

Research Article

Development of OECD-Compliant Multiple 2D/3D-QSAR Models Integrated with ADMET, Molecular Docking, and Molecular Dynamics for Prioritizing Phytocompounds as COX-2 Inhibitors

Yahyea Baktiar Laskar^{1,2*}, Pranab Behari Mazumder¹, and Nachimuthu Senthil Kumar²

¹Natural Product and Biomedicine Research Laboratory, Department of Biotechnology, Assam University, India

²Department of Biotechnology, Mizoram University, India

***Corresponding author**

Yahyea Baktiar Laskar, Department of Biotechnology, Mizoram University, India, Tel: 8812895227

Submitted: 25 May 2026

Accepted: 05 September 2026

Published: 07 September 2026

ISSN: 2379-089X

Copyright

© 2026 Laskar YB, et al.

OPEN ACCESS**Keywords**

- Cyclooxygenase-2
- Phytochemical
- QSAR
- Molecular Docking
- Molecular Dynamics

Abstract

Cyclooxygenase-2 (COX-2) is an inducible enzyme linked to inflammation and several diseases, including cancer and neurodegenerative disorders. Current synthetic COX-2 inhibitors carry serious risks of cardiovascular, gastrointestinal, hepatic, and renal issues. This study designed and validated four different predictive 2D- and 3D-QSAR models to assess the COX-2 inhibitory potential of chemicals, using selected molecular descriptors, in accordance with OECD guidelines. These models were used to screen a library of 54 phytocompounds and their analogues, and the results were further validated using ADMET screening, rodent toxicity prediction, molecular docking, and molecular dynamics simulations. The top-performing 2D models had $R^2=0.75$, $Q^2=0.63$, and $\text{pred}_R^2=0.51-0.53$, while the 3D models had $R^2=0.64$, $Q^2=0.55-0.69$, and $\text{pred}_R^2=0.58-0.67$, respectively. Of the four QSAR-prioritized compounds, apigenin-7-sulfate and genistein emerged as promising, toxicologically safe leads, with favorable binding affinity and stability within the COX-2 active site, compared to the well-known NSAID, celecoxib. Integrating multi-model QSAR, in silico pharmacology, and interaction analyses, this study provides insight into plant-based alternatives to COX-2 inhibitors and underscores the need for rigorous experimental/clinical validation in bioassays.

INTRODUCTION

Cyclooxygenase (COX), also known as prostaglandin H2 synthase, is the key rate-limiting enzyme involved in the biosynthesis of prostanoids from arachidonic acid [1]. At least two distinct isoforms of COX have been characterized, COX-1 and COX-2, encoded by the PTGS1 and PTGS2 genes, respectively. While COX-1 is typically expressed constitutively and functions as a 'house-keeping' enzyme, maintaining physiological homeostasis, COX-2 expression is context-dependent and may be constitutive or inducible, depending on tissue type. Notably, COX-2 is upregulated by pro-inflammatory cytokines and mediators produced by various immune cells during inflammatory responses [2]. In both acute and chronic inflammation, the induced expression of COX-2 plays a pivotal role in the elevated production of prostanoids, which are key contributors to the pathogenesis of numerous inflammatory

disorders, including cancer [3,4], rheumatoid arthritis, neurodegenerative diseases like Alzheimer's disease [5], Parkinson's disease [6], epilepsy [7], and various respiratory conditions [8].

The current therapeutic strategies for inhibiting or modulating COX-2 expression primarily comprise the use of selective COX-2 inhibitors or Coxibs (such as celecoxib, rofecoxib, meloxicam, and etoricoxib) and other non-steroidal anti-inflammatory drugs (NSAIDs) [9,10]. These agents act by blocking the enzymatic activity or gene expression of COX-2, thereby reducing the synthesis of pro-inflammatory prostanoids. The Coxibs, such as celecoxib and etoricoxib, have shown efficacy in managing inflammation-associated conditions and in controlling symptoms of various human diseases [11]. Additionally, emerging approaches targeting upstream signaling pathways, such as NF- κ B, MAPKs, and cytokine-mediated

transcriptional regulation, also offer promising avenues for more specific COX-2 modulation. However, despite their therapeutic benefits, long-term or high-dose use of these drugs has been associated with adverse effects, including increased risk of cardiovascular events [12], renal impairment, and gastrointestinal complications [10-13]. Consequently, the pursuit of effective and selective COX-2 inhibitors remains an active area of research, driven by the need to enhance therapeutic efficacy while minimizing adverse effects.

Naturally occurring compounds, particularly those derived from plant sources, continue to serve as a fundamental reservoir of bioactive molecules that are frequently explored for their therapeutic potential against a wide range of human diseases. Owing to their abundant availability, cost-effectiveness, and generally low toxicity, these compounds have either directly contributed to the development of numerous clinically-approved drugs or provided structural templates for the design of novel pharmacological agents [14-16]. Thus, in the present study, we designed and validated several predictive 2D- and 3D-QSAR models to evaluate the COX-2 inhibitory potential of chemicals, in accordance with the OECD guidelines of QSAR models. The newly developed models were then used to screen a small library of phytocompounds with known anti-inflammatory activity in bioassays to assess their efficiency as COX-2 inhibitors and to prioritize their experimental/clinical validation.

MATERIALS & METHODS

Biological and Experimental Data:

The experimental IC_{50} values for 50 known agonists, antagonists, and inhibitors of COX-2 were retrieved from the BindingDB [17] follow the methodology adopted in our previous work [18] on designing QSAR models to evaluate the bioactivity of plant-derived compounds. The selected compound series (**Supplementary Table 1**) exhibits substantial structural diversity and a broad range of biological activity, with IC_{50} values spanning 0.57-10,000 nM, thereby providing a suitable dataset for developing robust QSAR models. To minimize numerical disparity and enhance model reliability, the biological activity values (IC_{50}) or the dependent variable in this study were transformed into their logarithmic form [$\log_{10}(IC_{50})$].

Chemical Diversity and Scaffold Analysis

To evaluate the structural diversity in the compound dataset used in QSAR model development, the Tanimoto similarity-based clustering and Bemis-Murcko scaffold analysis were performed [19]. The molecular structure

files were prepared in SDF format and processed using RDKit (version 25.3.2, <https://www.rdkit.org/>) in Python. The extended-connectivity fingerprints (ECFP4) of 1024 bits were generated for each compound using the MorganGenerator function with a radius of 2. Pairwise Tanimoto similarity scores were computed for all compound pairs, and the resulting similarity matrix was visualized using a heatmap and a histogram to assess overall distribution. To explore compound clustering based on structural similarity, hierarchical clustering was performed using average linkage on a Tanimoto distance matrix (1 – similarity), and a dendrogram was constructed. Additionally, a fingerprint-based cluster map was generated to provide a 2D visual overview of compound grouping. For scaffold-level analysis, Bemis-Murcko scaffolds were extracted from each molecule to identify common chemotypes. Unique scaffolds were counted, and their frequencies were tabulated to assess the diversity of core frameworks within the dataset. These analyses were conducted in a Linux environment using matplotlib, seaborn, SciPy, pandas, and NumPy packages.

Molecular Modelling and Descriptor Calculation

Further computational analyses were conducted on a personal workstation powered by an Intel Core i5 8th Generation CPU @ 2.50 GHz, running 64-bit Windows 11. The 3D structures of the compounds were retrieved from the NCBI PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>) and energy-minimized using the Merck Molecular Force Field (MMFF) with MMFF charges to eliminate steric clashes and optimize geometry. The dielectric environment was prepared using a distance-dependent dielectric function with a constant value of 1, simulating a solvent-like effect on molecular interactions. The minimization process was executed with a maximum of 10,000 cycles and a convergence criterion set to an RMS gradient of 0.01 kcal/mol/Å. The analytical gradient type was employed for accurate energy calculations. The non-bonded interaction parameters were defined with cutoffs of 20.0 Å for electrostatic interactions and 10.0 Å for van der Waals (vdW) interactions, with vdW parameters updated every 10 steps to ensure consistent molecular stability during optimization.

The VLifeMDS package computes 99 atom-type count descriptors, 239 physicochemical (23 classes) descriptors, and more than 700 alignment-independent molecular descriptors to estimate the physicochemical and topological features of a molecule; these descriptors were used as independent variables in constructing a 2D model. To generate 3D field descriptors, the Gasteiger-Marsili method was used to assign partial atomic charges

to the molecules. A distance-dependent dielectric function with a dielectric constant of 1.0 was used to simulate the molecular environment. A carbon atom with a unit charge (+1.0) was selected as the probe to evaluate interaction energies across the molecular grid. The energy cutoffs were set at 10 kcal.mol⁻¹ for electrostatic and 30 kcal.mol⁻¹ for steric interactions to limit extreme values and ensure meaningful descriptor generation. The grid parameters, including grid spacing and dimensions, were defined as detailed in **Supplementary Table 2**, enabling the calculation of field values at regularly spaced intervals around the aligned molecular structures.

Training, Test, and External Validation Datasets

The training and test datasets were prepared according to the previously described method [18], with minor modifications. The dataset of compounds was divided into training, test, and validation sets using the sphere exclusion method, ensuring structural diversity and representative sampling across all subsets. The division was done in the ratio of 70:20:10, respectively, with 70% of the compounds used to build the QSAR model (training set), 20% for evaluating its predictive performance (test set), and the remaining 10% reserved for external validation. The division of compounds into respective datasets was performed iteratively, with several rounds of reassignment and analysis carried out to optimize statistical quality.

To ensure consistency and uniform distribution of biological activity across the subsets, each iteration was evaluated using UniColumn statistics, which calculated the mean, minimum, maximum, and standard deviation of the dependent variable and compared them between the training and test sets. This approach ensured that the training and test sets were comparable and free from significant bias in activity distribution. Further, the model development steps were also repeated until OECD-compliant performance metrics, in terms of coefficient of determination (R^2), cross-validated correlation coefficient (Q^2), external predictive power (pred_R^2), and Z-scores, were achieved. Only when both descriptor-based activity distributions and the model's performance indicators reached satisfactory levels was the final dataset division accepted.

2D-QSAR Model Development

Multiple linear regression (MLR) method: The MLR method, also referred to as Ordinary Least Squares (OLS) regression, was employed to establish a linear relationship between the dependent and independent variables using the least-squares curve-fitting approach [20]. The models were developed using the stepwise

forward-backward (SW-FB) variable selection method, and a cross-correlation threshold of 0.5 was applied to exclude highly intercorrelated variables and ensure model robustness. The descriptor inclusion and exclusion were guided by R^2 values and F-statistic thresholds set at 4.0 for inclusion and 3.99 for exclusion, respectively. A variance cutoff of 0.5 was used to eliminate descriptors with no variation. All descriptor values were auto-scaled before model building to standardize the data and eliminate scale bias. Internal validation was performed using 25 groups in cross-validation, and the model's stability was further assessed through 10 random iterations to confirm reproducibility and predictive reliability. The model generated a regression curve from which actual versus predicted bioactivities of the compounds were derived, based on the general regression equation defined for the MLR method, as shown previously [18].

Partial least squares (PLS) method: The MLR-derived model was further cross-validated using PLS regression, with identical input parameters. The PLS, a modified regression technique, is particularly advantageous when the number of molecular descriptors exceeds the number of data points, a situation in which traditional least-squares regression becomes less reliable. This method models the dependent variable matrix (Y) and the independent variable matrix (X) by extracting latent variables that maximize the covariance between them. The resulting regression equation incorporates these latent variables, which exhibit a significant correlation with the observed biological activity, thereby enhancing the model's predictive reliability and robustness [18].

3D-QSAR Model Development

Principal component regression (PCR) analysis: The PCR is an effective dimensionality reduction method in QSAR studies, particularly useful for addressing multicollinearity and outliers in datasets comprising numerous structural descriptors. In PCR, highly correlated descriptors are transformed into a new set of uncorrelated variables, known as principal components (PCs), thereby reducing the dataset's dimensionality while retaining the maximum variance. Each PC captures a distinct portion of the variance among the compounds and can be independently interpreted. This transformation facilitates the analysis of structure-activity relationships and helps identify structurally deviant compounds. The optimal subset of compounds contributing to model reliability can be further refined through stepwise regression [18]. The parameters used for PCR model development were identical to those applied in MLR and PLS models, with model building based on cross-validation as the primary criterion.

