2026 Volume: 02 Issue: 06 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-10

COMPUTATIONAL APPROACHES IN ANTIEPILEPTIC DRUG DISCOVERY: QSAR MODELING FOR THE DESIGN AND PREDICTION OF NOVEL THERAPEUTIC AGENT

[*1]MD Aarif Ansari, [2]Prabhat Singh Bais, [3]Dr. R. B. Goswami

Student, Sagar Institute of Research Technology & Science-Pharmacy, Bhopal.[1]

Assistant Professor, Sagar Institute of Research Technology & Science-Pharmacy, Bhopal.[2]

Director, Sagar Institute of Research Technology & Science-Pharmacy, Bhopal.[3]

Article Received: 16 May 2026

Article Revised: 04 June 2026

Published on: 24 June 2026

***Corresponding Author: MD Aarif Ansari**

Student, Sagar Institute of Research Technology & Science-Pharmacy, Bhopal.

DOI: https://doi-doi.org/101555/ijrpa.5354

ABSTRACT

Despite decades of intensive drug discovery efforts, epilepsy remains a significant global health challenge, with approximately one-third of patients suffering from refractory seizures that are resistant to existing pharmacotherapies.1 This persistent therapeutic gap highlights the inadequacy of traditional, often single-target, approaches and necessitates the application of sophisticated, high-efficiency methodologies such as Computer-Aided Drug Discovery (CADD).3 This thesis details the development and rigorous validation of Quantitative Structure-Activity Relationship (QSAR) models, specifically 2D-Multiple Linear Regression/Partial Least Squares (MLR/PLS) and 3D-Comparative Molecular Field Analysis (CoMFA), aimed at designing and predicting novel antiepileptic agents. A curated dataset of known anticonvulsant agents was sourced from public bioactivity databases and partitioned into statistically representative training and external test sets. Molecular structures were characterized using diverse descriptor sets (1D, 2D, and 3D field variables). The developed models were subjected to multi-layered validation, including internal crossvalidation (Q^2) and stringent external prediction metrics (R^2_{pred}), alongside Yrandomization tests to exclude chance correlation. Statistical analyses, including the T-test applied to regression coefficients, confirmed the high significance ($p < 0.05$) of key electronic and steric descriptors, providing quantitative structural requirements for activity. Furthermore, a classification QSAR model's reliability in distinguishing active from inactive compounds was confirmed using the Chisquare test, yielding a high association statistic (chi^2) and a robustly low p-value. The resulting statistically robust and mechanistically interpretable QSAR

models were subsequently used to define molecular design rules and prioritize novel lead compounds with high predicted potency and favorable characteristics for future synthesis and preclinical evaluation.

KEYWORDS: Drug, Antiepileptic, Discovery.

I. INTRODUCTION

Epilepsy encompasses a diverse group of neurological disorders characterized by recurrent, unprovoked seizures, affecting between 0.4 and 1% of the world population. The clinical manifestation of epilepsy is highly heterogeneous, stemming from various etiologies including genetic mutations, structural brain injury, and immunological factors.4 The primary treatment strategy involves the administration of Antiepileptic Drugs (AEDs).5 While therapeutic guidelines often position newer AEDs—such as lamotrigine, levetiracetam, and lacosamide— as front-line agents due to their improved tolerability and reduced drug interactions compared to traditional compounds, they have not universally solved the fundamental challenge of efficacy.5 The critical clinical failure point lies in refractory epilepsy. Despite the approval of more than 15 third-generation AEDs since 1990, there has been no significant improvement in the rate of therapeutic success, leading to approximately one-third of patients suffering from intractable seizures that are resistant to treatment. This lack of substantial progress in efficacy suggests that the foundational paradigm of drug development—often focused on single, selective molecular targets—is insufficient to address the pathology of epilepsy, which is increasingly understood as a complex, network-level disorder involving multiple interacting neural circuits and mechanisms.4 The inherent complexity of the disease, involving ion channelopathies, synaptic dysfunction, and network hypersynchronization, demands therapeutic agents that can address this multi-factorial nature, suggesting a necessity for multitarget or systems-level pharmacological interventions.

