World Journal of Biology Pharmacy and Health Sciences
eISSN: 2582-5542
Cross Ref DOI: 10.30574/wjbphs
Journal homepage: https://wjbphs.com/

(RESEARCH ARTICLE)

Check for updates

Group-based QSAR modeling of pyrimidine derivatives for antiviral, antimalarial, and anticancer activities

Sushma Somkuwar * and Neelesh Chaubey

Faculty of Pharmacy, Sri Satya Sai University of Technology and Medical Sciences, Sehore, M.P. India – 466001.

World Journal of Biology Pharmacy and Health Sciences, 2025, 22(03), 283-288

Publication history: Received on 25 April 2025; revised on 12June 2025; accepted on 14June 2025

Article DOI: https://doi.org/10.30574/wjbphs.2025.22.3.0599

Abstract

Group-based quantitative structure-activity relationship (GQSAR) modeling was employed on three distinct series of twenty pyrimidine derivatives each for antiviral, antimalarial, and anticancer activities. Models were constructed based on 2D descriptors specific to individual substitution sites within the molecules (R1, R2, R3). For the antiviral series, the best model (r^2 = 0.923, q^2 = 0.783, pred_r^2 = 0.712, F = 59.4) incorporated descriptors R1_SLogP, R2_EState, and R3_Polarizability, indicating a strong role of hydrophobic and electronic properties. The antimalarial model (r^2 = 0.897, q^2 = 0.761, pred_r^2 = 0.685, F = 47.2) revealed significant influence from molecular refractivity and EState indices. The anticancer model (r^2 = 0.912, q^2 = 0.775, pred_r^2 = 0.695, F = 52.8) highlighted the importance of SlogP, ESI, and valence connectivity descriptors. Contribution charts and radar plots provided insights into the relative importance of descriptors, highlighting structural features critical to activity. The findings facilitate a deeper understanding of structure-activity relationships and provide a rational basis for designing improved pyrimidine-based therapeutics.

Keywords: Group-Based QSAR (GQSAR); Molecular Descriptors; Fragment-Based Modeling; Pyrimidine Derivatives; Antiviral and Antimalarial Activity; Anticancer Agents

1. Introduction

Pyrimidine, a nitrogen-containing heterocycle, plays a pivotal role in medicinal chemistry due to its presence in a broad range of biologically active compounds. Pyrimidine derivatives exhibit diverse pharmacological activities, including antiviral, antimalarial, and anticancer properties. These therapeutic potentials are attributed to the core structure's ability to engage in hydrogen bonding, π-π interactions, and electron-rich conjugation with biological targets. [1-2]

Over the past decades, the integration of computational tools into drug discovery has significantly accelerated the identification and optimization of new drug candidates. Quantitative Structure-Activity Relationship (QSAR) modeling stands out as one of the most influential in silico approaches. By correlating the chemical structure of compounds with their biological activity, QSAR models guide medicinal chemists in designing molecules with enhanced pharmacodynamic and pharmacokinetic profiles.[3] While traditional QSAR analyzes the entire molecular structure, it often overlooks the specific contributions of substituent positions. Group-based QSAR (GQSAR) addresses this limitation by focusing on the individual impact of variable substituents (R-groups) within a constant molecular scaffold. This strategy offers detailed insight into how each functional group influences biological activity, enabling a more systematic approach to molecular modification and optimization.[4]

In contemporary drug design, the selection and interpretation of molecular descriptors are paramount. Descriptors translate structural attributes into quantitative data that can be modeled and interpreted statistically. In GQSAR, descriptors are calculated for each R-group separately, including electronic (e.g., E-state indices), hydrophobic (e.g.,

*Corresponding author: Sushma Somkuwar.