κ -Nearest neighbour molecular field (κ NN-MFA) analysis: The κ NN-MFA method, a non-parametric, distance-based learning approach, classifies data points based on their proximity to their ' κ ' nearest neighbours in the training set. The κ -Nearest Neighbour (κ NN) models were developed using three variable selection methods: SW-FB, simulated annealing (SA), and genetic algorithm (GA). For SW-FB, model optimization was guided by a cross-correlation limit (CCL) of 0.5, term selection based on Q^2 values and F-statistic thresholds of 4.0 (inclusion) and 3.99 (exclusion), a 0.5 variance cutoff, and auto-scaling. The number of neighbours was varied from 2 to 5, and predictions were made using a distance-based weighted-average approach. For the SA method, parameters included a CCL of 0.5, maximum and minimum temperatures of 100 and 0.01, respectively, a perturbation limit of 1, a temperature decrement step of 10, five iterations per temperature level, and a seed value of 0. Variable selection was based on Q^2 , with a variance cutoff of 2.0 and auto-scaling applied. The GA method employed a population size of 10, a chromosome length of 4, a crossover probability of 0.95, a mutation probability of 0.05, and a maximum of 1000 generations, with convergence criteria set to 0.01 and an early-stopping threshold of 5 generations. The CCL was set to 0.5, the variance cutoff to 2.0, and auto-scaling was used. In all three methods, descriptor selection was driven by Q^2 , and model predictions were computed using a distance-based weighted average.

Model Validation:

Internal validation: The generated 2D and 3D QSAR models were assessed for robustness and predictive performance. Internal stability was evaluated using the widely accepted leave-one-out (LOO) cross-validation method, wherein the cross-validated correlation coefficient (Q^2) was calculated. In this approach, each compound in the dataset was systematically excluded once, and the remaining compounds were divided into two equal subsets to construct a new model. The excluded compound's bioactivity was then predicted using this model. This iterative process was repeated for all compounds, and the Q^2 value was computed in accordance with the formula shown in our previously published report [18], thereby confirming the models' internal consistency.

$$Q^2 = 1 - \frac{\sum(Y_{pred} - Y_{act})^2}{\sum(Y_{act} - Y_{mean})^2}$$

where Y_{pred} and Y_{act} are the predicted and actual bioactivity [$\text{Log}_{10}(\text{IC}_{50})$] of the compounds in the training set, respectively. Y_{mean} is the average activity of all compounds in the training set.

External validation: The reliability and external

predictive performance of the QSAR models were assessed using a designated test set comprising compounds excluded from the model-building process. The external validation was conducted by predicting the biological activities of these test set compounds, and the predictive ability of the models was evaluated based on established statistical metrics, following the methodology reported in our previous work [18].

$$pred_R^2 = 1 - \frac{\sum(Y_{pred(test)} - Y_{test})^2}{\sum(Y_{test} - Y_{training})^2}$$

where $Y_{pred(Test)}$ is the predicted activity of the test compounds, Y_{Test} is the actual activity of test set compounds, and $Y_{training}$ is the mean activity of all compounds in the training set.

Randomization Test

The Y-randomization test was employed to evaluate the robustness and statistical significance of the developed QSAR models. This involved generating randomized datasets by shuffling the biological activity values of the training set compounds, then constructing the model using these permuted datasets. The performance of the original models was then compared with that of the models derived from the randomized data to assess the likelihood of chance correlations. The model's significance was quantified using the Z-score, calculated as described in our previous study [18].

$$Z\ score = \frac{Q_{org}^2 - Q_a^2}{Q_{std}^2}$$

where Q_{org}^2 is the Q^2 value calculated for the actual data set, Q_a^2 is the average Q^2 , and Q_{std}^2 is the standard deviation of Q^2 , calculated for various repetitions using different random data sets.

Applicability Domain (AD) Analysis

The AD analysis was performed to assess the reliability of the developed QSAR models and to identify compounds within the model's predictive space [21]. The AD was evaluated using the standardization approach based on Euclidean distance in the descriptor space. Each compound's leverage and standardized residual were calculated, and a Williams plot was used to visualize the AD. Compounds with leverage values greater than the warning leverage (h^*) and standardized residuals beyond ± 3 standard deviation units were considered influential or outliers, indicating that their predictions may be less reliable [22]. This analysis ensured that the majority of the compounds fell within the model's defined chemical space, supporting the robustness and generalizability of the models. To streamline and automate this analysis, custom Python scripts were used, enabling efficient computation

and visualization of applicability-domain metrics. Leverage indicates a compound's influence on the model and is derived from the Hat matrix (h_i). It is calculated as:

$$h_i = \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i$$

where $\mathbf{x}_i$ is the descriptor vector of the i -th compound, and $\mathbf{X}$ is the matrix of descriptors used in model building (training set).

The residual is the difference between the observed and predicted activity. It is expressed as:

$$r_i = Y_i - \hat{Y}_i$$

where Y_i is the observed value, and $\hat{Y}_i$ is the predicted value.

The standardized residual normalizes the residuals to enable comparison and is expressed as:

$$SR_i = \frac{r_i}{\sigma_r}$$

where σ_r is the standard deviation of residuals in the training set.

The warning leverage threshold (h^*) is calculated as:

$$h^* = \frac{3(p+1)}{n}$$

where p is the number of model descriptors (independent variables), and n is the number of compounds in the training set [22].

Bioactivity Prediction of Experimental Phytocompounds

A small library of 54 experimental phytocompounds was prepared from the literature, and their 3D structures were retrieved from the PubChem database, subsequently energy-minimized, and 2D/3D molecular descriptors were computed using the same procedure as for the known COX-2 inhibitors during model development. The predicted biological activities of these compounds were computed using the best-developed 2D and 3D QSAR models in the Generic Prediction module of VLifeMDS, as shown in our earlier work [18].

Computational ADMET and Rodent Toxicology Prediction

The QSAR-prioritized phytocompounds were subjected to comprehensive pharmacokinetic and toxicological profiling using multiple *in silico* pharmacological tools. The pharmacokinetics, drug-likeness, and medicinal chemistry friendliness of the selected compounds were predicted using the SwissADME server ([http://www.](http://www.swissadme.ch/)

[swissadme.ch/](http://www.swissadme.ch/)) of the Swiss Institute of Bioinformatics, University of Lausanne [23]. This tool provides a broad range of predictive parameters, including gastrointestinal absorption, blood-brain barrier (BBB) permeability, P-gp substrate status, and cytochrome P450 (CYP) inhibition profile. In addition, drug-likeness filters such as the Lipinski, Ghose, Veber, Egan, and Muegge rules were evaluated, among others. For additional ADMET profiling, the PreADMET server (<https://preadmet.webservice.bmdrc.org/toxicity/>) developed by Yonsei University, Republic of Korea, was employed. The server predicted human intestinal absorption (HIA), plasma protein binding (PPB), Caco2/MDCK cell permeability, and hERG Inhibition. Additional drug-likeness indicators such as bioavailability score and structural alerts (PAINS/Brenk) were also recorded.

To gain deeper insights into toxicity endpoints, the ProTox-III web server (<https://tox.charite.de/protox3/>), developed by Charité University of Medicine, Germany, was utilized [24]. This platform enabled the prediction of acute toxicity (LD50) in the rodent model, organ-specific toxicity, and a wide spectrum of toxicological endpoints, including mutagenicity, carcinogenicity, immunotoxicity, and cytotoxicity. The tool also provides insight into molecular initiating events, adverse outcome pathways (AOPs) from the Tox21 database, and possible interactions with known toxicity targets. The descriptions and validation metrics for the QSAR models employed by each server are available on their respective websites. All computational predictions were performed using default settings and interpreted comparatively to prioritize compounds with favorable pharmacokinetic and safety profiles for further development.

Protein-Ligand Interaction and Stability Prediction

Protein Structure Preparation & Energy Minimization: The crystallographic 3D structure of COX-2 was retrieved from the RCSB Protein Data Bank in PDB format (PDB ID: 6COX) [25]. Various researchers recently used the same PDB structure for their computational studies [26-28]. Before energy minimization, the protein structure was pre-processed to remove heteroatoms, such as crystallographic water molecules, ions, and non-receptor ligands, using Discovery Studio Visualizer [29]. The structure was checked, and any missing atoms were added. The cleaned protein file was then converted to a GROMACS-compatible format, applying the CHARMM36 all-atom force field and the TIP3P water model for solvation [30]. A cubic simulation box was generated around the protein with a minimum distance of 1.0 nm from the edges. The system was solvated by filling the box

with explicit water molecules. To neutralize the system and mimic physiological ionic strength, appropriate numbers of counterions (Na^+ or Cl^-) were added. Subsequently, the system topology was updated, and energy minimization was performed using the steepest descent algorithm for up to 50,000 steps or until convergence (maximum force < 1000 kJ/mol/nm). The minimization was carried out under periodic boundary conditions using the PME (Particle-Mesh Ewald) method for long-range electrostatics, and the LINear Constraint Solver (LINCS) algorithm was used to constrain bond lengths involving hydrogen atoms. The minimized structure was saved as a PDB file and subsequently used as the receptor structure for downstream molecular docking studies.

Molecular Docking Analysis: The molecular docking simulations were performed using AutoDock Vina and AutoDock 4.2 [31] to estimate the binding affinities and interaction profiles of the prioritized ligands and the reference drug, celecoxib, within the COX-2 active site. First, polar hydrogens were added, Kollman charges were assigned to the protein, and Gasteiger charges were computed for the ligands. All ligand torsions were allowed to rotate freely, whereas the receptor was treated as a rigid structure. A grid box was defined to encompass the protein's active site, centered at $X = 46.3725$, $Y = 50.6325$, and $Z = 51.4925$, which was large enough to cover the entire binding cavity. The docking simulations employed the Lamarckian genetic algorithm with the following optimized parameters: number of GA runs = 50, population size = 300, maximum number of energy evaluations = 2,500,000, maximum number of generations = 27,000, and other default parameters were maintained. Following docking, the output conformations were clustered using a root-mean-square deviation (RMSD) threshold of $2.0 \text{ \AA}$. The best-ranked pose in each cluster, defined as the one with the lowest binding energy, was selected for further analysis.

Molecular Dynamics Simulation: To evaluate the dynamic behavior and stability of the best-hit protein-ligand complex, molecular dynamics (MD) simulations were performed using the Desmond module of the Schrödinger Suite, as described in our previous report [18]. Briefly, the top-ranked docking pose of ligands in complex with COX-2 was used as the starting structure. The complex was embedded in an orthorhombic box, solvated with the TIP3P water model, and neutralized by adding appropriate counterions. Physiological ionic strength (0.15 M NaCl) was maintained to mimic biological conditions. The system was then equilibrated using default NVT (constant number of particles, volume, and temperature) and NPT (constant number of particles,

pressure, and temperature) ensembles. A 50 ns production run was carried out at 300 K and 1 atm, with trajectories recorded every 100 ps. The resulting simulation trajectory was analyzed to compute root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), radius of gyration (Rg), hydrogen bonding profiles, and molecular mechanics/generalized Born surface area (MM/GBSA) to assess the conformational stability and binding integrity of the complex over time. These dynamic parameters provided insights into the structural adaptability and sustained interaction of the best-hit ligand within the COX-2 binding pocket under physiological conditions.

RESULTS

QSAR Datasets and Bioactivity Distribution

For the 2D-QSAR models, the dataset had 33 compounds for training, 11 for testing, and 6 for external validation. Similarly, the 3D-QSAR models utilized 35 training compounds, 9 test compounds, and 6 for external validation. Before model building, invariant descriptors were removed, leaving 393 and 6655 descriptors for the 2D and 3D model development, respectively. For the 2D-QSAR models, the training set had a mean activity of 2.1994, a range of -0.24 to 4, and a standard deviation (SD) of 1.4362 (sum = 72.58). The test set displayed a lower mean activity (1.2009), a narrower range (0.24 to 3.49), and a slightly reduced SD of 1.1542 (sum = 13.21). For the 3D-QSAR models, the training set had a mean of 2.3617 (range: -0.24 to 4, SD = 1.3301, sum = 82.66), while the test set had a mean of 1.5311 (range: -0.19 to 3.42, SD = 1.4717, sum = 13.78). The UniColumn statistics, provided in Table 1, indicate that, in both 2D and 3D datasets, the test sets (including validation sets) are interpolative, i.e., their minimum and maximum activity values fall within the corresponding ranges of the training sets (Figure 1). This ensures that the models are not being extrapolated beyond their training domain. Furthermore, the observed differences in mean values suggest that the training sets contain relatively more active molecules. In contrast, the comparable SD values between the training and test sets reflect similar distributions of activity values around their

Table 1: UniColumn statistical analysis. The table presents a comparative summary of statistical parameters like average, maximum, minimum, standard deviation (SD), and sum, between the training and test sets with respect to the dependent variable ($\text{Log}_{10}[\text{IC}_{50}]$), in both 2D- and 3D-QSAR models.

