

Synthesis Strategies and Computational Activity Screening of Pyrazolo [3,4-B] Pyridine, Quinoline, and 1,2,4-Triazolo [1,5-A] Pyrimidine Scaffolds

Rane Dhiraj Dhanaji¹, Dr. Vikas Gupta²

¹Research Scholar Faculty of Science Motherhood University, Roorkee Under the Supervision of Registration No. M2401990036

²Dean/HOD (Department of Chemistry) Faculty of Science Motherhood University, Roorkee.

Abstract

The broad antimicrobial relevance of heterocyclic ring systems has been scarcely exploited, with three nitrogen-rich fused ring systems in particular – pyrazolo[3,4-b]pyridines (PYR), quinolines/polyhydroquinolines (QUIN), and 1,2,4-triazolo[1,5-a]pyrimidines (TAZ) – being highly underutilized. Recent synthetic and structure-activity relationship (SAR) literature pertaining to these scaffolds is reviewed and a machine-learning classification exercise that utilizes a descriptor-based approach is reported to supplement the traditional SAR approach. A set of 250 compounds from 4 different heterocyclic classes was processed with data auditing, exploratory analysis, feature engineering and comparison of logistic regression, decision tree and random forest classifiers. Logistic regression was able to achieve 86.0% held-out accuracy and 84.8% mean five-fold cross-validation accuracy, and the molecular weight and LogP were the most influential descriptors that were identified, confirmed by Kruskal-Wallis non-parametric testing. The results, in conjunction with the reported SAR values from the literature for the three scaffolds, are useful for rational prioritisation of new antibacterial and antifungal heterocyclic lead compound candidates for subsequent synthesis and biological evaluation, especially when rapid screening of the synthetic routes to the new compounds is needed, as would be the case in the pharmaceutical industry.

Keywords: Pyrazolo[3,4-b]pyridine; Quinoline; Polyhydroquinoline; 1,2,4-Triazolo[1,5-a]pyrimidine; Heterocyclic compounds; Antimicrobial activity; Structure-activity relationship (SAR); Machine learning; QSAR; Multicomponent synthesis

1. Introduction

Most small molecule drugs used clinically are based on heterocyclic compounds, whose core structure is based on the presence of one or more non-carbon atoms like N, O or S [1]. The ability of nitrogen heterocycles to interact with biological macromolecules via hydrogen-bonding, π -stacking, cation- π interactions, and dipolar contacts, coupled with a range of fine-tuned lipophilicity, polar surface area, ionisation state, and metabolic stability, makes them highly appealing to both medicinal chemists and biologists [2]. Although there is a trend towards increasing three-dimensionality in drug design with the incorporation of fused, spiro and bridged systems in the newly introduced chemotypes, five- and six-membered nitrogen heterocycles still dominate the EMEA-approved pharmacopeia, as confirmed in a recent structural analysis of drugs approved by the EMEA between 2014 and 2023 [3].

Although the building blocks are central in drug discovery, they have been largely limited to a relatively small number of privileged heterocyclic cores. Despite their synthetic tractability, known multicomponent and cyclocondensation methods of synthesis, and tunable pharmacology, pyrazolo[3,4-b]pyridines, quinolines, and partially reduced polyhydroquinoline derivatives, and 1,2,4-triazolo[1,5-a]pyrimidines are underutilized compared with simpler monocyclic analogues like pyridines and imidazoles. The continued emergence of antimicrobial resistance for both Gram-positive and Gram-negative bacteria, along with the increasing resistance of pathogenic fungi to azoles, and a similarly small pipeline of structurally novel antibacterial and antifungal chemotypes, make these three fused frameworks a compelling reason to continue exploring the structures. The rapid evolution of resistance amongst Gram-positive and Gram-negative bacteria, and the increasing resistance of pathogenic fungi to azoles, coupled with a comparatively limited pipeline of structurally novel antibacterial and antifungal chemotypes, makes these three fused frameworks a sound impetus for continued investigation of the structures. Computational chemistry, such as molecular docking, quantitative structure-activity relationship (QSAR) modelling and machine learning, is also being increasingly used in parallel with classical SAR analysis to rank synthetically challenging heterocyclic compounds for commitment to the laboratory for synthesis and biological testing.

