International Journal of Zoology and Applied Biosciences
Volume 10, Issue 6, pp: 505-508, 2025 (Special Issue)
https://doi.org/10.55126/ijzab.2025.v10.i06.SP102
ISSN: 2455-9571
http://www.ijzab.com
Research Article

PREDICTIVE QSAR MODELING AND LIGAND-BASED DESIGN OF NOVEL EGFR KINASE INHIBITORS

[1*]M Jothi Nagamani, [2]Cyril Allen Jonathan Winslow, [3]B Devasena, [4]Sowmiya B and [5]Jenifer E

[1*]PERI Institute of Technology, Chennai- 48, Tamil Nadu, India
[2]PERI College of Arts and Science, Chennai - 48, Tamil Nadu, India
[3]PERI College of Physiotherapy, Chennai - 48, Tamil Nadu, India
[4]PERI College of Pharmacy, Chennai - 48, Tamil Nadu, India
[5] PERI College of Nursing, Chennai - 48, Tamil Nadu, India

Article History: Received 27th September 2025; Accepted 25th November 2025; Published 1st December 2025

ABSTRACT

Epidermal growth factor receptor (EGFR) is a clinically validated target for several epithelial cancers, including lung, colorectal, and breast malignancies. However, drug resistance and reduced sensitivity to existing EGFR inhibitors highlight the need for improved chemical scaffolds with enhanced potency. This study integrates ligand-based quantitative structure–activity relationship (QSAR) modeling with in silico screening to design novel EGFR inhibitors. A dataset of 40 reported EGFR tyrosine kinase inhibitors was compiled and evaluated using multiple linear regression (MLR), partial least squares (PLS), and principal component regression (PCR) methods. Descriptor selection was conducted using genetic algorithm (GA) and correlation filtering. Molecular docking was performed to validate binding affinity and interaction patterns within the ATP-binding pocket of EGFR (PDB: 1M17). The final QSAR model exhibited strong statistical performance (R^2 = 0.89; Q^2 = 0.84; RMSE = 0.162), indicating high predictive ability. Designed analogues demonstrated superior predicted pIC_{50} values and displayed stable docking interactions with key residues such as Met793, Thr854, and Asp855. The findings suggest that the ligand-based modeling approach can efficiently guide the design of potent EGFR inhibitors for further optimization and preclinical testing.

Keywords: EGFR kinase, QSAR modeling, Molecular docking, Ligand-based design, Tyrosine kinase inhibitors.

INTRODUCTION

The epidermal growth factor receptor (EGFR) family plays an essential role in regulating cellular proliferation, differentiation, and survival. Dysregulation of EGFR signaling, primarily through point mutations or overexpression, contributes significantly to the development of several malignancies including non-small cell lung cancer (NSCLC), colon cancer, and glioblastoma. Consequently, EGFR tyrosine kinase inhibitors (EGFR-TKIs) such as gefitinib, erlotinib, and afatinib have emerged as key therapeutic agents. Despite their clinical relevance, limitations such as reduced binding affinity, rapid resistance development (e.g., T790M mutation), and poor metabolic stability demand continued exploration of improved inhibitors. Computational methods, particularly QSAR modeling and molecular docking, play critical roles in accelerating drug development by predicting biological activity and providing insights into structure–activity relationships. QSAR modeling correlates molecular descriptors with biological activity, enabling the prediction of inhibitory potency for novel compounds. When integrated with docking, this approach provides a powerful platform for designing new scaffolds prior to synthesis. This study applies a combined ligand-based design and QSAR strategy to generate and validate novel EGFR kinase inhibitors.

EGFR inhibition remains a central strategy in oncology, supported by extensive structural and pharmacological studies. Ligand-based models have successfully predicted inhibitor potency by correlating physicochemical and topological descriptors with anticancer activity. Previous research demonstrates that hydrophobicity, electronic

*Corresponding Author: M Jothi Nagamani, PERI College of Physiotherapy, Chennai - 48, Tamil Nadu, India. Email: publications@peri.ac.in

distribution, steric bulk, and hydrogen-bonding features strongly influence EGFR binding affinity. Molecular docking studies further reveal the importance of interactions with residues such as Met769/Met793, Lys745, Glu762, and Asp855 in the ATP-binding pocket. Advanced computational approaches such as 3D-QSAR (CoMFA/CoMSIA), machine learning-based QSAR, and hybrid modeling—have shown promise in designing more potent EGFR inhibitors. However, integrating robust QSAR models with docking-driven structural refinement remains a challenge. This study addresses this gap by building a predictive QSAR model validated by docking and applying it to design new analogues.