Dependent Variable, $\text{Log}_{10}(\text{IC}_{50})$	2D-QSAR		3D-QSAR	
	Training Set	Test Set	Training Set	Test Set
Average	2.1994	1.2009	2.3617	1.5311
Maximum	4.00	3.49	4.00	3.42
Minimum	-0.24	0.24	-0.24	-0.19
Standard Deviation	1.4362	1.1542	1.3301	1.4717
Sum	72.58	13.21	82.66	13.78

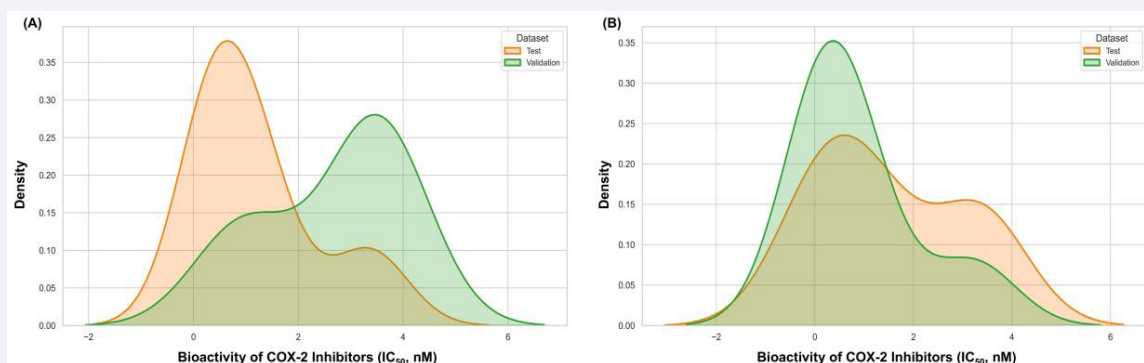


Figure 1 Bioactivity distribution plots for training, test, and validation datasets. (A) Distribution plot for the 2D-QSAR models and (B) distribution plot for the 3D-QSAR models, showing compound density versus bioactivity (Log₁₀ IC₅₀) values.

respective means. Collectively, these statistical parameters validate the representativeness and integrity of the dataset partitioning strategy for domain-appropriate modeling, thereby reinforcing the reliability of subsequent model development and validation.

Structural Similarity and Scaffold Diversity Analysis

To assess the structural diversity in the compound dataset used for the model development, pairwise Tanimoto similarity indices were computed. The resulting similarity scores ranged from 0.00 to 0.72, with a mean of 0.172 and a median of 0.123, indicating low pairwise structural similarity across the dataset. The Tanimoto similarity heatmap and histogram confirmed that the majority of compound pairs shared minimal structural overlap, with most similarity values clustering below 0.2. This suggests a chemically diverse dataset, which is desirable for building robust and generalizable QSAR models. The hierarchical clustering based on the similarity matrix revealed no dominant clusters, further supporting the high degree of structural heterogeneity. In addition, Bemis-Murcko scaffold analysis identified 30 unique scaffolds, with 7 being the most frequent, confirming chemical diversity at the core framework level and thereby minimizing the risk of scaffold bias in the models. These findings, shown in Figure 2, collectively support the representativeness and coverage of the chemical space by the compounds used for model development. The most dominant chemotypes within the dataset, i.e., the unique scaffolds with the highest frequencies, are shown in Figure 3.

Developed 2D-QSAR Models and Validation Metrics

Each selected 2D model, from both MLR and PLS approaches, showed a strong correlation between predicted and actual bioactivity values in the training set

($R^2 = 0.75$), suggesting that the models fit the training data well. The RMSE of 0.71 for the training set indicates a relatively low prediction error for these compounds. However, test-set performance was substantially lower ($R^2 = 0.32$, RMSE = 1.11). This indicates that the model's predictive ability on unseen data may be limited, suggesting the need to include more diverse descriptors, perform dimensionality reduction, or apply model regularization. These results for MLR and PLS models are visualized in the regression plots, provided in Figure 4, where predicted bioactivity values are plotted against experimentally observed values for both training and test sets. In addition, a radar plot, as shown in Figure 5, was used to compare the actual and predicted bioactivity values of the individual compounds in the training set. This visual representation helps in identifying compounds where the model performs consistently and where it may deviate across individual compounds.

Both MLR and PLS models converged on a consistent set of five descriptors (Table 4) that were identified as statistically and chemically relevant contributors to the inhibitory activity, expressed as Log₁₀(IC₅₀) values. The final MLR model yielded the following regression equation:

$$\text{Log}_{10}(\text{IC}_{50}) = -0.3827 (\pm 0.1411) \text{SsssCHcount} + 1.4457 (\pm 0.2491) \text{SssNHcount} - 0.4126 (\pm 0.1643) \text{T_N_O_5} + 0.6143 (\pm 0.2042) \text{SsCH3count} - 0.0178 (\pm 0.0008) \text{T_2_C_6} + 1.9458 (\text{Constant})$$

The corresponding PLS model provided a comparable equation:

$$\text{Log}_{10}(\text{IC}_{50}) = -0.4548 \text{SsssCHcount} + 1.4255 \text{SssNHcount} - 0.4154 \text{T_N_O_5} + 0.5132 \text{SsCH3count} - 0.0140 \text{T_2_C_6} + 1.9813 (\text{Constant})$$

The descriptors SsssCHcount, SssNHcount, and SsCH3count are E-State (electrotopological state) and topological. SsssCHcount is the count of -CH groups connected through three single bonds, negatively correlated with activity, suggesting steric bulk from

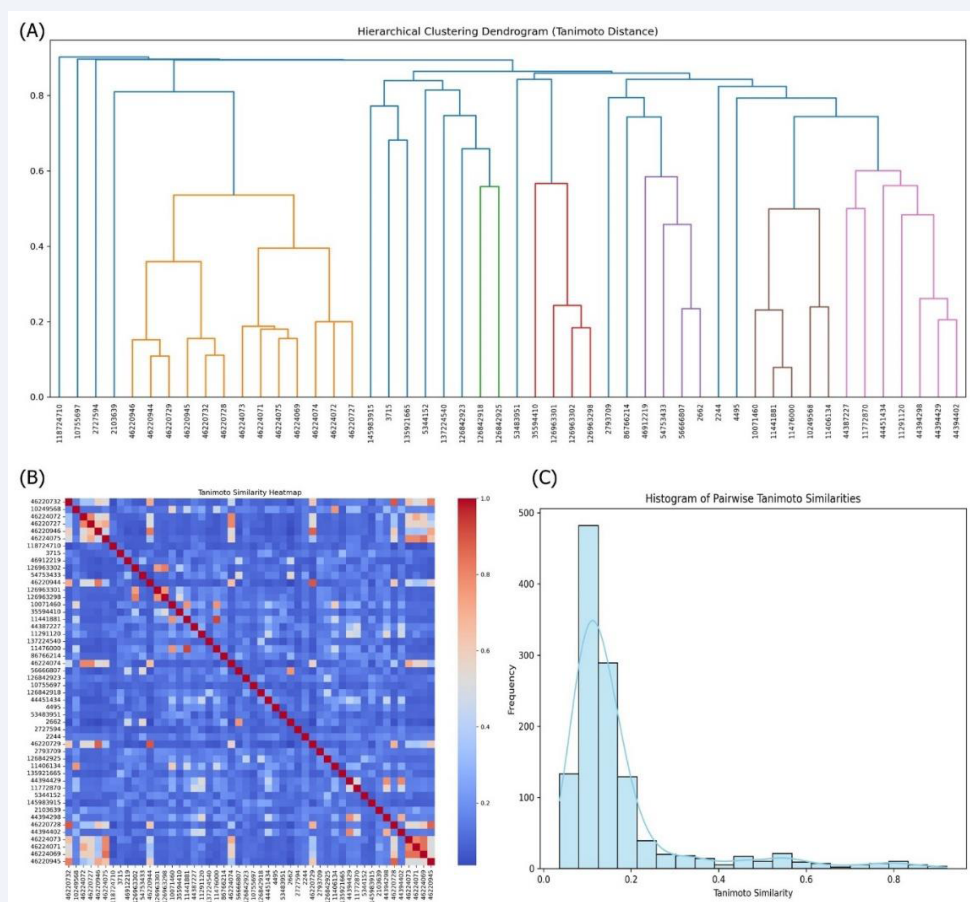


Figure 2 Structural similarity and diversity analysis of the QSAR dataset. **(A)** Tanimoto dendrogram representing hierarchical clustering of 50 compounds based on ECFP4 fingerprints (radius = 2, 1024 bits) and Tanimoto distance (1 – similarity). **(B)** Heatmap of pairwise Tanimoto similarity values showing low structural overlap among most compound pairs. **(C)** Histogram of pairwise Tanimoto similarity scores, illustrating the distribution of structural similarity across the dataset. The majority of compound pairs exhibited low similarity (mean = 0.172; median = 0.123), indicating a chemically diverse dataset suitable for robust QSAR model development.

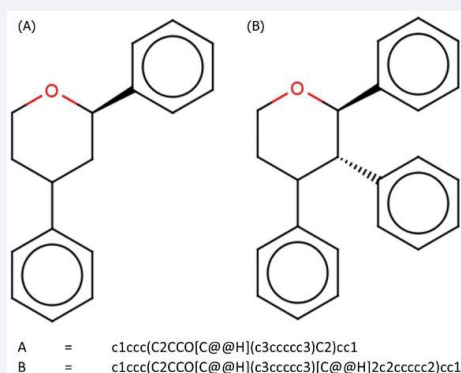


Figure 3 Representative scaffolds identified from the QSAR compound dataset using Bemis–Murcko scaffold analysis. The figure shows the chemical structures of the two most frequently occurring core scaffolds: Scaffold A (7 occurrences) and Scaffold B (6 occurrences). These scaffolds represent the dominant chemotypes in the dataset and provide insight into recurring molecular frameworks across structurally diverse compounds.

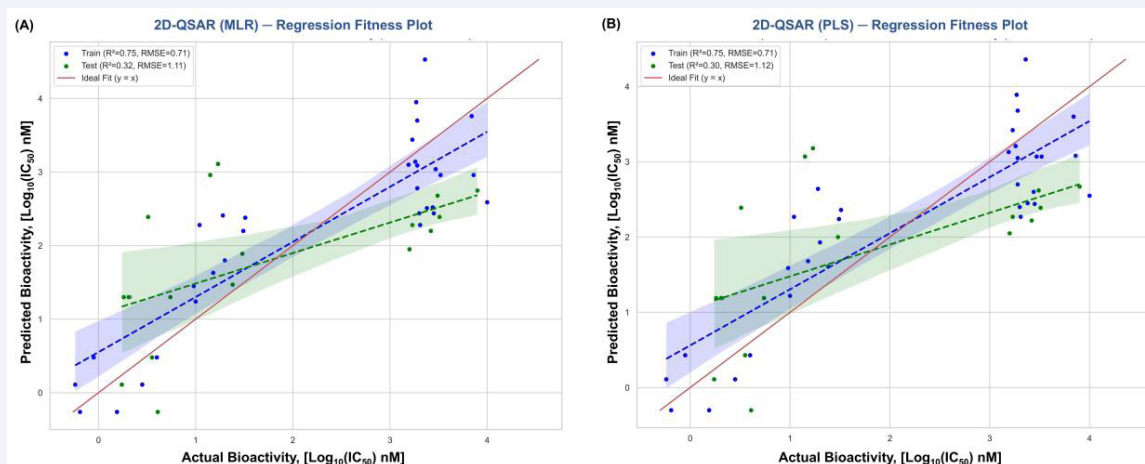


Figure 4 Regression Fitness Plots of 2D-QSAR models (MLR/PLS). (A) Fitness plot of MLR model showing predicted versus actual bioactivity values for 50 known COX-2 inhibitors. The plot includes data points for both the training set ($R^2 = 0.75$, $RMSE = 0.71$) and the test set ($R^2 = 0.32$, $RMSE = 1.11$). (B) Fitness plot of PLS model showing data points for both the training set ($R^2 = 0.75$, $RMSE = 0.71$) and the test set ($R^2 = 0.30$, $RMSE = 1.12$), illustrating the model's performance in fitting known data and predicting unseen data.

