities and predicted ADMET properties of the 17 primary hits.

Compound	Docking (kcal/mol)	z	MW	Consensus LogP	GI absorption	BBB permeant	PAINS	Brenk	Synth. access.	Bioavailability	Lipinski viol.
IMPHY001793	−11.04	−3.3	352.4	3.37	High	Yes	0	0	4.14	0.55	0
IMPHY000398	−10.9	−3.13	374.3	2.88	High	No	1	0	3.26	0.55	0
IMPHY005217	−10.73	−2.92	368.3	2.07	High	Yes	0	0	4.85	0.55	0
IMPHY004477	−10.43	−2.56	338.4	3.91	High	Yes	0	0	4.15	0.55	0
IMPHY005683	−10.43	−2.56	340.4	3.57	High	Yes	0	0	4.06	0.55	0
IMPHY008805	−10.41	−2.53	282.4	3.78	High	Yes	1	0	4.49	0.55	0
IMPHY013919	−10.35	−2.46	314.4	3.31	High	Yes	0	0	3.34	0.85	0
IMPHY012881	−10.3	−2.4	296.4	2.93	High	Yes	0	0	5.61	0.55	0
IMPHY015068	−10.21	−2.29	354.4	2.62	High	Yes	0	0	4.12	0.55	0
IMPHY009261	−10.18	−2.26	322.4	3.55	High	Yes	0	0	3.94	0.55	0
IMPHY013011	−10.11	−2.17	350.4	3.52	High	Yes	0	1	4.86	0.55	0
IMPHY013920	−10.1	−2.16	300.4	4.28	High	Yes	0	0	3.35	0.85	0
IMPHY010647	−10.04	−2.09	386.4	1.18	High	No	0	0	4.46	0.55	0
IMPHY015025	−10.02	−2.06	336.4	3.75	High	Yes	0	0	4.05	0.55	0
IMPHY005344	−9.986	−2.02	316.4	4.53	High	Yes	0	0	3.64	0.85	0
IMPHY005303	−9.984	−2.02	314.4	3.31	High	Yes	0	0	3.31	0.85	0
IMPHY013057	−9.974	−2.01	324.4	3.72	High	Yes	0	1	4.05	0.55	0

Compounds are the 17 with $z \leq -2$. Zero Lipinski violations is a consequence of the pre-screening filter; PAINS/Brenk alerts, BBB permeation, synthetic accessibility and consensus LogP were not filtered on.

to correlate at 0.60. Thus, permuting MW alone leaves the same size information intact in a different descriptor, so the model never misses it. Similarly, FractionCSP3, which sits around the middle in importance to classification (+0.0101), correlates 0.37 with RingCount. When descriptors are correlated, the permutation importance spreads credit across the group, meaning that individual rankings are less interpretable than group rankings. LogP, on the other hand, is independent of the ring cluster (r = −0.07 to 0.32), showing the model is reading two separate signals and not one bundled property. Furthermore, the largest importance value was a 0.025 AUC drop, suggesting that no single property determines binding. As the importance was measured on only held-out folds, the values are not inflated by the model having already seen the data before.

All five of the top hits contacted F178, F610, V612 and F628, three of the residues being phenylalanines of the DBP cluster. These are consistent with the canonical AcrB inhibitor binding mode, the hydrophobic trap that characterized inhibitors engage[36,56]. F617 was contacted by three of the five, and the G-loop was contacted by four of the five. Of the 15 residues contacted, 13 lie within the DBP, suggesting that the hits engage the pocket itself rather than just the peripheral surface. As per ADMET analysis, most hits are pharmacologically plausible with 13/17 displaying no structural alerts, 15 BBB-permeant, synthetic accessibility between 3.26 and 5.61, and consensus LogP between 1.18 and 4.53. The four that carry alerts (two PAINS and two Brenk) do warrant cau-

© The National High School Journal of Science 2026

tion as these alerts can interfere with downstream assays. The ADMET predictions require experimental confirmation, as the method was purely computational.