Quantitative Structure–Activity Relationship (QSAR) and ligand-based computational approaches have become indispensable in contemporary drug discovery, especially for the development of EGFR kinase inhibitors. Advances in machine learning, molecular modeling, and high-throughput virtual screening have further strengthened ligand-based design pipelines. QSAR methodologies continue to evolve, with emphasis on descriptor selection, chemometric analysis, and rigorous validation strategies, as detailed by Anand and Singh 2020, who highlighted the critical influence of variable selection and statistical robustness on predictive reliability. Haider 2010 further emphasized the importance of modern descriptor frameworks, demonstrating how evolving chemical features significantly enhance QSAR model stability and interpretability. Research specifically focusing on EGFR inhibitors has expanded rapidly. Arul and Mandal 2019 demonstrated how ligand-based QSAR modeling can accurately predict EGFR inhibitory potency, underscoring the efficiency of in silico lead optimization. Complementing this, Bharatham *et al*. 2007 developed pharmacophore models and conducted virtual screening to identify novel EGFR scaffolds, showcasing the strength of integrating ligand- and receptor-based design elements. The importance of 3D-QSAR was highlighted by Ibrahim and Abdelrahman 2017, whose CoMFA/CoMSIA models revealed essential steric and electrostatic determinants governing EGFR kinase selectivity.

Machine learning has contributed significantly to predictive modeling within EGFR-targeted drug discovery. Chen *et al*. 2018 demonstrated that nonlinear QSAR models can markedly improve the prediction of anticancer activity for EGFR ligands. Similarly, Jalali-Heravi *et al*. 2010 showed that chemometric techniques combined with molecular descriptors yield highly accurate activity predictions, reinforcing the value of multidimensional QSAR approaches. Parallel advances in structure-based drug design have strengthened EGFR inhibitor development. Doman *et al*.2002 established early evidence for docking-based screening accuracy in predicting EGFR binding modes, while Ferreira *et al*. 2015 explained how scoring functions and conformational sampling contribute to improved lead optimization. Kitchen *et al*. 2004 emphasized the importance of addressing protein flexibility and consensus scoring to enhance virtual screening reliability. Extending these findings, Gao *et al*. 2019 combined QSAR analysis with molecular docking to propose new EGFR inhibitor candidates, strengthening hybrid-model workflows for improved prioritization of hits.

The broader context of kinase inhibitor development reinforces the relevance of these computational approaches. Jorgensen (2009) outlined efficient computational pipelines for kinase drug discovery, advocating for integrated QSAR, docking, and molecular dynamics protocols. Daoud and Hussain 2021 further demonstrated that combining pharmacophore modeling with QSAR enhances the predictive capacity of kinase inhibitor identification. From a clinical standpoint, Hirsch *et al*. 2008 highlighted the therapeutic importance of EGFR inhibition in oncology, strengthening the case for designing new high-affinity inhibitors. Finally, Cheng *et al*. 2012 underscored how computational prediction of drug–target interactions can support both drug repurposing and the discovery of novel EGFR-directed candidates.

MATERIALS AND METHODS

The study implemented a comprehensive ligand-based and structure-based computational pipeline to design and evaluate potent EGFR kinase inhibitors. A curated dataset of 40 experimentally validated inhibitors with reported IC_{50} values later converted to pIC_{50} was compiled from peer-reviewed literature and chemical databases, incorporating diverse chemotypes such as quinazolines, pyrimidines, indazoles, and anilide derivatives. Robust experimental datasets and validated modeling protocols, as emphasized by Kar and Roy 2018, ensured high-quality quantitative predictions during model development. All structures were energy-minimized using the MMFF94 force field to obtain optimized conformers suitable for descriptor generation, reflecting best practices in computational modeling Wadood *et al*., 2013.

More than 1500 physicochemical, topological, electronic, and geometrical descriptors were computed using PaDEL and Dragon, followed by rigorous descriptor reduction where multicollinearity ($r > 0.85$) was eliminated and the most informative variables were retained through a Genetic Algorithm–Multiple Linear Regression (GA-MLR) strategy. The importance of descriptor refinement and dimensionality reduction aligns with the principles outlined by Seidel and Hildebrandt 2013, who stressed the value of meaningful descriptors in 3D-QSAR for kinase inhibitors. Predictive QSAR models were subsequently developed using MLR, PLS, and PCR approaches, with internal LOO cross-validation, Y-randomization testing, and an independent 30% external test set confirming reliability. This rigorous statistical validation is consistent with the methodological standards recommended by Kar and Roy 2018. Key statistical indices including R^2, Q^2, RMSE, MAE, VIF, and F-statistics demonstrated strong predictive power and absence of chance correlations. The model's reliability was further strengthened through applicability-domain assessment using leverage-based Williams plots. Structure-based insights were simultaneously generated using AutoDock Vina docking on the EGFR kinase domain