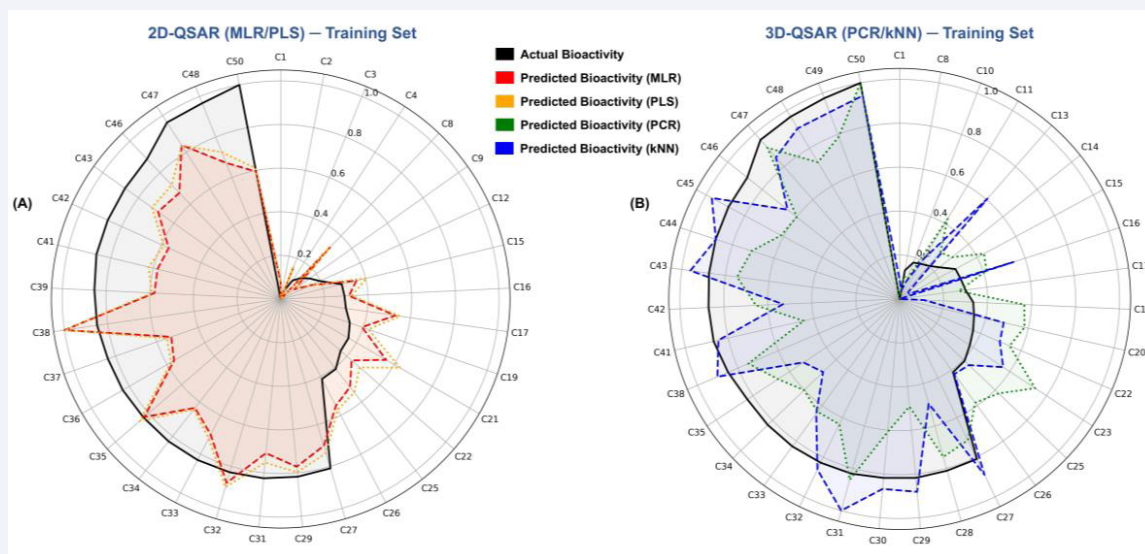


Figure 5 Radar plots of the training set data. The figures illustrate the robustness of the (A) 2D-QSAR models (MLR and PLS) and (B) 3D-QSAR models (PCR and kNN) by comparing actual versus predicted bioactivity values for compounds in the training set. The plot highlights the consistency and deviation patterns of the model's predictive performance across individual compounds.

such carbon environments may reduce COX-2 inhibition. However, *SssNHcount*, or the number of $-NH$ groups attached via two single bonds, positively correlated, indicating hydrogen-bond-donating potential may enhance binding affinity. Similarly, *SsCH3count* or the total number of $-CH_3$ groups connected through a single bond, positively correlated, implying a favorable contribution of small hydrophobic groups to activity. In addition, the

alignment-independent topological descriptors $T_N_O_5$ and $T_2_C_6$, which are negatively correlated, likely reflect an unfavourable electronic configuration or molecular conformation for receptor interaction. The coherence between MLR and PLS models, both in terms of descriptor selection and in their relative contributions to activity, enhances the reliability of these findings. Notably, the model underscores the significance of both local electronic

environments (as captured by E-state descriptors) and long-range topological features (captured by alignment-independent descriptors) in modulating COX-2 inhibitory potency (Figure 6).

The developed MLR and PLS models demonstrated satisfactory statistical strength, as reflected in multiple performance indices shown in Table 2. The coefficient of determination (R^2) was 0.7485 for MLR and 0.7451 for PLS, indicating that approximately 74.85% and 74.51% of the variance in the observed biological activity [$\text{Log}_{10}(\text{IC}_{50})$] could be explained by the respective models, which reflects good model fit for both approaches. The cross-validated coefficient (Q^2) values were 0.6280 (MLR) and 0.6307 (PLS), suggesting that both models retained strong internal predictive performance, explaining approximately 62.8% and 63.07% of the variability in activity under the leave-one-out cross-validation (LOO-CV) method, indicating reliable generalization within the training dataset. The F-statistic (F-test) values were

Table 2: Statistical validation metrics of the 2D-QSAR models (MLR and PLS). The table summarizes key validation metrics, including the number of compounds (n), degrees of freedom (DF), coefficient of determination (R^2), cross-validated correlation coefficient (Q^2), F-test value (F_test), and predictive R^2 for the test set (pred_ R^2). These metrics collectively assess the goodness-of-fit, internal robustness, predictive accuracy, and external predictability of each QSAR model.

Statistics	2D-MLR	2D-PLS
Optimum Components	--	3
n	33	33
DF	27	29
R^2	0.7485	0.7451
Q^2	0.6280	0.6307
F_test	16.0708	28.2532
pred_ R^2	0.5308	0.5157

16.07 for MLR and 28.25 for PLS. The higher F value for PLS implies greater overall model significance in terms of variance explained relative to residual error, reinforcing its robustness in regression fitting. The external predictive correlation coefficient (pred_ R^2) was 0.5308 for MLR and 0.5157 for PLS, indicating that each model achieved approximately 53.08% and 51.57% accuracy in predicting the activity of external test set compounds, respectively. These values are above the widely accepted threshold of 0.5, affirming the models' external predictive validity. In addition, the Z-score-based statistical validation provided in Table 3 further confirmed the statistical reliability of the developed models. For MLR, the Z-scores of R^2 , Q^2 , and pred_ R^2 were 9.81, 6.29, and 1.70, respectively, whereas for the PLS model, corresponding Z-scores were 7.32, 6.31, and 2.26, respectively. These values, all significantly greater than 1.65, underscore that the observed statistical parameters are unlikely to have arisen by chance, thereby validating the models' robustness. Together, these statistical metrics confirm that the newly developed 2D-QSAR models exhibit good internal consistency, strong cross-validation performance, and acceptable external predictive performance, with supporting evidence from both the MLR and PLS frameworks. This affirms the models' potential for reliable prediction and mechanistic interpretation of COX-2 inhibitory activity in structurally related compounds. A comparison of the actual bioactivity (IC_{50} in nM) values of 50 known COX-2 inhibitors with the predicted bioactivities obtained from the developed 2D-QSAR models is provided in Table 5.

To evaluate the reliability and predictive scope of the

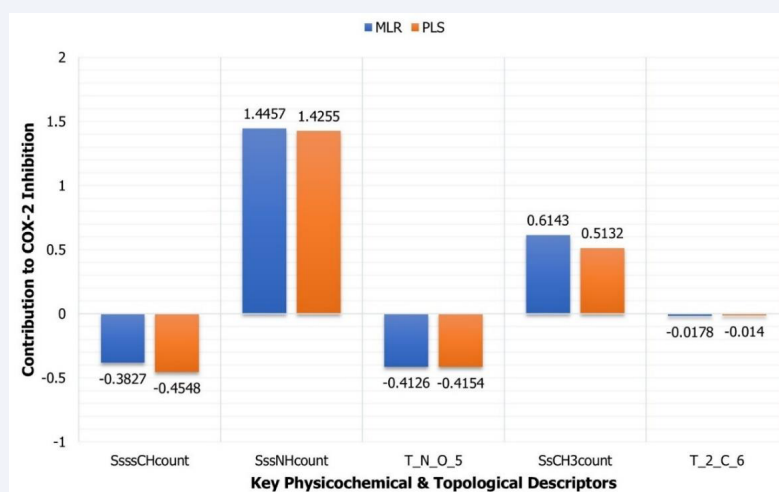


Figure 6 Contribution of key descriptors to COX-2 inhibitory activity. Bar plot showing the relative contribution (regression coefficients) of the five key physicochemical and topological descriptors influencing COX-2 inhibition, as identified by the 2D-QSAR (MLR and PLS) models.

Table 3: Z-score-based statistical validation of the developed 2D-QSAR models using Y-randomization test. The table shows Z-scores and best randomized values for R² (training set fit), Q² (cross-validation), and pred_R² (external validation), indicating the robustness, statistical significance, and predictive reliability of the model.

Log ₁₀ (IC ₅₀)	Z-Score R ²	Z-Score Q ²	Best Rand R ²	Best Rand Q ²	Z Score pred_R ²	best Rand pred_R ²
MLR	9.812	6.290	0.288	0.125	1.700	0.447
PLS	7.329	6.319	0.373	0.068	2.261	0.394

Table 4: Definition and interpretation of key 2D molecular descriptors. These descriptors contribute to the COX-2 binding activity of the compounds as identified in MLR and PLS models, and represent specific structural, topological, electronic, or physicochemical features that significantly influence the inhibitory potential.

2D Descriptor	Descriptor type	Definition/Interpretation
SsssCHcount	Estate Numbers	The total number of -CH groups connected with three single bonds.
SssNHcount	Estate Numbers	The total number of -NH groups connected by two single bonds.
T_N_O_5	Alignment Independent	The count of the number of Nitrogen atoms (single, double, or triple bonded) separated from the oxygen atom (single, double, or triple bonded) by 5 bonds in a molecule.
SsCH3count	Estate Numbers	The total number of -CH ₃ groups connected with a single bond.
T_2_C_6	Alignment Independent	The count of the number of double-bonded atoms separated from the carbon atom by 6 bonds in a molecule.

Table 5: Comparison of the actual bioactivity (IC₅₀ in nM) of the known COX-2 inhibitors with their predicted bioactivities from the developed 2D-QSAR models. The table shows the 50 compounds, their assigned dataset, and the assessment of their actual vs. predicted inhibitory concentrations, providing insight into the predictive performance and reliability of each model in estimating COX-2 inhibitory activity.

Sl. No.	Compound ID	Dataset	Actual Bioactivity, IC ₅₀ (nM)		Predicted Bioactivity, [Log ₁₀ (IC ₅₀)] (nM)	
			IC ₅₀	Log ₁₀ (IC ₅₀)	MLR	PLS
1	CHEMBL1096723	Training	0.57	-0.24	0.11	0.11
2	CHEMBL1096724	Training	0.65	-0.19	-0.26	-0.30
3	CHEMBL1095083	Training	0.90	-0.05	0.48	0.43
4	CHEMBL1094412	Training	1.55	0.19	-0.26	-0.30
5	CHEMBL1094097	Test	1.72	0.24	0.11	0.11
6	CHEMBL1095082	Test	1.82	0.26	1.30	1.19
7	CHEMBL1096037	Test	2.05	0.31	1.30	1.19
8	CHEMBL1096384	Training	2.07	0.32	1.30	1.19
9	CHEMBL1096725	Training	2.80	0.45	0.11	0.11
10	CHEMBL71833	Test	3.20	0.51	2.39	2.39
11	CHEMBL1096385	Test	3.57	0.55	0.48	0.43
12	CHEMBL1094096	Training	4.01	0.60	0.48	0.43
13	CHEMBL1095094	Test	4.03	0.61	-0.26	-0.30
14	CHEMBL1095084	Validation	5.49	0.74	1.30	1.19
15	CHEMBL3361124	Training	9.50	0.98	1.45	1.59
16	CHEMBL6	Training	10	1.00	1.24	1.22
17	CHEMBL1224098	Training	11	1.04	2.28	2.27
18	BDBM227581	Test	14	1.15	2.96	3.07
19	CHEMBL1835477	Training	15	1.18	1.63	1.68
20	BDBM227580	Validation	17	1.23	3.11	3.18
21	BDBM227577	Training	19	1.28	2.41	2.64
22	CHEMBL187137	Training	20	1.30	1.80	1.93
23	BDBM227570	Test	24	1.38	1.47	1.61
24	CHEMBL184737	Test	30	1.48	1.89	2.00
25	CHEMBL175686	Training	31	1.49	2.20	2.24
26	CHEMBL360720	Training	32	1.51	2.38	2.36
27	BDBM222445	Training	1560	3.19	3.10	3.13
28	CHEMBL365497	Validation	1600	3.20	1.95	2.05

29	BDBM123807	Training	1680	3.23	3.44	3.42
30	CHEMBL1835475	Training	1700	3.23	2.28	2.27
31	BDBM222446	Training	1830	3.26	3.14	3.21
32	CHEMBL9192	Training	1850	3.27	3.95	3.89
33	BDBM222441	Training	1890	3.28	3.09	3.05
34	CHEMBL259401	Training	1900	3.28	2.78	2.70
35	CHEMBL56367	Training	1920	3.28	3.70	3.68
36	CHEMBL2147422	Training	2000	3.30	2.44	2.40
37	CHEMBL118	Training	2050	3.31	2.28	2.27
38	BDBM50106737	Training	2300	3.36	4.53	4.36
39	CHEMBL25	Training	2400	3.38	2.51	2.45
40	CHEMBL412610	Validation	2630	3.42	2.20	2.22
41	BDBM222448	Training	2740	3.44	2.52	2.60
42	CHEMBL70835	Training	2800	3.45	2.44	2.44
43	ZINC27015737	Training	2950	3.47	3.04	3.07
44	CHEMBL187481	Test	3100	3.49	2.68	2.62
45	CHEMBL364663	Validation	3200	3.51	2.39	2.39
46	BDBM50035979	Training	3330	3.52	2.96	3.07
47	CHEMBL4276853	Training	6990	3.84	3.76	3.60
48	CHEMBL1702044	Training	7210	3.86	2.96	3.08
49	CHEMBL365288	Validation	8000	3.90	2.75	2.67
50	CHEMBL365615	Training	10000	4.00	2.59	2.55

developed 2D-QSAR models, applicability domain (AD) analysis was conducted using the leverage approach, visualized via the Williams plot, and the results are provided in **Supplementary Table 3**. The leverage threshold ($h_{\max}$) was calculated to be 0.360 based on the number of model descriptors and the number of training compounds. Compounds with leverage values higher than $h_{\max}$ were considered structural outliers, while compounds with standardized residuals beyond ± 3 were deemed response outliers. In the case of the MLR model, only one compound (CHEMBL6) from the training set exceeded the leverage threshold ($L=0.4214$) but remained within the acceptable range of standardized residuals ($SR=-1.05$). This compound was classified as a structural outlier but not a response outlier, indicating that although its molecular structure diverged from the modelled domain, its predicted activity remained statistically consistent (Figure 7). In contrast, the PLS model showed no outliers (Figure 7), with all compounds falling within the defined AD, i.e., within both the leverage cutoff and standardized residual range. This reflects a more uniformly distributed model domain, with better accommodation of compound diversity and activity prediction. Overall, the AD analysis confirmed the robustness of both models, with only one

mild outlier in the MLR model and no violations in the PLS model, indicating that the vast majority of compounds used for modeling and external validation were within the defined applicability domain. These outcomes support the predictive credibility of the developed QSAR models for structurally similar COX-2 inhibitory compounds.