Molecular dynamics simulations of the top-ranked candidate show that the ligand stayed in the pocket across the 100 ns, though contact loosened progressively from 3.10 to 4.54 Å between the first and last 20 ns. The elevated protein Cα RMSD (4–4.5 Å) can be attributed to the absence of a lipid bilayer in the simulation, as AcrB is an inner-membrane protein which was simulated in water for this study. Crucially, the pocket itself only moved 1.2–2.8 Å, so the global drift did not propagate into the binding site. MM-GBSA was influenced largely by van der Waals (-36.70 against -14.29 electrostatic); this is consistent with the hydrophobic, aromatic character that was seen in the descriptor analysis earlier. Given that the dynamics simulation began from the docked pose and that the ML classifier was trained on the docking scores, the results produced are not independent lines of evidence despite being mutually consistent. The simulation does not confirm whether the top-hit can inhibit AcrB, but does show that the docked pose is maintained under explicit solvent. ProLIF analysis of the trajectory detected π-stacking at F178 and F615; F178 is the residue reported to form a π–π interaction at the edge of the hydrophobic trap in the inhibitor-bound AcrB structure[35], so the simulation recovers that interaction independently.

In terms of limitations, this study was purely computational. No experimental wet-lab validation was conducted (like ethidium-bromide efflux or MIC-potentiation assays) which means inhibitory ability of any compound cannot be confirmed. The Vina docking scores in the study carry an uncertainty of ~2–3 kcal/mol[34]. The enrichment carried out in the study showed that the scoring function separated known actives from decoys poorly, with an ROC-AUC of 0.567, and EF1% and EF5% both being 0.00. The non-template actives were recovered at 2.80–3.81 Å which is above the conventional 2 Å success threshold, which may be due to box truncation, ligand flexibility, or both. Also, receptor flexibility preserved broad separation (ρ 0.810) but not ordering among near-tied compounds (ρ 0.053; 6/17 retained). The study relied on a homology model rather than a crystal structure, with side-chain rotamers being a particularly weak point (6.42% poor). Another factor is that the molecular dynamics simulation was carried out in a water box with no lipid bilayer, meaning that the run was only exploratory and cannot support any claims about the binding pose in AcrB's native environment. The dynamics was also only conducted for one compound, so the dynamic behavior of the other 16 hits is unexplored. Also, the two benchmark ligands were prepared twice, once for the main screen and once for the benchmark set; given that 3D coordinate generation is stochastic, this resulted in different starting geometry, as seen in PAβN and NMP scoring 0.48 and 0.17 kcal/mol apart between the two preparations. This source of variance is not captured by the seed-robustness analysis, which varied the docking seed rather than the starting conformer.

To conclude, in this study, a computational screening of drug-like IMPPAT phytochemicals against the S. Typhimurium AcrB DBP was done. Each candidate's docking scores were benchmarked against the known inhibitors PAβN and NMP. An ML model was built to identify the characteristics that can differentiate strong and weak docking scores classes. Results show that 254 phytochemicals scored better than PAβN, the top hit being Glabroisoflavanone B (IMPHY001793) at -11.040. ML was able to separate strong and weak docking-score classes (AUC 0.944), driven by ring architecture and lipophilicity. The protocol was validated by redocking a crystallographic ligand to 1.05 Å, and its limits were measured through decoy enrichment, seed robustness and conformer-ensemble docking. The top hits contact the same region (F178, F610, V612 and F628, the phenylalanine cluster) that PAβN also engages[10,56]. 13 out of the 17 top hits had no structural alerts, meaning most were ADMET-favorable. IMPHY001793's binding pose was maintained over 100 ns when measured through molecular dynamics. The study produces a prioritized list of candidate EPIs from an Indian phytochemical source, including an interpretable feature signature. Experimental validation and further computational studies like molecular dynamics on more than one hit could be included in future studies. The results justify wet-lab validation for the prioritized candidates, to further assess their potential as EPIs.