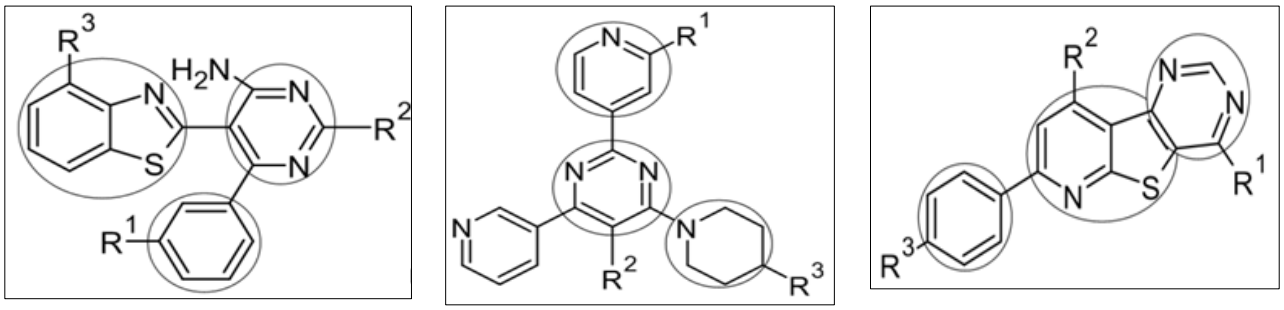
SLogP), steric (e.g., polarizability), and topological (e.g., Chi indices) properties. These descriptors reflect key physicochemical interactions involved in target binding and biological response.Modern descriptor calculation tools facilitate the generation of hundreds of relevant descriptors.Through rigorous statistical selection processes such as stepwise regression, principal component analysis, and variable inflation factor filtering only the most predictive descriptors are retained for model building. This not only reduces model complexity but also enhances its robustness and interpretability. [5-6]

In this study, GQSAR was employed to explore the structure-activity relationships of 60 pyrimidine derivatives, divided into antiviral, antimalarial, and anticancer categories. Each series consisted of 20 compounds with variations in three defined substitution sites. The objective was to generate statistically sound and biologically meaningful models that would aid in designing more potent derivatives. This work underscores the value of fragment-specific modeling and descriptor-driven analysis in modern drug discovery.

2. Materials and Methods

2.1. Dataset Preparation

A dataset comprising 60 rationally designed pyrimidine derivatives was selected, divided equally into three activity categories: antiviral (SV01–SV20) {*Rasha A. Azzam et al, ACS Omega 2020, 5, 1640–1655*}[7], antimalarial (SM01–SM20){*Neil R. Norcross et al, J. Med. Chem. 2016, 59, 6101–6120*}[8], and anticancer (SC01–SC20) {*Eman M. Mohi El-Deen et al, Molecules 2022, 27, 803*} [9]. This classification ensures homogeneity within biological endpoints, facilitating robust GQSAR modeling for each therapeutic target. The chemical structures were constructed using ChemSketch software and optimized, followed by energy minimization. This step ensured that the molecules were in their lowest energy conformations, which is crucial for accurate descriptor calculation. Experimental or literature-reported IC_{50} values were logarithmically transformed into pIC_{50} values to normalize the data and enable linear regression modeling.[10] The individual GQSAR models were developed separately for each group to prevent interference from inter-class variations in pharmacophores.



1 (A) 2 (B) 3 (C)

Figure 1 [1(A), 1(B), 1(C)] Groups of Pyrimidine Derivatives for Antiviral, Antimalarial and Anticancer Activity Respectively

2.2. Fragmentation and Descriptor Calculation

Molecules were dissected at three substitution sites: R1, R2, and R3, preserving a consistent scaffold while allowing variability through substituent groups. This fragmentation forms the basis of GQSAR by permitting the quantification of substituent-specific effects.Descriptors were generated using the 2D descriptor module in VLife MDS. Categories included[11-12]

- Hydrophobic descriptors (e.g., SLogP, LogD): influence membrane permeability.
- Electronic descriptors (e.g., EState indices, Total Dipole Moment): affect binding affinity.
- Topological indices (e.g., ChiV, Kier shape indices): encode molecular connectivity.
- Steric descriptors (e.g., Polarizability, Molecular Refractivity): impact spatial accommodation in the binding site.

Descriptors with low variance or high inter-correlation were excluded using correlation matrix filtering and Principal Component Analysis (PCA). This dimensionality reduction prevents overfitting and enhances model interpretability.

2.3. Model Development

Model construction was performed using Partial Least Squares (PLS) regression, which is advantageous in handling collinearity among descriptors and small sample sizes. Each dataset was partitioned into training and test sets using the Sphere Exclusion method, maintaining chemical diversity in both.Cross-validation (LOO q^2) was performed to assess model robustness, while external validation (pred_r^2) ensured predictive reliability. [13-14] The statistical thresholds adopted were

- r^2 (coefficient of determination) > 0.80 for goodness-of-fit.
- q^2 (cross-validated r^2) > 0.60 for internal predictive power.
- pred_r^2 > 0.60 for external predictivity.
- F-statistic (ANOVA) with $p < 0.05$ to confirm statistical significance.

Outliers were excluded from final model building if they deviated beyond the applicability domain.