Developed 3D-QSAR Models and Validation Metrics

To find the spatial determinants of COX-2 inhibitory potential of chemicals, 3D-QSAR models were developed using aligned molecules and their steric (S) and electrostatic (E) field descriptors. To assess training set robustness in the models, a radar plot comparing actual versus predicted $\text{Log}_{10}(\text{IC}_{50})$ values was provided in Figure 5, showing that the predicted activities align closely with the actual ones for most compounds. The PCR model was built using four key descriptors: S_{884} , E_{1977} , E_{2605} , and E_{951} . **Supplementary Table 4**. The resulting regression equation was:

$$\text{Log}_{10}(\text{IC}_{50}) = 249.7690 S_{884} - 0.0690 E_{1977} + 1.4933 E_{2605} - 0.1212 E_{951} + 3.2534 (\text{Constant})$$

This equation suggests that the steric field descriptor S_{884} is the most influential positive contributor, while E_{1977} and E_{951} are the most influential negative contributors.

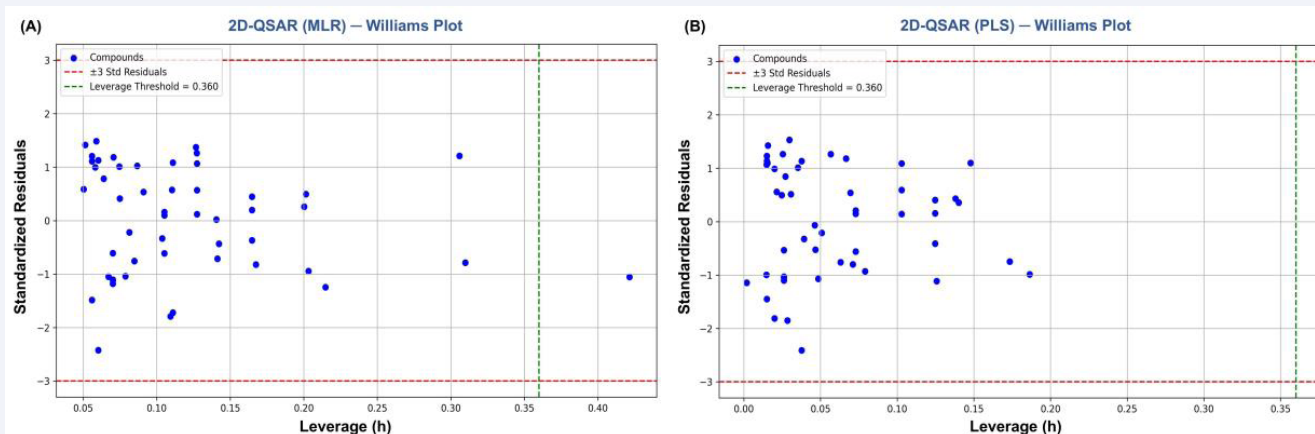


Figure 7 Williams plots for the 2D-QSAR model. (A) Standardized residuals vs. Leverage values for the MLR model. The horizontal dashed lines represent the ± 3 standardized residual threshold, indicating the cut-off for response outliers. The vertical dashed line indicates the leverage threshold ($h = 0.360$). **(B)** Standardized residuals vs. Leverage values for the PLS model. The plot illustrates that all compounds are within the defined leverage threshold ($h = 0.360$) and the ± 3 residual range.

The model explained 64% of the variance in the training data ($R^2 = 0.64$, $RMSE = 0.79$) but only 19% of the variance in the test set ($R^2 = 0.19$, $RMSE = 1.20$), indicating limited generalizability, as shown in Figure 8. Nevertheless, internal validation metrics ($R^2 = 0.72$, $Q^2 = 0.56$, $pred_R^2 = 0.67$) and favorable Z-scores (10.00 for R^2 , 7.85 for Q^2 , and 2.49 for $pred_R^2$) of the PCR model confirm that the model is statistically significant in accordance with the OECD guidelines and not a result of chance correlations (Table 6). The optimum number of components in the PCR model was determined to be 3.

The kNN model, optimized with $k = 2$ nearest neighbours, performed better in both internal and external predictive performance. It accounted for 70% of the variance in the training set ($R^2 = 0.70$, $RMSE = 0.72$) and 56% in the test set ($R^2 = 0.56$, $RMSE = 0.89$), shown in Figure 8. The cross-validation and external prediction were also strong, with $Q^2 = 0.70$, $pred_R^2 = 0.58$, and an external validation R^2 of 0.91, indicating high predictive reliability (Table 6). The model identified four steric field points with their ranges: S659 (-0.0016 to -0.0005), S1369 ($+30.00$ to $+30.00$), S2747 (-0.0025 to -0.0010), and S2825 (-0.0162 to -0.0143) as critical influencers of the COX-2 inhibitory activity of chemicals. The 3D field point map, provided in Figure 9, highlights the spatial locations around the aligned molecules where these steric interactions occur. A comparison of the actual bioactivity (IC_{50} in nM) values of known COX-2 inhibitors used for model building against the predicted bioactivities obtained from the developed 3D-QSAR models is provided in Table 7.

The applicability domain (AD) of both models was determined using the leverage (L) method with a standardized residual (SR) cutoff of ± 3 , and the results are shown in **Supplementary Table 5**. For the PCR model, the warning leverage value (h^*) was calculated as 0.240. Two compounds from the model building dataset, CHEMBL1096384 (L/SR = 0.273/ -0.698) and ZINC27015737 (L/SR = 0.330/1.309), were flagged as mild outliers, suggesting they lie slightly beyond the ideal chemical space (Figure 10). For the kNN model, the AD threshold was 0.300, and again, two compounds, CHEMBL175686 (L/SR = 0.99/ -1.19) and ZINC27015737 (L/SR = 0.96/ -0.23), were also identified near the boundary (Figure 10). The structural outliers in both the PCR and kNN models were in the training datasets used for model development. Overall, over 95% of the compounds were within the model's reliable predictive domain for COX-2 inhibitory activity.

QSAR-guided Bioactivity Prediction of Phytocompounds

A total of 54 plant-derived compounds were selected, and their COX-2 inhibitory potential was predicted using the newly developed and best-performing 2D- and 3D-QSAR models. The prediction results obtained using the 2D-QSAR models (MLR and PLS) are provided in Table 8, while those from the 3D-QSAR models (PCR and kNN) are presented in Table 9. These tables provide a comparative overview of the inhibitory potential across different compounds as estimated by the respective models. Among all models, the kNN model demonstrated

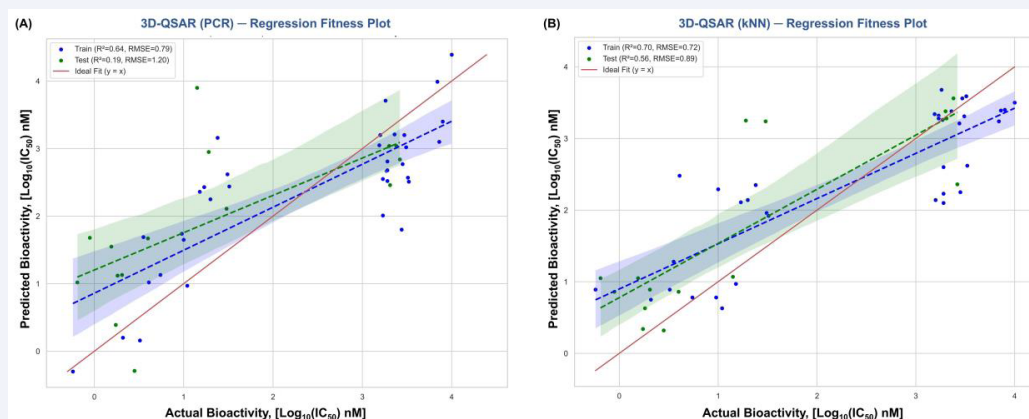


Figure 8 Regression Fitness Plots of 3D-QSAR models (PCR/kNN). (A) Fitness plot of PCR model showing predicted versus actual bioactivity values for 50 known COX-2 inhibitors. The plot includes data points for both the training set ($R^2 = 0.64$, $RMSE = 0.79$) and the test set ($R^2 = 0.19$, $RMSE = 1.20$). (B) Fitness plot of kNN model showing data points for both the training set ($R^2 = 0.70$, $RMSE = 0.72$) and the test set ($R^2 = 0.56$, $RMSE = 0.89$), illustrating the model's performance in fitting known data and predicting unseen data. These results suggest that the kNN model exhibits both robustness during training and a reasonable level of external predictive performance.

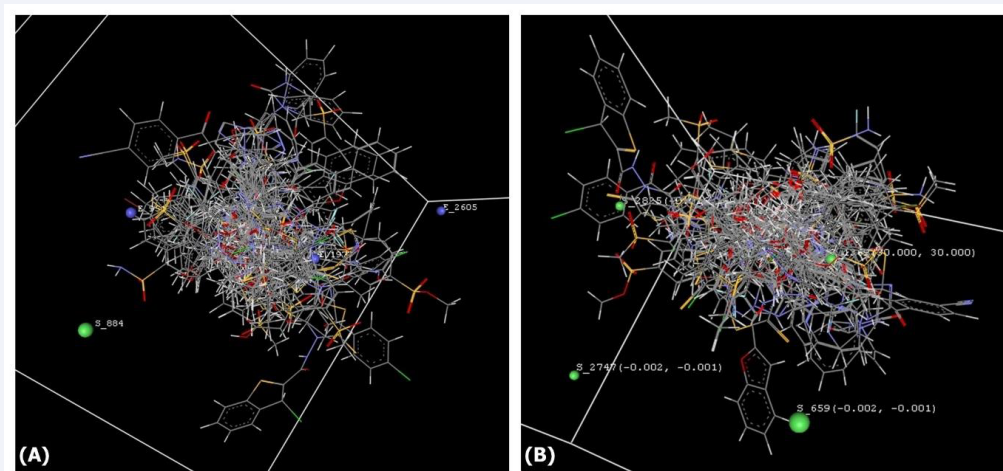


Figure 9 3D field point maps of (A) 3D-PCR model, (B) kNN model. The maps illustrate the spatial distribution of steric and electrostatic interactions influencing the COX-2 inhibitory potential of the aligned compounds. The maps highlight regions where these molecular interactions contribute positively or negatively to activity, providing visual insight into key pharmacophoric features.

Table 6: Statistical validation metrics of the 3D-QSAR models (PCR and kNN). The table presents the same set of statistical metrics as described in 2D models, enabling a comparative evaluation of model performance and predictive reliability across different modeling approaches.

Statistics	PCR	kNN
k Nearest neighbour	--	2
Optimum Components	3	--
n	35	35
Degree of freedom	31	30
R^2	0.6360	--
Q^2	0.5570	0.6997
F_test	18.0525	--
pred_ R^2	0.6709	0.5800

the highest statistical robustness, as indicated by superior internal and external validation parameters. Therefore, the predictions obtained from the kNN model were considered the most reliable for further analysis.

Computed Pharmacokinetic Parameters of the QSAR-prioritized Compounds

To evaluate the pharmacokinetic behavior and drug-likeness of the QSAR-prioritized phytocompounds, *in silico* ADMET profiling was executed using the SwissADME web server. A total of twenty (20) phytocompounds,

Table 7: Comparison of the predictive accuracy and generalization capability of the 3D-QSAR models. The table compares the experimental or actual bioactivity values for 50 known COX-2 inhibitors with the predicted bioactivities obtained from the developed 3D-QSAR models, demonstrating each model's ability to capture 3D structural features relevant to COX-2 inhibition.