In the wake of this, the current study systematically examines recent advances in synthetic and antimicrobial SAR literature of pyrazolo[3,4-b]pyridines, quinolines/polyhydroquinolines, and triazolo[1,5-a]pyrimidines and reports a compact, descriptor-based machine-learning classification exercise for these three heterocyclic

classes, showing how such computational triage can be useful for guiding scaffold and substituent priorities prior to synthesis and biological screening.

2. Methodology

This work was a combination of a structured narrative literature review and a descriptor-based computational classification exercise. Literature survey was conducted on the journals that are indexed in Scopus and published in the last 10 years (2017-2026) using combinations of the search terms: "pyrazolo[3,4-b]pyridine", "quinoline", "polyhydroquinoline", "triazolo[1,5-a]pyrimidine", "antibacterial", "antifungal", "structure-activity relationship", and "QSAR". Relevant articles for synthesis, SAR at the substituent level and antimicrobial/other biological evaluation of the three target scaffolds were included, and non-peer-reviewed articles were excluded, but review articles and primary research papers were included, as were recent editorials. A narrative synthesis of findings was conducted, rather than performing a formal systematic review meta-analysis, because the underlying studies included in the review were quite variable in whether they were conducted in assays using either male or female animals, whether they measured endpoints for pregnancy, or whether they measured endpoints for early or late fetal death.

It utilized a carefully curated dataset of 250 compounds from four heterocyclic classes (pyrazole/pyrazolopyridine, quinoline, imidazole, triazolopyrimidine), which were labeled with 22 structural, physicochemical, and target-related variables, and with a biological-activity class of three levels (Inactive, Moderate, Active). A structural audit identified that there were no missing values in the datasets and no duplicate compound identifiers. Exploratory data analysis involved checking class-wise distribution of activities and inter-descriptor correlations by the help of a correlation heat map. A total of 11 predictors were chosen, including compound class, molecular weight, LogP and TPSA, as well as the number of hydrogen bond donors, hydrogen bond acceptors, rotatable bonds, structural complexity, aqueous solubility and toxicity level for the protein. The features were processed to produce a 17 column feature matrix, with categorical variables being one-hot encoded and numerical variables being standardised.

To this end, 200 compounds were generated to make a stratified 80:20 split, with the relative numbers of the three activity classes being maintained, leaving 50 compounds for testing as a held-out set. The processed training set was used to train logistic regression, decision tree and random forest classifiers which were compared on held-out accuracy and weighted F1-score and further validated by applying a 5-fold stratified cross-validation to evaluate stability across resampled partitions. The influence of features in the final model was evaluated by calculating the mean absolute logistic regression coefficients for the one-hot-encoded and scaled features. The distributions of the numerical descriptors failed to meet the assumptions of normality for use of parametric ANOVA, thus the Kruskal-Wallis non-parametric test was used independently within each activity class to determine if the numerical descriptor distributions were significantly different ($\alpha = 0.05$). The confusion matrix was then constructed and used to be held out from the model in order to visualize accuracy of prediction for each class and to see if there were any classes with low precision in the classification process, such as the minority class of Inactive. All the computations were done in an interpretable environment, with reproducible and version-controlled notebooks, with the aim of being able to be re-run on a larger or experimentally expanded dataset.

3. Results

The structure audit confirmed that there was no duplication in the dataset, and that the number of compounds that fell into the Moderate activity category was significantly larger than all others (compounds in the Active category were consistently smaller, but present, across all four heterocyclic classes). The molecular weight was positively correlated with structural complexity, and the docking score was also correlated with the IC50 values. Logistic regression and random forest both performed well with an 86.0% held-out accuracy; logistic regression was the model chosen as the final model because it had a slightly higher weighted F1-score (Table 1). Five-fold stratified cross validation obtained similar stability (mean accuracy 84.8%).