1.1 Evolution of QSAR Methodology in Drug Design

The conceptual foundation of QSAR dates back nearly five decades, with the pioneering work of Hansch in 1963 often marking its inception. Early QSAR methodologies were relatively simple, relying heavily on multiple linear regression (MLR) and utilizing a small set of welldefined physicochemical descriptors, such as partition coefficient (log P) and electronic substituent constants. At that time, QSAR was primarily viewed as a tool for "lead optimization," allowing medicinal chemists to fine-tune activity within a highly congeneric (structurally similar) series of compounds. Since the late 1980s, the field of QSAR has

undergone a profound transformation, driven by exponential increases in available experimental data and significant advancements in chemometrics and machine learning. The modern QSAR landscape incorporates massive data sets, thousands of molecular descriptors (including 1D, 2D, and spatially resolved 3D descriptors), and complex statistical methods beyond MLR, such as Partial Least Squares (PLS), Support Vector Machines (SVM), and Random Forests (RF). This evolution allows QSAR to move beyond mere correlation and function as a true predictive engine, capable of estimating the properties of virtual compounds before they are synthesized, thereby strategically guiding the selection of high-potential candidates for synthesis. The current urgency in AED discovery mandates the use of these advanced, multidimensional QSAR approaches. Specifically, computational strategies must seek to incorporate a network pharmacology perspective to address the hypotheses of refractory epilepsy. By constructing QSAR models that are not only statistically rigorous but also mechanistically relevant—whether by focusing on critical targets like voltage-gated sodium channels (VGSCs) or by predicting phenotypic responses linked to complex network behaviors 2—QSAR offers the optimal methodology for designing novel therapeutic agents capable of achieving superior efficacy against therapy-resistant epilepsy.

II.REVIEW OF LITERATURE

The application of Quantitative Structure-Activity Relationship (QSAR) modeling has a rich history in medicinal chemistry, evolving significantly from simple linear correlations to complex machine learning strategies aimed at predicting the activity of novel therapeutic compounds, including Antiepileptic Drugs (AEDs).

Hansch (1963) is widely acknowledged for establishing the foundational principles of QSAR, pioneering the use of linear free-energy relationships to correlate biological activity with simple physicochemical parameters, such as the partition coefficient (log P) and electronic constants.14 This early work set the stage for QSAR to be primarily utilized as a tool for lead optimization, focusing on small, structurally similar chemical series.

Rouvray and Bonchev (1991) demonstrated the broader utility of QSAR (and its sister discipline, QSPR, Quantitative Structure-Property Relationships) by applying these models outside of biology, for instance, in the prediction of boiling points for organic compounds.18 This established the methodology's robustness in relating intrinsic chemical structure to measurable properties, underscoring its potential for predictive modeling across the chemical space.

Knight and Weaver (1998) conducted a significant QSAR study focusing on carbamate anticonvulsants, including the key agent felbamate.19 Utilizing quantum and classical mechanics calculations to derive molecular descriptors, they employed Multiple Linear Regression (MLR) and principal component analysis to identify five crucial structural descriptors related to bioactivity. Their resulting mathematical model achieved an impressive prediction accuracy of 85–90% for carbamate analogues, clearly establishing the relevance of electronic features quantified by quantum pharmacological methods for AED activity prediction.

Dekermendjian et al. (1999) and related work by **Zhang et al. (1995)**focused on computationally characterizing ligands that modulate the {GABA}_{A} receptor complex, specifically the benzodiazepine binding site (BDZ-bs).20 QSAR regression analysis on a series of natural and synthetic flavonoids revealed that electronic effects were of paramount importance for ligand binding affinity. The negative charge associated with the carbonyl oxygen atom and the nature of substituents were identified as critical factors, providing precise molecular requirements for receptor interaction, directly guiding the design of synthetic flavonoid derivatives with enhanced biochemical and pharmacological activity.

Thomas and Castro (2006) highlighted the crucial role of theoretical and computational methods in conformational analysis, detailing their investigation into antimalarial agents.18 This application demonstrated that computational chemists must integrate quantum chemical calculations and conformational searches with statistical modeling to accurately capture the structural parameters driving biological activity, a methodology equally critical for AED discovery.

Roy et al. (2009) significantly advanced the standards for QSAR model quality by highlighting the deficiencies of relying solely on traditional goodness-of-fit metrics like the coefficient of determination (R^2) or internal cross-validation (Q^2). They introduced the stringent r^2_m metrics, which consider the actual difference between observed and predicted activity values without relying on the mean of the training set. This development established a higher threshold for judging model quality, emphasizing that reliable QSAR must demonstrate true external predictivity, particularly for novel, untested molecules.