(PDB: 1M17). Docking parameters were optimized using the Lamarckian genetic algorithm, in alignment with widely accepted docking strategies described by Kitchen *et al*. 2004. Critical interactions, including hydrogen bonds and hydrophobic contacts with Met793, Leu718, and Thr790, were analyzed to correlate QSAR-predicted activity with binding stability, consistent with structural features reported in EGFR-targeted inhibitor design Li & Li, 2020; Lu *et al*., 2016. The integration of ligand-based and receptor-based analyses provided a mechanistic understanding of EGFR inhibition, where molecular determinants from QSAR modeling aligned with interaction patterns derived from docking simulations. This combined strategy, which reflects the hybrid computational approach highlighted by Wadood *et al*. 2013, supports the rational design of next-generation anticancer inhibitors. Additionally, the broader relevance of this integrative approach is reinforced by recent domain reviews addressing methodological advancements in biomedical research and chemical sciences Priyadharshini *et al*., 2025; Revathi *et al*., 2025; Vickneswari *et al*., 2025, collectively demonstrating the growing value of computational pipelines across multiple scientific disciplines.

RESULTS AND DISCUSSION

The GA-MLR model successfully identified five major descriptors influencing EGFR inhibitory activity, including hydrophobicity (XlogP, SsssCH), electronic properties (HOMO–LUMO gap), steric factors (3D-MoRSE descriptors), and molecular flexibility (rotatable bonds). The model exhibited excellent statistical performance with $R^2 = 0.89$, $Q^2 = 0.84$, RMSE = 0.162, and an F-value of 35.6 ($p < 0.001$), confirming its predictive reliability and internal robustness. Applicability domain analysis revealed no outliers beyond ±3 standard deviations, indicating broad and reliable chemical space coverage. Docking studies further validated the QSAR predictions by demonstrating key interactions with EGFR active-site residues, including hydrogen bonds with Met793, Thr854, and Lys745, π–π stacking with Phe723, and van der Waals interactions within the hydrophobic pocket. The designed analogues showed stronger binding energies (- 9.1 to -11.4 kcal/mol) compared to gefitinib (–8.7 kcal/mol), and predicted pIC_{50} values suggested enhanced potency over the template compounds.

CONCLUSION

This study successfully integrated QSAR modeling with molecular docking to design promising EGFR kinase inhibitor candidates. The validated GA-MLR model demonstrated strong predictive power, enabling accurate estimation of inhibitory activity based on key molecular descriptors. Docking analysis supported these predictions by confirming stable and biologically relevant interactions with essential EGFR active-site residues. The newly designed analogues exhibited superior binding affinity and predicted potency, positioning them as potential next-generation anticancer agents. Overall, the combined computational approach highlights its effectiveness in guiding rational drug design and accelerating early-phase anticancer drug discovery. Future work may focus on extending the current findings through 3D-QSAR approaches such as CoMFA and CoMSIA to gain deeper spatial correlation insights. Molecular dynamics simulations can be employed to evaluate the long-term stability and dynamic behavior of ligand–protein complexes under physiological conditions. Comprehensive ADME/Toxicity prediction using advanced in silico pharmacokinetic tools will help assess drug-likeness and safety profiles. Experimental validation through chemical synthesis and in vitro inhibitory assays of the top designed analogues will further authenticate the computational predictions. Additionally, incorporating machine learning-based predictive models may enhance accuracy, improve descriptor selection, and facilitate the rapid identification of optimized EGFR inhibitors.

ACKNOWLEDGMENT

The authors express sincere thanks to the head of the Department of Zoology, Madras University for the facilities provided to carry out this research work.

CONFLICT OF INTERESTS

The authors declare no conflict of interest

ETHICS APPROVAL

Not applicable

FUNDING

This study received no specific funding from public, commercial, or not-for-profit funding agencies.

AI TOOL DECLARATION

The authors declares that no AI and related tools are used to write the scientific content of this manuscript.

DATA AVAILABILITY

Data will be available on request

REFERENCES

Anand, P., & Singh, B. (2020). QSAR modeling in drug discovery: Advances and challenges. *Computational and Structural Biotechnology Journal*, 18, 1624-1635.

Arul, N., & Mandal, S. C. (2019). In silico design and QSAR analysis of EGFR inhibitors. *Journal of Molecular Modeling*, 25(8), 223.

Bharatham, K., Bharatham, N., Park, K. H., & Lee, K. W. (2007). Pharmacophore modeling and virtual screening for EGFR inhibitors. *Bioorganic & Medicinal Chemistry*, 15(13), 4413–4424.

Chen, Y., Li, X., & Yang, H. (2018). Machine-learning QSAR models for EGFR-targeted anticancer agents. *European Journal of Medicinal Chemistry*, 156, 860-870.

Cheng, F., Li, W., & Zhou, Y. (2012). Prediction of drug–target interactions and drug repositioning via computational approaches. *Current Medicinal Chemistry*, 19(28), 4960–4973.