Sl. No.	Compounds	Dataset	Experimental, Log ₁₀ (IC ₅₀), nM	Predicted Bioactivity, Log ₁₀ (IC ₅₀), nM	
				PCR	kNN
1	CHEMBL1096723	Training	-0.24	-0.30	0.89
2	CHEMBL1096724	Test	-0.19	1.02	1.05
3	CHEMBL1095083	Validation	-0.05	1.68	0.86
4	CHEMBL1094412	Validation	0.19	1.55	1.05
5	CHEMBL1094097	Test	0.24	0.39	0.34
6	CHEMBL1095082	Test	0.26	1.12	0.63
7	CHEMBL1096037	Validation	0.31	1.13	0.89
8	CHEMBL1096384	Training	0.32	0.20	0.75
9	CHEMBL1096725	Validation	0.45	-0.29	0.32
10	CHEMBL71833	Training	0.51	0.16	0.89
11	CHEMBL1096385	Training	0.55	1.69	1.28
12	CHEMBL1094096	Test	0.60	1.67	0.86
13	CHEMBL1095094	Training	0.61	1.02	2.48
14	CHEMBL1095084	Training	0.74	1.13	0.78
15	CHEMBL3361124	Training	0.98	1.74	0.78
16	CHEMBL6	Training	1.00	1.65	2.29
17	CHEMBL1224098	Training	1.04	0.97	0.63
18	BDBM227581	Validation	1.15	3.90	1.07
19	CHEMBL1835477	Training	1.18	2.36	0.97
20	BDBM227580	Training	1.23	2.43	2.11
21	BDBM227577	Test	1.28	2.95	3.25
22	CHEMBL187137	Training	1.30	2.25	2.14
23	BDBM227570	Training	1.38	3.16	2.35
24	CHEMBL184737	Test	1.48	2.11	3.24
25	CHEMBL175686	Training	1.49	2.62	1.96
26	CHEMBL360720	Training	1.51	2.44	1.92
27	BDBM222445	Training	3.19	3.05	3.34
28	CHEMBL365497	Training	3.20	3.20	2.14
29	BDBM123807	Training	3.23	2.01	3.32
30	CHEMBL1835475	Validation	3.23	2.55	3.28
31	BDBM222446	Training	3.26	3.71	3.68
32	CHEMBL9192	Training	3.27	2.67	3.26
33	BDBM222441	Training	3.28	2.68	2.60
34	CHEMBL259401	Training	3.28	2.52	2.10
35	CHEMBL56367	Training	3.28	2.81	2.23
36	CHEMBL2147422	Training	3.30	3.04	3.38
37	CHEMBL118	Test	3.31	2.46	3.28
38	BDBM50106737	Training	3.36	3.21	3.38
39	CHEMBL25	Test	3.38	3.04	3.56

40	CHEMBL412610	Test	3.42	2.84	2.36
41	BDBM222448	Training	3.44	1.80	3.21
42	CHEMBL70835	Training	3.45	2.77	2.25
43	ZINC27015737	Training	3.47	3.20	3.56
44	CHEMBL187481	Training	3.49	3.02	3.31
45	CHEMBL364663	Training	3.51	2.57	3.59
46	BDBM50035979	Training	3.52	2.51	2.62
47	CHEMBL4276853	Training	3.84	3.99	3.24
48	CHEMBL1702044	Training	3.86	3.10	3.39
49	CHEMBL365288	Training	3.90	3.40	3.40
50	CHEMBL365615	Training	4.00	4.39	3.50

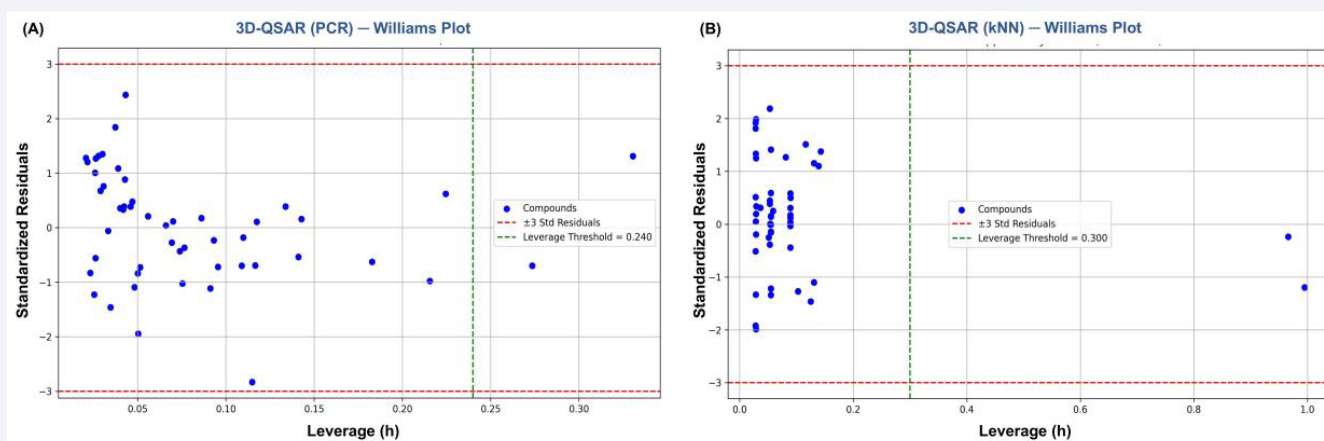


Figure 10 Williams plots for the 3D-QSAR models. **(A)** Standardized Residuals vs. Leverage values for the PCR model. The horizontal dashed lines represent the ± 3 standardized residual threshold, indicating the cut-off for response outliers. The vertical dashed line indicates the leverage threshold ($h = 0.240$). **(B)** Standardized Residuals vs. Leverage values for the kNN model. The vertical leverage threshold ($h = 0.300$) and horizontal residual limits (± 3) are indicated by dashed lines. The distribution confirms that over 95% of the compounds lie within the defined applicability domain for both models.

prioritized through the 3D kNN regression model, were subjected to this analysis alongside celecoxib, a clinically-approved COX-2 inhibitor, included as a reference compound. The evaluation included key parameters such as gastrointestinal (GI) absorption, blood-brain barrier (BBB) permeability, P-glycoprotein (P-gp) substrate status, cytochrome P450 (CYP) isoform (CYP1A2, CYP2C19, CYP2C9, CYP2D6, and CYP3A4) inhibition, skin permeability ($\text{Log } K_p$), and compliance with established drug-likeness rules namely, Lipinski (Pfizer), Ghose (Amgen), Veber (GSK), Egan (Pharmacia), and Muegge (Bayer). Additionally, bioavailability score, PAINS (Pan Assay Interference Structures) alerts, Brenk structural alerts, and lead-likeness violations were also assessed. Among the 20 candidates, only 4 compounds (C11, C27, C31, and C37) demonstrated favorable ADMET properties,

with acceptable drug-likeness and toxicological profiles (Table 10). The remaining compounds exhibited significant violations of multiple criteria, including excessive Lipinski rule violations, high PAINS/Brenk alerts, and low bioavailability scores, and were therefore excluded from subsequent analyses (**Supplementary Table 6**). Celecoxib, the reference compound, exhibited high GI absorption with no BBB permeability or P-gp interaction. It inhibited CYP1A2 and CYP2C9, which may indicate potential drug-drug interaction liabilities. Celecoxib passed most drug-likeness filters but violated the Ghose rule and showed a lead-likeness violation, consistent with its known pharmacokinetic complexity.

The compound C11, despite low GI absorption, displayed an otherwise favorable pharmacokinetic profile, including no CYP inhibition, no P-gp substrate behavior,

Table 8: The predicted COX-2 inhibitory potential of the plant-derived phytochemicals using 2D-QSAR models. The table presents the COX-2 inhibitory potential of the compounds, expressed as $\text{Log}_{10}(\text{IC}_{50})$ in nM, providing a preliminary assessment of their bioactivity and aiding in identifying promising COX-2 inhibitors.

Identifier	Compounds	Influential Molecular Descriptors					MLR Prediction	PLS Prediction
		SsssCHcount	SssNHcount	T_N_O_5	SsCH3count	T_2_C_6		
C1	3, 4-Difluorobenzylidene curcumin	0	0	0	1	51	1.65	1.78
C2	4', 5-DI-N-Propoxy-7-gem-difluoromethylenated genistein	1	0	0	2	35	2.17	2.06
C3	4', 5-Di-O-acetyl Genistein	0	0	0	2	34	2.57	2.53
C4	4', 7-DI-Gem-difluoromethylenated genistein	2	0	0	0	24	0.75	0.74
C5	4-(4-Hydroxybenzylidene) curcumin	0	0	0	2	70	1.93	2.03
C6	5'-Prenyl-glabridin	1	0	0	4	35	3.40	3.09
C7	6, 8-Di-DMA-chrysin	0	0	0	4	57	3.39	3.24
C8	6, 8-dibenzyl-chrysin	0	0	0	0	85	0.43	0.80
C9	6-isopropyl-chrysin	1	0	0	2	29	2.28	2.15
C10	Apigenin 7, 4'-dimethyl ether	0	0	0	2	27	2.69	2.63
C11	Apigenin 7-sulfate	0	0	0	0	28	1.45	1.59
C12	Apigenin-7-olate	0	0	0	0	24	1.52	1.65
C13	Apigenin	0	0	0	0	24	1.52	1.65
C14	Chrysin-7-O-beta-D-glucoronide	5	0	0	0	38	-0.64	-0.82
C15	Chrysine A	0	0	0	4	39	3.71	3.49
C16	Chrysine B	0	0	0	4	42	3.65	3.45
C17	Chrysin	0	0	0	0	24	1.52	1.65
C18	Curcumin 5-8	1	1	1	3	34	3.83	3.60
C19	Curcumin pyrazole	0	0	0	2	32	2.60	2.56
C20	Curcumin sulfate	0	0	0	2	35	2.55	2.52
C21	Curcumin-d6	0	0	0	2	30	2.64	2.59
C22	Curcumin	0	0	0	2	30	2.64	2.59
C23	Didemethyl Curcumin	0	0	0	0	28	1.45	1.59
C24	Difluorinated Curcumin	0	0	0	2	70	1.93	2.03
C25	Epigallocatechin Gallate	2	0	0	0	54	0.22	0.32
C26	Fisetin 3-glucoside	5	0	0	0	42	-0.72	-0.88
C27	Fisetin 3-methyl ether	0	0	0	1	28	2.06	2.10
C28	Fisetin	0	0	0	0	24	1.52	1.65
C29	Genistein 7-sulfate	0	0	0	0	25	1.50	1.63
C30	Genistein-d4	0	0	0	0	21	1.57	1.69
C31	Genistein	0	0	0	0	21	1.57	1.69
C32	Glabridin dimethyl ether	1	0	0	4	26	3.56	3.22
C33	Glabridin	1	0	0	2	25	2.35	2.20
C34	Hydroxytyrosol 3'-sulfate	0	0	0	0	5	1.86	1.91
C35	Hydroxytyrosol Acetate	0	0	0	1	8	2.42	2.38
C36	Hydroxytyrosol Decanoate	0	0	0	1	11	2.36	2.34
C37	Hydroxytyrosol	0	0	0	0	0	1.95	1.98
C38	Isomangiferin	5	0	0	0	27	-0.45	-0.67
C39	Mono-Propargyl Curcumin	1	0	0	2	38	2.11	2.02
C40	Monoglycinoyl curcumin	0	0	0	2	35	2.55	2.52
C41	Morusin	0	0	0	4	62	3.30	3.17
C42	N-(3-Nitrophenylpyrazole) Curcumin	0	0	2	2	64	1.21	1.28
C43	Naringenin	1	0	0	0	24	1.14	1.19
C44	O7-Nitrooxyethyl chrysin	0	0	0	0	30	1.41	1.56
C45	Olmelin	0	0	0	1	22	2.17	2.19
C46	Quercetin 3-sulfate	0	0	0	0	36	1.30	1.48
C47	Quercetin	0	0	0	0	24	1.52	1.65
C48	Racemosic acid	10	0	0	1	23	-1.68	-2.37
C49	Resveratrol Hexanoic Acid	0	0	0	0	28	1.45	1.59
C50	Resveratrol-3-O-sulfate	0	0	0	0	25	1.50	1.63
C51	Resveratrol	0	0	0	0	20	1.59	1.70
C52	Tetrahydro Curcumin-d6	0	0	0	2	24	2.75	2.67
C53	Trihydroxychalcone	0	0	0	0	22	1.55	1.67
C54	[18FP]-curcumin	0	0	0	2	35	2.55	2.52

Table 9: The predicted COX-2 inhibitory potential of the phytochemicals using 3D-QSAR models. The table presents the COX-2 inhibitory potential of the compounds, expressed as $\text{Log}_{10}(\text{IC}_{50})$ in nM, providing a structure-based assessment of their bioactivity and aiding compound prioritization for further experimental validation.