Data and Code Availability

The docking results for all 863 compounds, the receptor model, the machine-learning dataset, the docked poses of the top hits and the molecular dynamics analysis outputs are deposited at Zenodo (https://doi.org/10.5281/zenodo.23356930), together with the scripts that generated them. The scripts are also reproduced in full in the Supplementary Information. The molecular dynamics trajectory files are available from the author on request.

Acknowledgments

I would like to thank Professor Monika Sharma for her constant guidance, helping me improve my methodology and style of writing despite her busy schedule. Furthermore, I am deeply indebted to both my parents for their unwavering emotional, technical, and financial support during this process. I acknowledge that I used Claude for the generation of certain figures, to proof-read writing, and to assist with the writing

© The National High School Journal of Science 2026

and debugging of certain code files. There are no competing interests or external funding sources to disclose.

References

1 GBD 2021 Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. The Lancet. Vol. 404, pg. 1199–1226, 2024, https://doi.org/10.1016/S0140-6736(24)01867-1.

2 I. N. Okeke, M. E. A. de Kraker, T. P. Van Boeckel, C. K. Kumar, H. Schmitt, A. C. Gales, S. Bertagnolio, M. Sharland, R. Laxminarayan. The scope of the antimicrobial resistance challenge. The Lancet. Vol. 403, pg. 2426–2438, 2024, https://doi.org/10.1016/S0140-6736(24)00876-6.

3 L. Huang, C. Wu, H. Gao, C. Xu, M. Dai, L. Huang, H. Hao, X. Wang, G. Cheng. Bacterial multidrug efflux pumps at the frontline of antimicrobial resistance: an overview. Antibiotics. Vol. 11, pg. 520, 2022, https://doi.org/10.3390/antibiotics11040520.

4 J. Kobylka, M. S. Kuth, R. T. Müller, E. R. Geertsma, K. M. Pos. AcrB: a mean, keen, drug efflux machine. Annals of the New York Academy of Sciences. Vol. 1459, pg. 38–68, 2020, https://doi.org/10.1111/nyas.14239.

5 GBD 2017 Non-Typhoidal Salmonella Invasive Disease Collaborators. The global burden of non-typhoidal Salmonella invasive disease: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet Infectious Diseases. Vol. 19, pg. 1312–1324, 2019, https://doi.org/10.1016/S1473-3099(19)30418-9.

6 V. Urdaneta, J. Casadesús. Adaptation of Salmonella enterica to bile: essential role of AcrAB-mediated efflux. Environmental Microbiology. Vol. 20, pg. 1405–1418, 2018, https://doi.org/10.1111/1462-2920.14047.

7 L. J. V. Piddock. Multidrug-resistance efflux pumps — not just for resistance. Nature Reviews Microbiology. Vol. 4, pg. 629–636, 2006, https://doi.org/10.1038/nrmicro1464.

8 O. Lomovskaya, M. S. Warren, A. Lee, J. Galazzo, R. Fronko, M. Lee, J. Blais, D. Cho, S. Chamberland, T. Renau, R. Leger, S. Hecker, W. Watkins, K. Hoshino, H. Ishida, V. J. Lee. Identification and characterization of inhibitors of multidrug resistance efflux pumps in Pseudomonas aeruginosa: novel agents for combination therapy. Antimicrobial Agents and Chemotherapy. Vol. 45, pg. 105–116, 2001, https://doi.org/10.1128/AAC.45.1.105-116.2001.

9 J. A. Bohnert, W. V. Kern. Selected arylpiperazines are capable of reversing multidrug resistance in Escherichia coli overexpressing RND efflux pumps. Antimicrobial Agents and Chemotherapy. Vol. 49, pg. 849–852, 2005, https://doi.org/10.1128/AAC.49.2.849-852.2005.

10 A. V. Vargiu, H. Nikaido. Multidrug binding properties of the AcrB efflux pump characterized by molecular dynamics simulations. Proceedings of the National Academy of Sciences. Vol. 109, pg. 20637–20642, 2012, https://doi.org/10.1073/pnas.1218348109.