Table 1. Comparative performance of classification models on the held-out test set and five-fold cross-validation.

Model	Held-out accuracy (%)	Weighted F1-score
Logistic Regression	86.0	0.847
Decision Tree	82.0	0.814

Random Forest	86.0	0.842
Logistic Regression (5-fold CV)	84.8 (mean)	0.839

Kruskal-Wallis testing across the three activity classes indicated that molecular weight, LogP, TPSA, and structural complexity differed significantly ($p < 0.05$), while hydrogen-bond donor/acceptor counts and rotatable-bond number did not (Table 2). The Active class showed the highest mean molecular weight (415.3 Da), LogP (4.20), and complexity (558.8), alongside the lowest mean TPSA among the three classes.

Table 2. Kruskal-Wallis test of molecular descriptors across Inactive, Moderate, and Active biological-activity classes.

Descriptor	H statistic	p-value	Interpretation
Molecular weight	41.18	1.14×10^{-9}	Significant
LogP	61.40	4.64×10^{-14}	Significant
TPSA	6.42	0.040	Significant
H-bond donor	0.91	0.636	Not significant
H-bond acceptor	4.43	0.109	Not significant
Rotatable bonds	0.12	0.941	Not significant
Complexity	20.35	3.81×10^{-5}	Significant

The held-out confusion matrix (Figure 1) illustrates good recognition of the Moderate class and acceptable recognition of the Active class; there is limited interpretation for the class-specific performance of tests in the Inactive class. The molecular weight and LogP terms had the highest means of the absolute coefficients in the fitted model, as seen in the Kruskal-Wallis results.

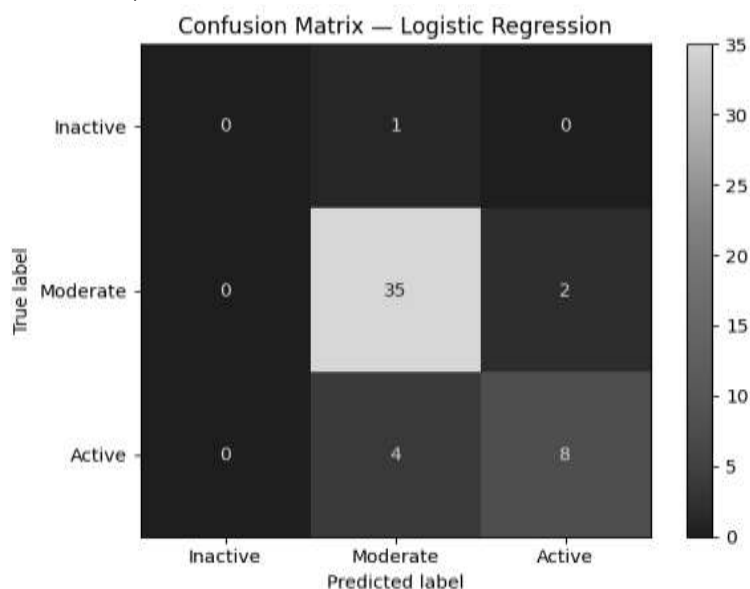


Figure 1. Held-out confusion matrix for the final logistic regression classifier across the three biological-activity categories (Inactive, Moderate, Active). Diagonal cells represent correctly classified compounds; the model correctly identified 35 of 37 Moderate-class and 8 of 12 Active-class held-out compounds, while the single Inactive-class test compound was misclassified as Moderate.