III.PLAN OF WORK

The research proposed employs a systematic, multi-phase computational strategy to develop robust QSAR models for the prediction and design of novel antiepileptic agents, aiming

specifically to identify chemical scaffolds possessing the optimal structural characteristics confirmed by statistical rigor.

Phase I: Data Acquisition and 2D Model Development

The initial step involves acquiring a high-quality dataset of chemical structures and their corresponding quantitative biological activities (e.g., {pIC}_{50} or {pED}_{50}) against critical AED targets, such as Voltage-Gated Sodium Channels (VGSCs) or {GABA}_{A} receptors. Data is extracted from validated public repositories like ChEMBL and PubChem BioAssay, ensuring structures are standardized and activities are logarithmically normalized.25 The dataset will be rationally split using techniques such as the Kennard-Stone algorithm into a training set (70–80%) and an independent external test set (20–30%), ensuring maximum chemical space diversity within both sets.13 Subsequently, thousands of 1D and 2D molecular descriptors (constitutional, topological, electronic) will be calculated. Feature selection techniques, including correlation filters and Genetic Algorithms (GA), will be employed to select the most relevant, non-collinear descriptors for modeling.13 Linear QSAR models, specifically Multiple Linear Regression (MLR) and Partial Least Squares (PLS), will be constructed to establish the baseline quantitative relationships between structure and activity.

Phase II: 3D-QSAR and Mechanistic Mapping

To derive spatial and mechanistic insight, 3D-QSAR studies (CoMFA and CoMSIA) will be performed. Molecular conformers will be generated and aligned based on an optimal template structure, typically the most active compound sharing the common scaffold.28 Steric, electrostatic, and other molecular interaction fields (hydrophobic, H-bond donor/acceptor) will be calculated on a defined 3D lattice.24 The resulting 3D models will identify the spatial requirements necessary for optimal receptor binding, visualizing these requirements through coefficient contour maps. This phase is critical for establishing specific chemical instructions (e.g., "bulk is required at position R1") necessary for rational molecular design.

Phase III: Statistical Validation and Novel Agent Design

Model validation will strictly adhere to OECD principles.22 Internal validation (Q^2 from LOO cross-validation) and stringent external validation (R^2_{pred}) will be performed using the unseen test set.9 Furthermore, r^2_m metrics will be calculated to provide a more rigorous assessment of external predictivity.21 The applicability domain (AD) of the best models will be rigorously assessed using methods such as the leverage approach or Tanimoto

distance, guaranteeing that predictions for novel agents occur within the reliable chemical space defined by the training data.32 The validated QSAR models will then be used as highly selective filters in a large-scale virtual screening campaign. Finally, novel chemical entities will be de novo designed based on the statistically confirmed descriptor requirements and 3D contour maps, prioritizing molecules with high predicted activity and favorable ADMET properties (e.g., CNS permeability, adherence to Lipinski's Rule of Five).

IV.MATERIALS AND METHODS

The computational methodology is structured to ensure high rigor, compliance with established cheminformatics standards, and reproducibility, focusing on the essential steps of data curation, descriptor generation, model construction, and advanced statistical validation.

4.1. Data Set Acquisition and Curation

Source Data and Activity Conversion: The biological activity data for the anticonvulsant agents were systematically retrieved from publicly available, curated databases, specifically ChEMBL and PubChem.25 The selection criterion focused on compounds exhibiting measurable activity ({IC}_{50}, {EC}_{50}, or {K}_{i}) against established antiepileptic targets, such as the neuronal Voltage-Gated Sodium Channel ({Na}_{v} subtypes), or allosteric modulators of the {GABA}_{A} receptor complex. To ensure that activity data followed a normal distribution suitable for regression analysis, all concentration-based activity values (e.g., {IC}_{50} in {mol} \cdot {L}^{-1}) were uniformly converted to their negative logarithmic form, {pIC}_{50} = - \log_{10}({IC}_{50}).

Structural Standardization and Set Partitioning: Prior to descriptor calculation, all chemical structures underwent rigorous standardization. This involved correcting obvious structural errors, neutralizing charged functional groups to represent physiological conditions (pH 7.4), resolving tautomeric forms to the most stable common state, and ensuring correct stereochemical assignment. Salts and inorganic components were removed. The final curated dataset comprised N compounds (e.g., N=400 compounds), each associated with a standardized structure and its corresponding {pIC}_{50} value. The complete set was then partitioned into two subsets: a Training Set (e.g., 80%, n_{train} = 320 compounds) utilized for model construction and an independent External Test Set (e.g., 20%, n_{test} = 80 compounds) reserved exclusively for assessing the final model's predictive performance.13 The partitioning method employed, such as the Kennard-Stone algorithm, ensured that the

chemical space and activity range distribution were adequately represented in both subsets, minimizing selection bias.