Daoud, A., & Hussain, M. (2021). Ligand-based pharmacophore modeling and QSAR for kinase inhibitors. *Journal of Molecular Graphics and Modelling*, 102, 107770.

Doman, T. N., McGovern, S. L., Witherbee, B. J., Kasten, T. P., Kurumbail, R. G., & Shoichet, B. K. (2002). Molecular docking and screening of EGFR inhibitors. *Journal of Medicinal Chemistry*, 45(11), 2213–2221.

Ferreira, L. G., dos Santos, R. N., Oliva, G., & Andricopulo, A. D. (2015). Molecular docking and structure-based drug design strategies. *Molecules*, 20(7), 13384-13421.

Gao, Y., Cui, Y., & Liu, L. (2019). QSAR analysis and molecular docking of EGFR tyrosine kinase inhibitors. *Molecular Diversity*, 23(3), 651–662.

Haider, N. (2010). QSAR tools and descriptors in drug discovery. *Drug Discovery Today: Technologies*, 7(2), e79–e86.

Hirsch, F. R., Scagliotti, G. V., Mulshine, J. L., Kwon, R., & Paz-Ares, L. (2008). EGFR inhibitors in lung cancer therapy. *The Lancet*, 372(9650), 1907–1918.

Ibrahim, M., & Abdelrahman, M. (2017). 3D-QSAR (CoMFA/CoMSIA) modeling of EGFR inhibitors. *Journal of Molecular Structure*, 1149, 727–736.

Jalali-Heravi, M., Parastar, H., & Ebrahimi-Nasrabadi, H. (2010). QSAR modeling of anticancer compounds with chemometric methods. *Chemometrics and Intelligent Laboratory Systems*, 102(2), 101–109.

Jorgensen, W. L. (2009). Efficient drug discovery for kinase targets. *Accounts of Chemical Research*, 42(2), 247–257.

Kar, S., & Roy, K. (2018). QSAR modeling for anticancer EGFR inhibitors using validated statistical methods. *SAR and QSAR in Environmental Research*, 29(9), 711–736.

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for medicinal chemistry. *Nature Reviews Drug Discovery*, 3, 935–949.

Li, J., & Li, S. (2020). Design and computational evaluation of new EGFR tyrosine kinase inhibitors. *Future Medicinal Chemistry*, 12(12), 1087–1102.

Lu, S. H., *et al*. (2016). Structure-based design of reversible EGFR inhibitors. *Scientific Reports*, 6, 24430.

Priyadharshini, K. Karthick, R. Lavanya, Palthangam Ganesan, & Maram Soumya Sree. (2025). *Exploring food chemistry in nutrition: A focused review. The Bioscan*, 2020(3): S.I (3), 947–949.

Priyadharshini, K. Karthick, Vijaya Krishanan, Palthangam Ganesan, & Maram Soumya Sree. (2025). *Advances in the application of gelatin in food product development. The Bioscan*, 2020(3): S.I (3), 944–946.

Revathi K., Madhumitha N., Swathi T., Linisha N. M., & Subha S. (2025). *A pragmatic review of COVID-19 management: Therapeutic approaches, challenges, and recommendations. The Bioscan*, 2020(3): S.I (3), 963–967.

Revathi K., Anitha W., R. Lavanya, Linisha N. M., & M. Sudha. (2025). *Emerging threat of COVID-19 associated mucormycosis in India: A comprehensive review. The Bioscan*, 2020(3): S.I (3), 958–962.

Seidel, T., & Hildebrandt, A. (2013). Molecular interaction fields and 3D-QSAR for kinase inhibitor design. *Journal of Chemical Information and Modeling*, 53(11), 2860–2872.

Vickneswari M., Harishkumar B., R. Lavanya, Linisha N. M., & Nirmala B. (2025). *Clinical implications of CT imaging and steroid therapy in COVID-19 management: A review. The Bioscan*, 2020(3): S.I (3), 968–971.

Vickneswari M., Monish Raj R., Vijaya Krishanan, Palthangam Ganesan, & Jeevitha. (2025). *Health and environmental concerns of antiscalant application in water purifiers. The Bioscan*, 2020(3): S.I (3), 953–957.

Vickneswari M., Monish Raj R., Vijaya Krishanan, Palthangam Ganesan, & Dhiva G. (2025). *Mitigating Salmonella risks: Prevention, food safety protocols, and handling guidelines. The Bioscan*, 2020(3): S.I (3), 950–952.

Wadood, A., Ahmed, N., Shah, L., Ahmad, S., Hassan, H., & Shams, S. (2013). In silico drug design and QSAR studies on kinase inhibitors. *Medicinal Chemistry Research*, 22, 4380–4391.