11 M. Duffey, R. P. Jumde, R. M. A. da Costa, H. K. Ropponen, B. Blasco, L. J. V. Piddock. Extending the potency and lifespan of antibiotics: inhibitors of Gram-negative bacterial efflux pumps. ACS Infectious Diseases. Vol. 10, pg. 1458–1482, 2024, https://doi.org/10.1021/acsinfecdis.4c00091.

12 M. Stavri, L. J. V. Piddock, S. Gibbons. Bacterial efflux pump inhibitors from natural sources. Journal of Antimicrobial Chemotherapy. Vol. 59, pg. 1247–1260, 2007, https://doi.org/10.1093/jac/dkl460.

13 R. P. Vivek-Ananth, K. Mohanraj, A. K. Sahoo, A. Samal. IMPPAT 2.0: an enhanced and expanded phytochemical atlas of Indian medicinal plants. ACS Omega. Vol. 8, pg. 8827–8845, 2023, https://doi.org/10.1021/acsomega.3c00156.

14 A. S. El-Demerdash, S. A. Kamel, E. Y. T. Elariny, H. Henidi, Y. Mahran, H. Alahdal, A. M. Saleh, R. A. Ibrahim. Natural inhibitors of Salmonella MDR efflux pumps AcrAB and AcrD: an integrated in silico, molecular, and in vitro investigation. International Journal of Molecular Sciences. Vol. 25, pg. 12949, 2024, https://doi.org/10.3390/ijms252312949.

15 O. O. Oyedara, O. A. Fadare, E. Franco-Frías, N. Heredia, S. García. Computational assessment of phytochemicals of medicinal plants from Mexico as potential inhibitors of Salmonella enterica efflux pump AcrB protein. Journal of Biomolecular Structure and Dynamics. Vol. 41, pg. 1776–1789, 2023, https://doi.org/10.1080/07391102.2021.2024261.

16 J. Mehta, R. Rolta, K. Dev. Role of medicinal plants from North Western Himalayas as an efflux pump inhibitor against MDR AcrAB-TolC Salmonella enterica serovar typhimurium: in vitro and in silico studies. Journal of Ethnopharmacology. Vol. 282, pg. 114589, 2021, https://doi.org/10.1016/j.jep.2021.114589.

17 The UniProt Consortium. UniProt: the universal protein knowledgebase in 2025. Nucleic Acids Research. Vol. 53, pg. D609–D617, 2025, https://doi.org/10.1093/nar/gkae1010.

18 R. M. Johnson, C. Fais, M. Parmar, H. Cheruvara, R. L. Marshall, S. J. Hesketh, M. C. Feasey, P. Ruggerone, A. V. Vargiu, V. L. G. Postis, S. P. Muench, V. N. Bavro. Cryo-EM structure and molecular dynamics analysis of the fluoroquinolone resistant mutant of the AcrB transporter from Salmonella. Microorganisms. Vol. 8, pg. 943, 2020, https://doi.org/10.3390/microorganisms8060943.

19 T. Eicher, H. Cha, M. A. Seeger, L. Brandstätter, J. El-Delik, J. A. Bohnert, W. V. Kern, F. Verrey, M. G. Grütter, K. Diederichs, K. M. Pos. Transport of drugs by the multidrug transporter AcrB involves an access and a deep binding pocket that are separated by a switch-loop. Proceedings of the National Academy of Sciences. Vol. 109, pg. 5687–5692, 2012, https://doi.org/10.1073/pnas.1114944109.

20 A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research. Vol. 46, pg. W296–W303, 2018, https://doi.org/10.1093/nar/gky427.

21 G. Studer, C. Rempfer, A. M. Waterhouse, R. Gumienny, J. Haas, T. Schwede. QMEANDisCo—distance constraints applied on model quality estimation. Bioinformatics. Vol. 36, pg. 1765–1771, 2020, https://doi.org/10.1093/bioinformatics/btz828.

22 V. B. Chen, W. B. Arendall III, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica Section D. Vol. 66, pg. 12–21, 2010, https://doi.org/10.1107/S0907444909042073.