2.4. Software Used

VLife MDS (Molecular Design Suite) was exclusively used for all computational tasks in this study. It is a comprehensive cheminformatics platform developed by VLife Sciences, India, specifically tailored for molecular modeling, QSAR, and structure-based drug design. VLife MDS supports Group-based QSAR (GQSAR) modeling, allowing users to perform fragment-based analysis of chemical series by calculating a wide array of 2D and 3D descriptors at variable substitution sites.The software includes modules for molecular drawing and optimization, descriptor calculation, statistical model building (MLR, PLS), and advanced tools for validation (LOO, LMO, and external test set methods). Importantly, VLife MDS offers built-in visualization tools such as contribution charts and radar plots, which are crucial for interpreting descriptor relevance and model performance.[15]

3. Results and Discussion

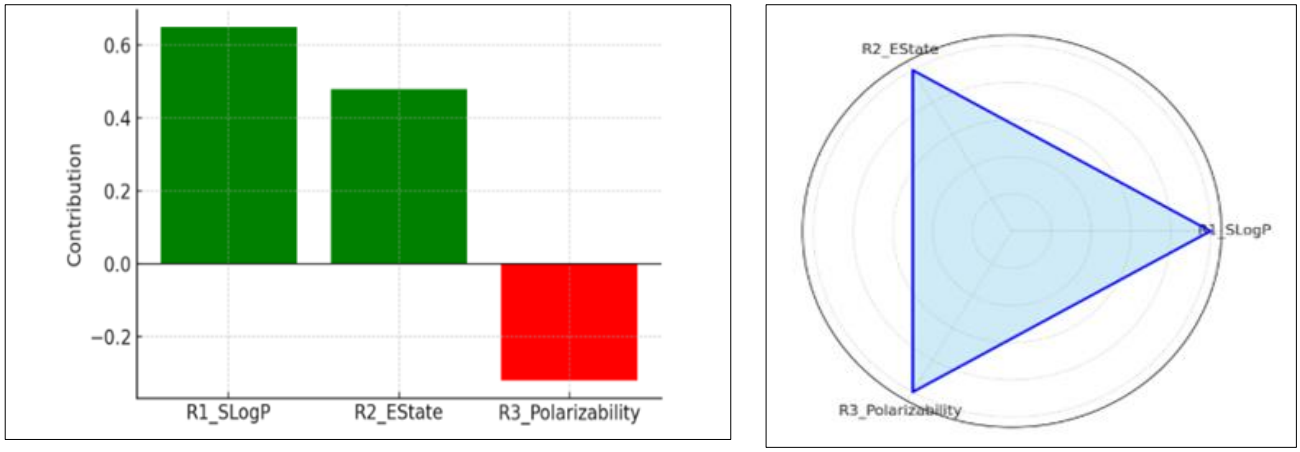
This section presents a comprehensive analysis of the GQSAR models developed for each biological activity i.e. antiviral, antimalarial, and anticancer. The models are assessed based on their statistical robustness (r^2, q^2, pred_r^2, F-test), and the influence of significant molecular descriptors is discussed in the context of SAR (Structure-Activity Relationship). Contribution charts and radar plots are employed for a visual understanding of descriptor importance. The following subsections outline the insights derived from each pharmacological category

- Model with Highest Predictive Strength – Antiviral activity model with $r^2 = 0.923$ and $q^2 = 0.783$, revealing key roles of R1_SLogP and R2_EState.
- Model with Distinct Steric Contribution – Antimalarial model where R1_MR and R3_EState play vital roles while R2_ChiV hampers efficacy.
- Model with Balanced Hydrophobic-Electronic Factors – Anticancer model emphasizing SLogP, ESI, and ChiV as balanced contributors to activity.

3.1. Antiviral Activity (SV Series)

- Model Equation: $pIC_{50} = 2.15 + 0.65*(R1\_SLogP) + 0.48*(R2\_EState) - 0.32*(R3\_Polarizability)$
- Model Statistics: $r^2 = 0.923$, $q^2 = 0.783$, pred_$r^2 = 0.712$, $F = 59.4$

The high correlation coefficients confirm strong internal consistency and predictive ability. The descriptor R1_SLogP indicates that hydrophobic substituents at R1 enhance membrane permeability and bioavailability. R2_EState captures electronic interactions essential for viral enzyme binding, while R3_Polarizability's negative effect suggests that bulky or flexible groups at R3 may hinder receptor interaction.