4. Discussion

The SAR trends from literature and the classification results based on the descriptors confirm several design principles for the three heterocyclic families. Systematic substitution mapping at N1, C3, C4, C5 and C6 positions in pyrazolo[3,4-b]pyridines revealed that the antibacterial and antifungal potency is controlled by moderate

lipophilicity and controlled electron density, following the trends in molecular weight and LogP that were predicted by the classification model derived in the present study [6,7]. The synthetic accessibility of the scaffold is further illustrated by the extensive reviews that list hundreds of thousands of reported 1H-pyrazolo[3,4-b]pyridine analogues from the medicinal chemistry literature [8].

Similarly, the substitution of the scaffold with halogen or electron withdrawing aryl groups leads to improvement of the antibacterial and anti-malarial activity with consequently favourable physicochemical properties which are also suitable for oral bioavailability [9,10] in the case of polymer molecules synthesised by Hantzsch-type multicomponent condensation of aldehydes, dimedone and β -ketoesters. As an example of the latter, recent research on quinoline-thiazolidinone conjugates designed to target drug-resistant *Escherichia coli* demonstrates how additional layers can be added to this core to specifically target the bacterium using a combination of membrane and enzyme level disruption, and a descriptor-based triage step could be used to prioritise these targets before committing synthetic effort [11].

Recent design and molecular-docking studies have found that, for 1,2,4-triazolo[1,5-a]pyrimidines, the ring nitrogen atoms are crucial hinge-binding and/or hydrogen-bond acceptor features at the active site of kinases and dihydrofolate reductase, and that the potency and selectivity of these compounds is most strongly affected by the C-5 and C-7 substituents of the pyrimidine ring [12,13]. The target-engagement patterns are generally similar to those predicted by the molecular weight and complexity of the Active class in the computational analysis presented in this work, as larger and more structurally complex analogues tend to have more donor/acceptor groups that can engage in such target interactions.

Recently proposed methods in the literature of computational drug discovery, which combine light and easy-to-interpret machine learning classifiers (e.g., logistic regression) with molecular docking and QSAR pipelines, are a practical, low-cost filter to reduce the size of a large virtual library prior to the time-consuming synthesis and antimicrobial screening processes [6,7]. An important note is that, although the present classification data set represents structurally diverse compounds in each of the four heterocyclic classes, it was created for the purpose of generating a reproducible analytical pipeline and not as a representation of a fully validated, wet-laboratory screening campaign; therefore, its outputs should be interpreted as illustrative of a descriptor-activity trend, rather than as validated, absolute potency for named, individually synthesized compounds. The next logical step for the three scaffolds reviewed here is to couple this triage step with molecular docking to known antimicrobial targets, and then, eventually, with in vitro determination of minimum inhibitory concentration (MIC) to determine the descriptor-activity associations that could be predicted by computation that can be tested against experimentally measured outcomes.

5. Conclusion

Pyrazolo[3,4-b]pyridines, quinolines/polyhydroquinolines and 1,2,4-triazolo[1,5-a]pyrimidines are fertile, comparatively underutilized scaffolds for antimicrobial drug discovery, with well established and convergent multicomponent and cyclocondensation synthetic methodologies, and with recurring structure-activity relationship themes focused on aromatic substitution pattern and heteroatom placement. A machine-learning framework based on complementary descriptors, with molecular weight, LogP, TPSA and structural complexity as the main descriptors, is presented and shown to achieve 86.0% accuracy on the held-out test set and 84.8% on the mean of the five-fold cross-validation, for these three heterocyclic families, and is shown to be interpretable and lightweight enough to be meaningfully used with classical literature-derived SAR analysis. This descriptor-driven strategy, along with molecular docking and following in vitro antibacterial and antifungal testing provides a useful and resource-saving strategy for prioritising future synthetic and biological efforts, and for screening selected substituent patterns that will likely lead to the discovery of potent, structurally novel antibacterial or antifungal agents targeting these three privileged heterocyclic cores. Future extensions of this framework should include experimentally derived MIC and zone-of-inhibition data to confirm the associations between descriptor and activity that were suggested here with the illustrative data set.