23 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison. Open Babel: an open chemical toolbox. Journal of Cheminformatics. Vol. 3, pg. 33, 2011, https://doi.org/10.1186/1758-2946-3-33.

24 J. A. Bohnert, S. Schuster, M. Szymaniak-Vits, W. V. Kern. Determination of real-time efflux phenotypes in Escherichia coli AcrB binding pocket phenylalanine mutants using a 1,2'-dinaphthylamine efflux assay. PLoS ONE. Vol. 6, pg. e21196, 2011, https://doi.org/10.1371/journal.pone.0021196.

25 Schrödinger, LLC. The PyMOL molecular graphics system, version 3.1.6.1. 2026.

26 C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews. Vol. 46, pg. 3–26, 2001, https://doi.org/10.1016/S0169-409X(00)00129-0.

27 D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple. Molecular properties that influence the oral bioavailability of

© The National High School Journal of Science 2026

drug candidates. Journal of Medicinal Chemistry. Vol. 45, pg. 2615–2623, 2002, https://doi.org/10.1021/jm020017n.

28 W. J. Egan, K. M. Merz, J. J. Baldwin. Prediction of drug absorption using multivariate statistics. Journal of Medicinal Chemistry. Vol. 43, pg. 3867–3877, 2000, https://doi.org/10.1021/jm000292e.

29 A. K. Ghose, V. N. Viswanadhan, J. J. Wendoloski. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. Journal of Combinatorial Chemistry. Vol. 1, pg. 55–68, 1999, https://doi.org/10.1021/cc9800071.

30 G. R. Bickerton, G. V. Paolini, J. Besnard, S. Muresan, A. L. Hopkins. Quantifying the chemical beauty of drugs. Nature Chemistry. Vol. 4, pg. 90–98, 2012, https://doi.org/10.1038/nchem.1243.

31 T. A. Halgren. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. Journal of Computational Chemistry. Vol. 17, pg. 490–519, 1996, https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P.

32 S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E. E. Bolton. PubChem 2025 update. Nucleic Acids Research. Vol. 53, pg. D1516–D1525, 2025, https://doi.org/10.1093/nar/gkae1059.

33 J. Eberhardt, D. Santos-Martins, A. F. Tillack, S. Forli. AutoDock Vina 1.2.0: new docking methods, expanded force field, and Python bindings. Journal of Chemical Information and Modeling. Vol. 61, pg. 3891–3898, 2021, https://doi.org/10.1021/acs.jcim.1c00203.

34 O. Trott, A. J. Olson. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry. Vol. 31, pg. 455–461, 2010, https://doi.org/10.1002/jcc.21334.

35 R. Nakashima, K. Sakurai, S. Yamasaki, K. Hayashi, C. Nagata, K. Hoshino, Y. Onodera, K. Nishino, A. Yamaguchi. Structural basis for the inhibition of bacterial multidrug exporters. Nature. Vol. 500, pg. 102–106, 2013, https://doi.org/10.1038/nature12300.

36 H. Sjuts, A. V. Vargiu, S. M. Kwasny, S. T. Nguyen, H. S. Kim, X. Ding, A. R. Ornik, P. Ruggerone, T. L. Bowlin, H. Nikaido, K. M. Pos, T. J. Opperman. Molecular basis for inhibition of AcrB multidrug efflux pump by novel and powerful pyranopyridine derivatives. Proceedings of the National Academy of Sciences. Vol. 113, pg. 3509–3514, 2016, https://doi.org/10.1073/pnas.1602472113.

37 D. Santos-Martins, Y. He, J. Eberhardt, P. Sharma, N. Bruciaferri, M. Holcomb, M. A. Llanos, A. Hansel-Harris, A. P. Barkdull, A. F. Tillack, G. Bianco, M.-L. Paulsen, J. Mato, I. Taneja, S. Forli. Meeko: molecule parametrization and software interoperability for docking and beyond. Journal of Chemical Information and Modeling. Vol. 65, pg. 13045–13050, 2025, https://doi.org/10.1021/acs.jcim.5c02271.