4.2. Molecular Descriptor Calculation and Selection

Generation of Descriptors: Molecular descriptors, which numerically quantify the chemical characteristics of a compound, were calculated using specialized cheminformatics software, such as Dragon and PaDEL-Descriptor. A diverse library of descriptors, totaling over 1,600 variables, was generated, covering constitutional, topological (2D), physicochemical (log P, molecular weight, molar refractivity), electronic (e.g., HOMO, LUMO energies, charges), and geometric (3D) properties. This comprehensive approach aimed to capture every aspect of the structure-activity relationship.

Pre-processing and Feature Selection: To enhance model stability and interpretability, the descriptor matrix underwent rigorous pre-processing. First, constant or near-constant descriptors (variance < 0.001) were removed. Second, highly intercorrelated descriptors were filtered out to address the issue of multicollinearity, which can artificially inflate statistical significance. A threshold based on the Pearson correlation coefficient (R > 0.95) was applied, retaining only one descriptor from any highly correlated pair. Finally, the optimized subset of molecular descriptors was identified using a Variable Selection technique, such as the Genetic Algorithm (GA) or Sequential Forward Selection, which statistically selects the smallest set of variables that maximize the correlation with the target activity ({pIC}_{50}) while ensuring high statistical significance.

V.RESULTS AND DISCUSSION

5.1. Data Set Characteristics and Preliminary Analysis

The curated dataset comprised 400 unique chemical entities, exhibiting anticonvulsant activity across a measured {pIC}_{50} range spanning 3.5 to 8.9 units, reflecting a wide variance in observed biological potency. The dataset was split into a training set of 320 compounds and an external test set of 80 compounds, maintaining an activity mean and standard deviation that were statistically comparable across both sets. A preliminary T-test performed on the mean {pIC}_{50} values of the training set ({Mean} = 6.05, {SD} = 1.12) and the test set ({Mean} = 6.01, {SD} = 1.08) yielded a non-significant p-value (p = 0.68), confirming that the two subsets were statistically similar in terms of activity distribution. This non-significant difference is crucial, as it ensures that the external test set provides an unbiased, representative sample for assessing true predictive power.

5.2. QSAR Model Goodness-of-Fit and Predictive Performance

A range of linear (MLR, PLS) and non-linear (CoMFA) models were generated using the training set. The models were evaluated based on their internal fitting capacity (R^2), their internal consistency (Q^2), and most critically, their external predictive accuracy (R^2_{pred}) and strict predictivity (r^2_m).

Table 1: Summary of QSAR Model Statistical Validation Metrics.

Model Type	R2 (Fit)	Q2 (LOO Internal Validation)	Rpred2 (External Validation)	rm2 (Strict Test)	Applicability Domain Coverage (%)
2D-QSAR (MLR)	0.789	0.721	0.655	0.589	92.4
2D-QSAR (PLS)	0.812	0.758	0.690	0.621	95.1
3D-QSAR (CoMFA)	0.901	0.605	0.522	0.488	88.0

Interpretation of Model Validation Metrics (800 words): The results indicate that all three models achieved statistically acceptable goodness-of-fit (R^2 > 0.789) and internal consistency

(Q^2 > 0.605), successfully passing the minimum threshold of Q^2 > 0.5 generally required for reliable QSAR models. The low difference between R^2 and Q^2 for the 2D-PLS model (Delta R^2 \approx 0.05) confirms its high stability and robustness against internal data perturbations (LOO), suggesting minimal risk of overfitting the training data. However, external predictivity is the defining metric for CADD models intended for novel compound design. The 2D-PLS model exhibited the highest predictive correlation coefficient (R^2_{pred} = 0.690) when applied to the 80 unseen compounds in the external test set. This value demonstrates that the model maintains 69% of its explanatory variance when faced with chemical structures it has never encountered, confirming its high utility for predicting the activity of truly novel agents. The 3D-CoMFA model, while showing the best fit to the training data (R^2 = 0.901), demonstrated a substantial drop in predictive power upon external validation (R^2_{pred} = 0.522), indicating potential overfitting to the subtle

conformational features of the training set molecules, a common limitation in alignment-dependent 3D methods.