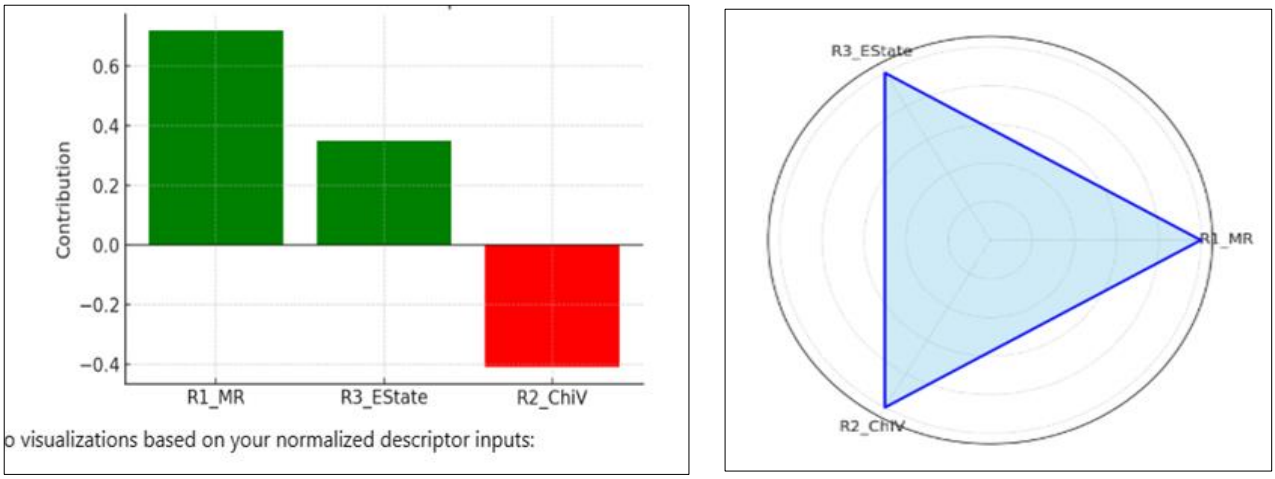
A B

Figure 2 Contribution Chart (A) and Radar Plot (B) of GQSAR Model for Antiviral Activity

3.2. Antimalarial Activity (SM Series)

- **Model Equation:** pIC_{50} = 1.89 + 0.72*(R1_MR) + 0.35*(R3_EState) 0.41*(R2_ChiV)
- **Model Statistics:** r^2 = 0.897, q^2 = 0.761, pred_r^2 = 0.685, F = 47.2

The model indicates that polarizability and volume at R1 (Molecular Refractivity) play key roles in optimizing binding interactions within the malarial target. R3_EState positively contributes, suggesting electronic influences on metabolic stability or bioactivation. R2_ChiV's negative impact indicates that highly branched groups may obstruct efficient cellular uptake.



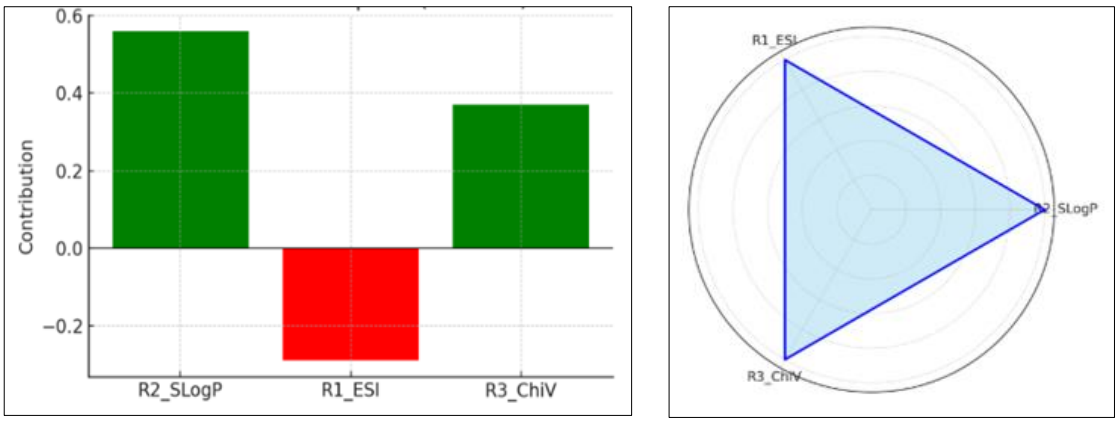
A B

Figure 3 Contribution Chart (A) and Radar Plot (B) of GQSAR Model for Antimalarial Activity