38 M. M. Mysinger, M. Carchia, J. J. Irwin, B. K. Shoichet. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. Journal of Medicinal Chemistry. Vol. 55, pg. 6582–6594, 2012, https://doi.org/10.1021/jm300687e.

39 RDKit: open-source cheminformatics. https://www.rdkit.org, 2026.

40 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay. Scikit-learn: machine learning in Python. Journal of Machine Learning Research. Vol. 12, pg. 2825–2830, 2011.

41 G. W. Bemis, M. A. Murcko. The properties of known drugs. 1. Molecular frameworks. Journal of Medicinal Chemistry. Vol. 39, pg. 2887–2893, 1996, https://doi.org/10.1021/jm9602928.

42 L. Breiman. Random forests. Machine Learning. Vol. 45, pg. 5–32, 2001, https://doi.org/10.1023/A:1010933404324.

43 M. Ojala, G. C. Garriga. Permutation tests for studying classifier performance. Journal of Machine Learning Research. Vol. 11, pg. 1833–1863, 2010.

44 A. Daina, O. Michielin, V. Zoete. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific Reports. Vol. 7, pg. 42717, 2017, https://doi.org/10.1038/srep42717.

45 J. B. Baell, G. A. Holloway. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. Journal of Medicinal Chemistry. Vol. 53, pg. 2719–2740, 2010, https://doi.org/10.1021/jm901137j.

46 R. Brenk, A. Schipani, D. James, A. Krasowski, I. H. Gilbert, J. Frearson, P. G. Wyatt. Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. ChemMedChem. Vol. 3, pg. 435–444, 2008, https://doi.org/10.1002/cmdc.200700139.

47 P. R. Arantes, M. D. Polêto, C. Pedebos, R. Ligabue-Braun. Making it rain: cloud-based molecular simulations for everyone. Journal of Chemical Information and Modeling. Vol. 61, pg. 4852–4856, 2021, https://doi.org/10.1021/acs.jcim.1c00998.

48 P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L. P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, V. S. Pande. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. PLoS Computational Biology. Vol. 13, pg. e1005659, 2017, https://doi.org/10.1371/journal.pcbi.1005659.

49 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case. Development and testing of a general amber force field. Journal of Computational Chemistry. Vol. 25, pg. 1157–1174, 2004, https://doi.org/10.1002/jcc.20035.

50 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. Journal of Chemical Theory and Computation. Vol. 11, pg. 3696–3713, 2015, https://doi.org/10.1021/acs.jctc.5b00255.

51 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein. Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics. Vol. 79, pg. 926–935, 1983, https://doi.org/10.1063/1.445869.

52 T. Darden, D. York, L. Pedersen. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics. Vol. 98, pg. 10089–10092, 1993, https://doi.org/10.1063/1.464397.

53 D. R. Roe, T. E. Cheatham. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. Journal of Chemical Theory and Computation. Vol. 9, pg. 3084–3095, 2013, https://doi.org/10.1021/ct400341p.

54 C. Bouysset, S. Fiorucci. ProLIF: a library to encode molecular interactions as fingerprints. Journal of Cheminformatics. Vol. 13, pg. 72, 2021, https://doi.org/10.1186/s13321-021-00548-6.

55 B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, A. E. Roitberg. MMPBSA.py: an efficient program for end-state free energy calculations. Journal of Chemical Theory and Computation. Vol. 8, pg. 3314–3321, 2012, https://doi.org/10.1021/ct300418h.

56 A. D. Kinana, A. V. Vargiu, T. May, H. Nikaido. Aminoacyl β-naphthylamides as substrates and modulators of AcrB multidrug efflux pump. Proceedings of the National Academy of Sciences. Vol. 113, pg. 1405–1410, 2016, https://doi.org/10.1073/pnas.1525143113.

Supplementary Information

The online version contains supplementary material available at https://nhsjs.com/?p=61275