Compounds/ Identifier	3D Field Descriptors				kNN Prediction	3D Field Descriptors				PCR Prediction
	S_{659}	S_{1369}	S_{2747}	S_{2825}		S_{884}	E_{1977}	E_{2605}	E_{951}	
C1	-0.0004	-0.0845	-0.0235	-0.1409	2.14	-0.0011	5.5800	-0.0965	0.1726	2.42
C2	-0.0007	-0.2202	-0.0033	30.0000	3.60	-0.0031	4.5485	0.1921	-0.1444	2.47
C3	-0.0006	-0.2701	-0.0028	30.0000	3.60	-0.0023	-1.0260	0.0175	0.4061	2.74
C4	-0.0004	-0.2295	-0.0032	-0.8740	3.24	-0.0017	0.8887	0.2088	0.0618	3.06
C5	-0.0007	-0.0415	-0.0068	-0.2589	2.26	-0.0024	-1.0613	0.0886	0.0301	2.86
C6	-0.0009	-0.1243	-0.0054	30.0000	3.61	-0.0043	-3.5618	0.1197	0.0729	2.61
C7	-0.0004	-0.2070	-0.0064	0.7079	2.16	-0.0012	-2.9802	-0.0397	0.1039	3.10
C8	-0.0008	-0.1953	-0.0067	1.7130	2.26	-0.0031	1.1597	0.0696	0.1025	2.49
C9	-0.0004	-0.2927	-0.0045	-0.7278	2.16	-0.0013	1.0192	0.0605	-0.2191	2.99
C10	-0.0004	-0.2405	-0.0055	-0.8790	2.15	-0.0013	-0.3481	-0.0353	0.4363	2.85
C11	-0.0003	-0.2198	-0.0056	-0.9258	2.15	-0.0012	-0.3561	0.1157	-0.1456	3.18
C12	-0.0003	-0.1763	-0.0042	-0.3000	2.15	-0.0010	0.8792	0.0382	0.3310	2.96
C13	-0.0003	-0.1757	-0.0045	-0.3607	2.15	-0.0010	0.1790	0.0329	0.1732	3.02
C14	-0.0007	-0.1672	-0.0043	-0.3970	2.27	-0.0026	2.4602	-0.1102	0.2030	2.24
C15	-0.0005	-0.2214	-0.0041	11.2645	3.55	-0.0016	10.0000	0.1414	0.5513	2.31
C16	-0.0004	-0.2390	-0.0055	30.0000	3.59	-0.0013	8.9851	-0.0895	0.0610	2.16
C17	-0.0003	-0.1683	-0.0038	-0.3028	2.15	-0.0010	-0.1684	0.0261	0.1326	3.04
C18	-0.0006	-0.1743	-0.0024	30.0000	3.60	-0.0029	0.6465	0.1236	0.4720	2.61
C19	-0.0009	-0.0486	-0.0050	-0.1110	2.26	-0.0040	1.1840	0.1179	0.1802	2.32
C20	-0.0004	-0.0436	-0.0643	-0.1221	2.13	-0.0010	-1.8768	0.1245	0.0101	3.31
C21	-0.0003	-0.0285	-0.0498	-0.2838	2.13	-0.0007	0.6435	0.0759	0.1158	3.13
C22	-0.0007	-0.0508	-0.0035	-0.1794	2.26	-0.0028	3.6361	0.1290	0.0875	2.48
C23	-0.0007	-0.0615	-0.0030	-0.1577	2.26	-0.0030	0.1564	0.1226	0.0500	2.68
C24	-0.0007	-0.0413	-0.0065	30.0000	3.60	-0.0025	0.5402	-0.1155	0.1213	2.41
C25	-0.0004	-0.0773	-0.0155	30.0000	3.59	-0.0015	8.0101	0.0870	0.2276	2.44
C26	-0.0005	-0.0613	-0.0077	30.0000	3.60	-0.0019	-10.0000	-0.2234	0.1454	3.12
C27	-0.0003	-0.1265	-0.0044	-0.7678	2.14	-0.0010	0.8575	-0.0497	0.3144	2.84
C28	-0.0003	-0.1647	-0.0040	-0.1144	2.15	-0.0009	0.4472	0.0070	0.3235	2.98
C29	-0.0004	-0.2492	-0.0030	-0.4490	3.24	-0.0015	2.1935	0.1672	-0.2144	3.00
C30	-0.0004	-0.2531	-0.0021	-0.4082	3.24	-0.0016	0.9908	0.0836	0.2365	2.89
C31	-0.0004	-0.2665	-0.0019	-0.4385	3.24	-0.0015	0.5623	0.0860	0.2403	2.95
C32	-0.0007	-0.2378	-0.0039	30.0000	3.60	-0.0030	-2.8621	-0.0044	0.0719	2.70
C33	-0.0006	-0.2605	-0.0030	30.0000	3.60	-0.0026	-2.9131	0.0736	0.1186	2.90
C34	-0.0003	-0.0728	-0.0025	-0.0615	2.15	-0.0009	2.7214	0.1506	-0.1892	3.09
C35	-0.0002	-0.0897	-0.0035	-0.1093	2.15	-0.0007	1.6988	0.0846	-0.2969	3.11
C36	-0.0002	-0.0301	-0.0245	-0.3461	2.13	-0.0006	-1.3949	0.0767	-0.1027	3.33
C37	-0.0002	-0.0900	-0.0016	-0.0702	2.14	-0.0007	2.4966	0.0358	-0.2575	3.00
C38	-0.0006	-0.0690	-0.0075	30.0000	3.60	-0.0026	2.5698	-0.1148	-0.1475	2.28
C39	-0.0010	-0.1073	-0.0040	30.0000	3.61	-0.0046	8.7216	0.2099	0.2566	1.77
C40	-0.0008	-0.1172	-0.0051	-0.1494	2.26	-0.0031	0.8662	0.1996	0.1663	2.70
C41	-0.0005	-0.0937	-0.0047	8.0188	3.55	-0.0018	5.6683	-0.0465	0.0453	2.34
C42	-0.0008	-0.0549	-0.0095	-0.7274	2.26	-0.0033	5.2120	-0.0082	0.4572	2.01
C43	-0.0003	-0.2212	-0.0047	1.9620	2.16	-0.0009	2.2249	-0.0727	0.1258	2.76
C44	-0.0004	-0.2358	-0.0050	0.6716	3.24	-0.0015	3.3180	0.0506	-0.4754	2.79
C45	-0.0004	-0.2664	-0.0021	-0.7187	3.24	-0.0016	-2.2652	0.0424	0.2656	3.05
C46	-0.0003	-0.2312	-0.0054	30.0000	3.59	-0.0010	1.5270	-0.0012	0.3549	2.85
C47	-0.0003	-0.1662	-0.0044	-0.1672	2.15	-0.0009	1.1234	0.0559	0.2457	3.00
C48	-0.0006	-0.3530	-0.0096	-0.7325	2.27	-0.0022	-0.4622	0.3133	-0.1638	3.21
C49	-0.0010	-0.0697	-0.0045	-0.0600	3.33	-0.0042	-1.0424	0.1040	0.1234	2.42
C50	-0.0006	-0.0607	-0.0032	-0.0776	2.27	-0.0028	1.4460	0.0149	0.1736	2.46
C51	-0.0005	-0.0929	-0.0023	-0.0373	3.24	-0.0021	0.0238	-0.0038	0.2920	2.69
C52	-0.0003	-0.0452	-0.0179	-0.3841	2.13	-0.0007	7.4214	0.0376	0.1093	2.60
C53	-0.0004	-0.1329	-0.0019	-0.1277	3.24	-0.0018	2.1260	0.1240	-0.4083	2.90
C54	-0.0008	-0.0974	-0.0064	-0.0884	2.26	-0.0032	-0.3882	0.2541	0.1146	2.84

Table 10: Predicted AMDET properties of the QSAR-prioritized compounds. The table shows the predicted pharmacokinetic, drug-likeness, and medicinal chemistry parameters for 20 phytocompounds selected by kNN prediction, along with the reference drug, celecoxib, using the SwissADME server.

Molecule	Pharmacokinetics											Druglikeness						Medicinal Chemistry		
	GI absorption	BBB permeant	P-gp substrate	CYP1A2 inhibitor	CYP2C19 inhibitor	CYP2C9 inhibitor	CYP2D6 inhibitor	CYP3A4 inhibitor	CYP 2D6_inhibition	CYP 2C19_inhibition	log K _p (cm/s)	Lipinski (Pfizer)	Ghose (Amgen)	Veber (GSK)	Egan (Pharmacia)	Muegge (Bayer)	Bioavailability Score	PAINS	Brenk	Leadlikeness
Celecoxib	High	X	X	✓	X	✓	X	X	X	✓	-6.21	0	1	0	0	0	0.55	0	0	1
C1	Low	X	X	X	X	✓	X	✓	X	✓	-5.48	0	3	0	1	1	0.55	0	3	3
C5	High	X	X	X	X	✓	X	✓	X	✓	-5.65	0	1	0	0	0	0.55	0	3	3
C6	High	✓	X	X	✓	✓	X	✓	X	✓	-4.56	0	0	0	0	1	0.55	0	1	2
C7	High	X	X	X	✓	X	X	X	X	✓	-3.8	0	1	0	1	1	0.55	0	1	2
C8	High	X	X	X	X	X	X	X	X	✓	-4.08	0	2	0	1	1	0.55	0	0	2
C10	High	✓	X	✓	✓	✓	✓	✓	X	✓	-5.51	0	0	0	0	0	0.55	0	0	1
C11	Low	X	X	X	X	X	X	X	X	X	-6.71	0	0	1	1	0	0.56	0	1	1
C14	Low	X	✓	X	X	X	X	X	X	X	-7.89	0	0	1	1	1	0.11	0	0	1
C20	Low	X	X	X	X	✓	✓	X	X	X	-7.18	0	0	1	1	0	0.56	0	3	2
C24	High	X	X	X	X	✓	X	✓	X	✓	-5.38	0	3	0	0	1	0.55	0	3	3
C25	Low	X	X	X	X	X	X	X	X	✓	-8.27	2	0	1	1	3	0.17	1	1	1
C26	Low	X	X	X	X	X	X	X	X	✓	-8.92	2	0	1	1	3	0.17	1	1	1
C27	High	X	X	✓	X	X	✓	✓	X	✓	-6.5	0	0	0	0	0	0.55	1	1	0
C31	High	X	X	✓	X	X	✓	✓	X	✓	-6.05	0	0	0	0	0	0.55	0	0	0
C32	High	✓	✓	X	✓	✓	✓	✓	X	✓	-5.23	0	0	0	0	0	0.55	0	0	2
C33	High	✓	✓	✓	✓	✓	✓	✓	X	✓	-5.52	0	0	0	0	0	0.55	0	0	1
C36	High	✓	X	✓	X	X	✓	X	X	✓	-5.39	0	0	1	0	0	0.55	1	1	2
C37	High	X	X	X	X	X	X	X	X	✓	-7.75	0	1	0	0	1	0.55	1	1	1
C38	Low	X	X	X	X	X	X	X	X	✓	-9.14	2	1	1	1	3	0.17	1	2	1
C48	Low	X	X	X	X	X	X	X	X	X	-11.3	3	2	1	1	4	0.11	0	1	1

“✓” = Yes, “X” = No

and full compliance with major drug-likeness filters, except for single violations in the Veber and Egan filters. The presence of one Brenk alert and one lead-likeness violation warrant consideration but does not preclude further optimization. The compound C27 demonstrated high GI absorption and inhibited CYP1A2, CYP2D6, and CYP3A4, suggesting a higher risk of metabolic interactions. It also showed PAINS and Brenk alerts, which may raise concerns regarding assay interference or off-target effects. Nonetheless, it complied with all major drug-likeness filters and did not violate lead-likeness criteria, making it a viable candidate for structural refinement. The compound C31 emerged as the most pharmacokinetically promising,

combining high GI absorption, no interactions with P-gp or CYP2 family enzymes, and no alerts under PAINS or Brenk filters. Importantly, it did not violate any of the drug-likeness or lead-likeness rules, indicating a well-balanced physicochemical profile with a strong potential for further development. The compound C37 also exhibited high GI absorption and a clean CYP/P-gp interaction profile. However, it presented multiple structural alerts (PAINS and Brenk) and violations in Ghose, Muegge, and lead-likeness criteria, which may limit its drug-likeness and synthetic feasibility. Its low Log K_p value suggests poor skin permeability, although this may not be a primary concern for orally administered compounds.