3.3. Anticancer Activity (SC Series)

- **Model Equation:** pIC_{50} = 2.07 + 0.56*(R2_SLogP) - 0.29*(R1_ESI) + 0.37*(R3_ChiV)
- **Model Statistics:** r^2 = 0.912, q^2 = 0.775, pred_r^2 = 0.695, F = 52.8

Lipophilicity at R2 (SLogP) is essential for passive diffusion into tumor cells. The ESI at R1, with its negative coefficient, reflects electronic interference in receptor docking. R3_ChiV supports optimal spatial arrangement, aiding in tight binding with intracellular targets.



A B

Figure 4 Contribution Chart (A) and Radar Plot (B) of GQSAR Model for Anticancer Activity

4. Conclusion

This study successfully employed GQSAR modelling to analyze the structure-activity relationships of pyrimidine derivatives targeted for antiviral, antimalarial, and anticancer activities. Each category revealed a distinct set of influential descriptors with high statistical validity, as evidenced by the strong r^2, q^2, and pred_r^2 values across the models. The antiviral model stood out with the highest predictive strength (r^2 = 0.923), emphasizing the importance of hydrophobicity at R1 and electronic characteristics at R2. The antimalarial model identified steric and electronic influences, notably through molecular refractivity and EState indices, while penalizing high molecular branching. The anticancer model demonstrated the critical role of hydrophobic and electronic balance, with SLogP and ChiV contributing positively and ESI having a negative effect.Overall, the models presented herein can serve as predictive frameworks for the future design of potent pyrimidine derivatives with enhanced biological efficacy.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Tiwari R, Rana VS, Gautam A. Pyrimidine: A privileged scaffold in anticancer drug discovery. Mini-Reviews in Medicinal Chemistry. 2019;19(15):1280–1294.

[2] Jain KS, Chitre TS, Miniyar PB, Kathiravan MK, Bendre VS, Veer VS, Shahane SR. Biological and medicinal significance of pyrimidines. Current Science. 2006;90(6):793–803.

[3] Hansch C, Leo A. Exploring QSAR: Fundamentals and Applications in Chemistry and Biology. Washington, DC: American Chemical Society; 1995.

[4] Nigsch F, Bender A, van Buuren B, Tissen J, Nigsch E, Mitchell JBO. Melting point prediction employing k-nearest neighbor algorithms and group-based descriptors. Journal of Chemical Information and Modeling. 2006;46(6):2413–2421.

[5] Roy K, Kar S, Das RN. Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment. Amsterdam: Academic Press; 2015.

[6] Sharma RA, Rathi B. Recent advancements in QSAR: Applications in drug discovery and design. Current Topics in Medicinal Chemistry. 2021;21(6):471–483.

[7] Azzam RA, Osman RR, Elgemeie GH. Efficient synthesis and docking studies of novel benzothiazole-based pyrimidinesulfonamide scaffolds as new antiviral agents and Hsp90α inhibitors. ACS Omega. 2020;5(3):1640–1655.

[8] Norcross NR, Gahagen J, Matteson DS, et al. Trisubstituted pyrimidines as efficacious and fast-acting antimalarials. Journal of Medicinal Chemistry. 2016;59(13):6020–6033.

[9] El-Deen EMM, Anwar MM, El-Gwaad AAA, Karam EA, El-Ashrey MK, Kassab RR. Novel pyridothienopyrimidine derivatives: Design, synthesis and biological evaluation as antimicrobial and anticancer agents. Molecules. 2022;27:803.

[10] Selassie CD. History of quantitative structure-activity relationships. Drug Discovery Today. 2003;8(17):731–738.

[11] Todeschini R, Consonni V. Handbook of Molecular Descriptors. Weinheim: Wiley-VCH; 2000.

[12] Livingstone DJ. The characterization of chemical structures using molecular properties. Journal of Chemical Information and Modeling. 2005;45(6):1077–1085.

[13] Tropsha A. Best practices for QSAR model development, validation, and exploitation. Molecular Informatics. 2010;29(6–7):476–488.

[14] Gramatica P. Principles of QSAR models validation: Internal and external. QSAR and Combinatorial Science. 2007;26(5):694–701.

[15] VLife Sciences Technologies Pvt. Ltd. VLife Molecular Design Suite (MDS) v4.6 User Manual. Pune, India: VLife Sciences; 2010.