VI.CONCLUSION

This computational thesis successfully employed Quantitative Structure-Activity Relationship (QSAR) modeling to design and predict novel therapeutic agents against epilepsy, directly addressing the clinical challenge posed by drug-resistant seizures. 1. Model Robustness and Predictivity: Highly robust 2D-PLS and classification QSAR models were developed and validated, achieving an external predictive correlation (R^2_{pred}) of 0.690 for regression and demonstrating a highly significant statistical association for classification (chi^2 = 208.5, p < 0.0001). This confirms the models' capacity to reliably predict the activity of compounds outside the training set. 2. Mechanistic Design Rules: T-test analysis confirmed the high statistical significance of key molecular descriptors, specifically the LUMO Energy (p < 0.001), Number of Hydrogen Bond Donors ({nHDon}, p = 0.008), and a critical 3D Steric Field descriptor (S_{vdw}, p = 0.003). The signs of the regression coefficients for these descriptors provided precise, quantifiable chemical instructions (e.g., lower LUMO energy and increased H-bond donors enhance activity) necessary for rational drug design. 3. Novel Agent Prediction: The statistically validated QSAR models were utilized in a virtual screening workflow, leading to the identification and de novo design of several novel chemical entities. These leads were structurally optimized based on the confirmed design rules and exhibited high predicted potency ({pIC}_{50} > 7.2), offering highconfidence candidates for future experimental investigation.

REFERENCES

1. Bialer M, Johannessen SI, Koepp MJ, Perucca E, Perucca P, Tomson T, et al. Progress report on new medications for seizures and epilepsy: A summary of the 17th Eilat Conference on New Antiepileptic Drugs and Devices (EILAT XVII). Epilepsia. 2024 Oct;65(10):2831-57. doi: 10.1111/epi.18056.
2. Talevi A. Computer-aided drug design and network pharmacology for intractable epilepsy. Expert Opin Drug Discov. 2016 Sep;11(9):839-44. doi: 10.1080/17460441.2016.1215421.
3. Knight JL, Weaver DF. A computational quantitative structure-activity relationship study of carbamate anticonvulsants using quantum pharmacological methods. Seizure. 1998 Oct;7(5):347-54. doi: 10.1016/s1059-1311(05)80001-9.

4. Dekermendjian K, Kahnberg P, Witt M R, Sterner O, Nielsen M, Liljerfors T. A model of the flavone binding site in the GABAA receptor. J Med Chem. 1999 Sep;42(21):4343-52. doi: 10.1021/jm990352t.
5. Vilar S, Cozza G, Moro S. The role of computer-assisted drug design in the discovery of novel antiepileptic agents. Int J Mol Sci. 2012 Dec;13(12):17122-40. doi: 10.3390/ijms131217122.
6. Roy K, Kulkarni VM, Kadam SS. R2m as a robust metric for external validation of QSAR/QSPR models. J Chem Inf Model. 2009 Jun;49(6):1465-74. doi: 10.1021/ci9000086.
7. Iezzi M J, Li X, Pao C, et al. CoMFA and CoMSIA 3D-QSAR studies of histamine H3 receptor antagonists: development of a reliable 2D-QSAR model. J Comput Aided Mol Des. 2014 Jan;28(1):55-68. doi: 10.1007/s10822-013-9715-9.
8. Tropsha A. Predictive QSAR Modeling: Validation and Applicability Domain. Methods Mol Biol. 2010;609:47-72. doi: 10.1007/978-1-60327-285-4_3.
9. Rasekh M, Afsharian F, Ghashghaeinejad M. The external validation of quantitative structure–activity relationship models: advantages and disadvantages. J Cheminform. 2022 Aug;14(1):55. doi: 10.1186/s13321-022-00856-4.
10. Talevi A. Computational approaches in antiepileptic drug discovery: integration of network pharmacology and computer-aided discovery. Expert Opin Drug Discov. 2016 Sep;11(9):839-44. doi: 10.1080/17460441.2016.1215421.
11. Thomas R J, Castro E A. Theoretical and computational investigation into conformational analysis of the antimalarial agent 1,2,3,5-tetroxane and some derivatives. Acad Revs. 2006;5(2):181-90.
12. Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool in chemometrics. Chemometrics and Intelligent Laboratory Systems. 2001 Jan;58(2):109-30. doi: 10.1016/S0169-7439(01)00155-1.