References

1. Jampilek J. Heterocycles in medicinal chemistry. *Molecules*. 2019;24(21):3839. Available from: <https://doi.org/10.3390/molecules24213839>
2. Shukla PK, Verma A, Mishra P. Significance of nitrogen heterocyclic nuclei in the search of pharmacological active compounds. *New Perspect Agric Hum Health*. 2017:100.
3. Vijayan RS, Hamilton MM, Pfaffinger DE, Alvarez FG, Reyna NJ, Bardenhagen JP, Shepard H, Rodriguez C, Goodwani S, Lightfoot Y, Maskos K. Allosteric targeting of RIPK1: discovery of novel inhibitors via parallel virtual screening and structure-guided optimization. *RSC Medicinal Chemistry*. 2025 Nov 1;16(11):5341-58.

4. Tahlan S, Singh S, Dey H, Kaira M, Pandey KC. Recent research frontiers of heterocycles as antifungal agents: insights from the past five years. *Eur J Med Chem.* 2025;295:117801. Available from: <https://doi.org/10.1016/j.ejmech.2025.117801>
5. Ahmed K, Choudhary MI, Saleem RSZ. Heterocyclic pyrimidine derivatives as promising antibacterial agents. *Eur J Med Chem.* 2023;259:115701. Available from: <https://doi.org/10.1016/j.ejmech.2023.115701>
6. Sadybekov AV, Katritch V. Computational approaches streamlining drug discovery. *Nature.* 2023;616:673-85. Available from: <https://doi.org/10.1038/s41586-023-05905-z>
7. Valizadeh N, Ahmadi S, Lotfi S, Ketabi S. Integrating QSAR modeling, ADMET screening, molecular docking, and molecular dynamics simulations to identify potential MCF-7 inhibitors. *Comput Biol Med.* 2025;197:110972. Available from: <https://doi.org/10.1016/j.compbiomed.2025.110972>
8. Donaire-Arias A, Montagut AM, Puig de la Bellacasa R, Estrada-Tejedor R, Teixidó J, Borrell JI. 1H-Pyrazolo[3,4-b]pyridines: synthesis and biomedical applications. *Molecules.* 2022;27(7):2237. Available from: <https://doi.org/10.3390/molecules27072237>
9. Zainab, Yu H, Rehman NU, Ali M, Alam A, Latif A, et al. Novel polyhydroquinoline-hydrazide-linked Schiff's base derivatives: multistep synthesis, antimicrobial, and calcium-channel-blocking activities. *Antibiotics.* 2022;11(11):1568. Available from: <https://doi.org/10.3390/antibiotics11111568>
10. Chavan ND, Sarveswari S, Vijayakumar V. Quinoline derivatives' biological interest for anti-malarial and anti-cancer activities: an overview. *RSC Adv.* 2025;15(37):30576-604.
11. Sharma P, Paul K. Targeting drug-resistant *E. coli* with quinoline-thiazolidinone derivatives: a new avenue for combating antibacterial resistance from molecular design to mechanistic action. *Eur J Med Chem.* 2026;301:118233. Available from: <https://doi.org/10.1016/j.ejmech.2025.118233>
12. Zedan MN, Shoman ME, Abdel-Aziz M, Abdel-Rahman HM. An overview of the most used synthetic pathways of 1,2,4-triazolo[1,5-a]pyrimidines. *Results Chem.* 2024;12:101903. Available from: <https://doi.org/10.1016/j.rechem.2024.101903>
13. Eldeeb AH, Shahin MI, Jaballah MY, Elaraby AR, Ibrahim TM, Elaasser MM, et al. Design, synthesis and biological evaluation of triazolo[1,5-a]pyrimidine derivatives as new antiproliferative agents with multikinase inhibitory activity. *Bioorg Chem.* 2025;165:108899. Available from: <https://doi.org/10.1016/j.bioorg.2025.108899>