To complement the SwissADME profiling, additional ADMET parameters were predicted using the PreADME server, specifically assessing human intestinal absorption (HIA%), Caco-2 and MDCK cell permeability, plasma protein binding (PPB%), and potential for hERG channel inhibition, which is a key predictor of cardiotoxicity (Table 11). Celecoxib, the reference molecule, showed a high HIA of 96.69%, consistent with its known oral bioavailability. Its Caco-2 permeability (0.499) and MDCK permeability (45.05) were within acceptable ranges, indicating moderate passive diffusion across intestinal and renal membranes. It demonstrated high plasma protein binding (91.08%) and was predicted to pose a medium risk of hERG inhibition, which may be a concern for cardiac safety. Among the selected phytocompounds, C11 showed an HIA of 83.98%, with moderate Caco-2 (1.260) and low MDCK permeability (10.73). Despite this, it exhibited complete plasma protein binding (100%) and a medium hERG risk, which necessitates caution due to potential off-target cardiac effects. The compound C27 showed favorable absorption and permeability, with an HIA of 88.19%, high Caco-2 (3.931) and MDCK (89.13) permeability, and moderate PPB (83.47%). It also exhibited a medium risk of hERG inhibition, aligning with its PAINS and CYP inhibition alerts observed in earlier screening. C31 continued to demonstrate a robust pharmacokinetic profile with HIA of 88.12%, the highest Caco-2 permeability (5.747) among all tested compounds, and moderate MDCK permeability (39.43). Its PPB of 89.74% suggests suitable plasma retention, and it, too, presented a medium hERG risk. In contrast, the compound C37, while achieving high MDCK permeability (93.87) and Caco-2 permeability (19.37), had the lowest HIA (79.62%) among the five compounds and an unusually low PPB (3.04%), indicating weak plasma protein interaction. Interestingly, it was the only compound to show a low risk of hERG inhibition, potentially reducing concerns regarding cardiac toxicity.

These additional pharmacokinetic and toxicity assessments reinforce the earlier SwissADME findings. The compound C31 again emerged as the most balanced

candidate, combining high intestinal absorption, favorable membrane permeability, adequate plasma protein binding, and acceptable hERG risk. Compound C27, despite strong permeability, requires structural refinement to address potential CYP-mediated drug-drug interactions and hERG-related cardiotoxicity. The compounds, C11 and C37, demonstrated mixed profiles and may warrant selective optimization based on their respective liabilities, particularly regarding protein binding and cardiac safety. To further substantiate the safety profiles of the top four kNN-selected compounds, we conducted a comprehensive *in silico* toxicity risk assessment using the ProTox-III server. This platform integrates 61 structure-activity relationship (SAR)-based predictive models, derived from extensive rodent model datasets. The evaluation encompassed acute toxicity (LD_{50}), organ-specific toxicity, toxicological endpoints, molecular initiating events, and both Tox21 nuclear receptor signaling and stress response pathways. All four phytocompounds, i.e., C11, C27, C31, and C37, were classified under Toxicity Class V, indicating they “may be harmful if swallowed” based on the Globally Harmonized System (GHS) of chemical classification. The predicted median lethal doses (LD_{50}) for all compounds ranged from 2500 to 5000 mg/kg (oral, rodent), suggesting relatively low acute oral toxicity.

A detailed breakdown of the ProTox-III predictions, provided in Table 12. The result showed that the majority of these compounds exhibited acceptable safety margins across various organ toxicity profiles and toxicological endpoints. Some mild alerts were predicted for interactions with nuclear receptor pathways or stress response elements (e.g., oxidative stress or mitochondrial dysfunction), but none were flagged as high-confidence liabilities. Importantly, apart from C27, no compound showed strong predictions for mutagenicity or carcinogenicity, two key exclusionary criteria in early drug discovery. Nevertheless, certain molecular initiating events (MIEs) and moderate activity within stress response pathways suggest potential off-target interactions, underscoring the need for further structural optimization to mitigate residual toxicity risks.

Table 11: Additional ADMET parameters of selected compounds from the PreADMET server. HIA: human intestinal absorption (0–20% = poorly absorbed, 20–70% = moderate, 70–100% = well absorbed); Caco-2 and MDCK permeability: <4 = low, 4–70 = moderate, >70 = high permeability (nm/sec); PPB: plasma protein binding, with >90% indicating strong binding and <90% indicating weak binding; hERG: prediction of cardiac toxicity potential.

Compound	HIA (%)	Caco2	MDCK	PPB (%)	hERG Inhibition
Celecoxib	96.69	0.499	45.049	91.08	Medium Risk
C11	83.98	1.260	10.725	100.00	Medium Risk
C27	88.19	3.931	89.125	83.47	Medium Risk
C31	88.12	5.747	39.433	89.74	Medium Risk
C37	79.62	19.365	93.871	3.04	Low Risk

Table 12: Predicted rodent toxicity risk profiles of the prioritized compounds from the ProTox-III server. The results are based on 61 rodent model-based SAR-predictive models covering acute toxicity, organ toxicity, toxicological endpoints, molecular initiating events, adverse outcome pathways (Tox21), and toxicity targets. The values in parentheses represent the confidence probability of each prediction, expressed as a fraction of 1 (i.e., 1 = 100%).

Classification	Target	Compounds			
		C11	C27	C31	C37
	Predicted LD ₅₀ (mg/kg)	3919	5000	2500	2820
	Predicted Toxicity Class	V	V	V	V
Organ toxicity	Target	Prediction (Probability)			
	Hepatotoxicity	Inactive (0.70)	Inactive (0.73)	Inactive (0.69)	Inactive (0.91)
	Neurotoxicity	Inactive (0.90)	Inactive (0.86)	Inactive (0.85)	Inactive (0.82)
	Nephrotoxicity	Active (0.51)	Active (0.59)	Active (0.63)	Inactive (0.54)
	Respiratory toxicity	Active (0.54)	Active (0.82)	Active (0.84)	Inactive (0.59)
	Cardiotoxicity	Inactive (0.51)	Inactive (0.71)	Inactive (0.88)	Inactive (0.51)
Toxicity endpoints	Carcinogenicity	Inactive (0.72)	Active (0.59)	Inactive (0.69)	Inactive (0.75)
	Immunotoxicity	Inactive (0.99)	Active (0.81)	Inactive (0.97)	Inactive (0.97)
	Mutagenicity	Inactive (0.57)	Active (0.69)	Inactive (0.74)	Inactive (0.71)
	Cytotoxicity	Inactive (0.64)	Inactive (0.91)	Inactive (0.91)	Inactive (0.9)
	BBB-barrier	Active (0.67)	Inactive (0.55)	Active (0.56)	Inactive (0.54)
	Ecotoxicity	Inactive (0.5)	Inactive (0.58)	Active (0.53)	Inactive (0.72)
	Clinical toxicity	Inactive (0.51)	Active (0.5)	Inactive (0.54)	Inactive (0.53)
	Nutritional toxicity	Inactive (0.66)	Active (0.56)	Active (0.67)	Inactive (0.72)
Tox21-Nuclear receptor signalling pathways	Aryl Hydrocarbon Receptor (AhR)	Inactive (0.73)	Active (0.86)	Active (1)	Inactive (0.88)
	Androgen Receptor (AR)	Inactive (0.97)	Inactive (0.99)	Inactive (0.99)	Inactive (0.98)
	Androgen Receptor Ligand Binding Domain (AR-LBD)	Inactive (0.98)	Inactive (0.88)	Inactive (0.99)	Inactive (1)
	Aromatase	Inactive (0.87)	Inactive (0.86)	Active (0.82)	Inactive (0.96)
	Estrogen Receptor Alpha (ER)	Inactive (0.55)	Active (0.66)	Active (1)	Inactive (0.87)
	Estrogen Receptor Ligand Binding Domain (ER-LBD)	Inactive (0.67)	Active (0.81)	Active (1)	Inactive (0.86)
Tox21-Stress response pathways	Peroxisome Proliferator Activated Receptor Gamma (PPAR-Gamma)	Inactive (0.92)	Inactive (0.96)	Inactive (0.75)	Inactive (0.99)
	Nuclear factor (erythroid-derived 2)-like 2/ antioxidant responsive element (nrf2/ARE)	Inactive (0.82)	Inactive (0.91)	Inactive (0.99)	Inactive (0.95)
	Heat shock factor response element (HSE)	Inactive (0.82)	Inactive (0.91)	Inactive (0.99)	Inactive (0.95)
	Mitochondrial Membrane Potential (MMP)	Inactive (0.65)	Active (0.84)	Active (0.99)	Inactive (0.86)
	Phosphoprotein (Tumor Suppressor) p53	Inactive (0.72)	Inactive (0.88)	Active (1)	Inactive (0.96)
	ATPase family AAA domain-containing protein 5 (ATAD5)	Inactive (0.89)	Active (0.51)	Active (1)	Inactive (0.98)
Molecular Initiating Events	Thyroid hormone receptor alpha (THRα)	Inactive (0.81)	Inactive (0.92)	Inactive (0.9)	Inactive (0.9)
	Thyroid hormone receptor beta (THRβ)	Inactive (0.74)	Inactive (0.97)	Inactive (0.78)	Inactive (0.78)
	Transthyretin (TTR)	Active (0.77)	Active (0.88)	Inactive (0.97)	Inactive (0.97)
	Ryanodine receptor (RYR)	Inactive (0.87)	Inactive (0.92)	Inactive (0.98)	Inactive (0.98)
	GABA receptor (GABAR)	Inactive (0.76)	Inactive (0.64)	Inactive (0.96)	Inactive (0.96)
	Glutamate N-methyl-D-aspartate receptor (NMDAR)	Inactive (0.91)	Inactive (0.99)	Inactive (0.92)	Inactive (0.92)
	alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionate receptor (AMPA)	Inactive (0.99)	Inactive (1)	Inactive (0.97)	Inactive (0.97)
	Kainate receptor (KAR)	Inactive (1)	Inactive (1)	Inactive (0.99)	Inactive (0.99)
	Acetylcholinesterase (AChE)	Inactive (0.73)	Inactive (0.55)	Inactive (0.69)	Inactive (0.91)
	Constitutive androstane receptor (CAR)	Inactive (0.97)	Inactive (1)	Inactive (0.98)	Inactive (0.98)
	Pregnane X receptor (PXR)	Inactive (0.5)	Inactive (0.61)	Inactive (0.92)	Inactive (0.92)
	NADH-quinone oxidoreductase (NADHox)	Inactive (0.86)	Inactive (0.84)	Inactive (0.97)	Inactive (0.97)
	Voltage-gated sodium channel (VGSC)	Inactive (0.83)	Inactive (0.94)	Inactive (0.95)	Inactive (0.95)
	Na ⁺ /I ⁻ symporter (NIS)	Inactive (0.91)	Inactive (0.95)	Inactive (0.98)	Inactive (0.98)

Overall, these ProTox-III results reinforce the findings from SwissADME and PreADME, positioning C11 and C31 as the most promising lead compounds, although with scope for medicinal chemistry refinement. The acceptable toxicity risk profiles of all four candidates provide a strong rationale for their progression into experimental/clinical validation, subject to iterative structure-based optimization strategies aimed at improving both efficacy and safety.

Molecular Interactions of the Selected Ligands with COX-2

To elucidate the binding interactions of the top-ranked phytocompounds with the COX-2 active site, a two-stage molecular docking strategy was employed. An initial virtual screening using AutoDock Vina was conducted to prioritize potential ligands, followed by more detailed docking simulations performed using AutoDock 4.2. This approach enabled the selection of four compounds, namely, C11, C27, and C31, based on their favorable docking scores and interaction profiles, along with the reference drug celecoxib. The results of the docking simulations are summarized in Table 13, and the interaction landscapes are visually represented in Figure 11 and Figure 12. The control, celecoxib, demonstrated the most favorable binding profile, with a binding energy of -9.95 kcal/mol and a predicted pKi of 0.05 nM, reflecting its high affinity. It formed five hydrogen bonds with key active-site residues, including Arg120, Phe518, Ser353, Gln192, and Leu352, alongside 14 weak interactions, corroborating its stable and extensive binding.

Among the phytocompounds, compound C11 exhibited a binding energy of -7.92 kcal/mol and a pKi of 1.56 nM, making it the best-performing natural compound in terms of binding affinity. It established five hydrogen bonds, notably with residues Ala199, Gln203, Asn382, and His388, and engaged in four weak interactions, suggesting a relatively focused but potent interaction pattern. The compound C31 followed closely, with a binding energy of -7.08 kcal/mol and a pKi of 6.51 nM. It formed four hydrogen bonds with Trp387, Tyr385, Thr206, and Asn382, and exhibited four additional weak interactions. The involvement of Trp387 and Tyr385, both important for COX-2 selectivity, highlights its potential as a selective binder. The compound C27 showed a binding energy of -6.86 kcal/mol and a pKi of 9.44 nM, forming four hydrogen bonds, primarily with Asn382, Ala199, and His388, and 11 weak interactions, indicating a broader but somewhat less stable binding conformation compared to C11 and C31. Overall, while celecoxib demonstrated superior binding metrics, C11 and C31 emerged as promising phytochemical

leads, capable of engaging the COX-2 binding site through multiple stabilizing interactions. The presence of critical binding residues in hydrogen bonding and hydrophobic contacts further supports their candidacy for structure-based optimization and experimental validation.

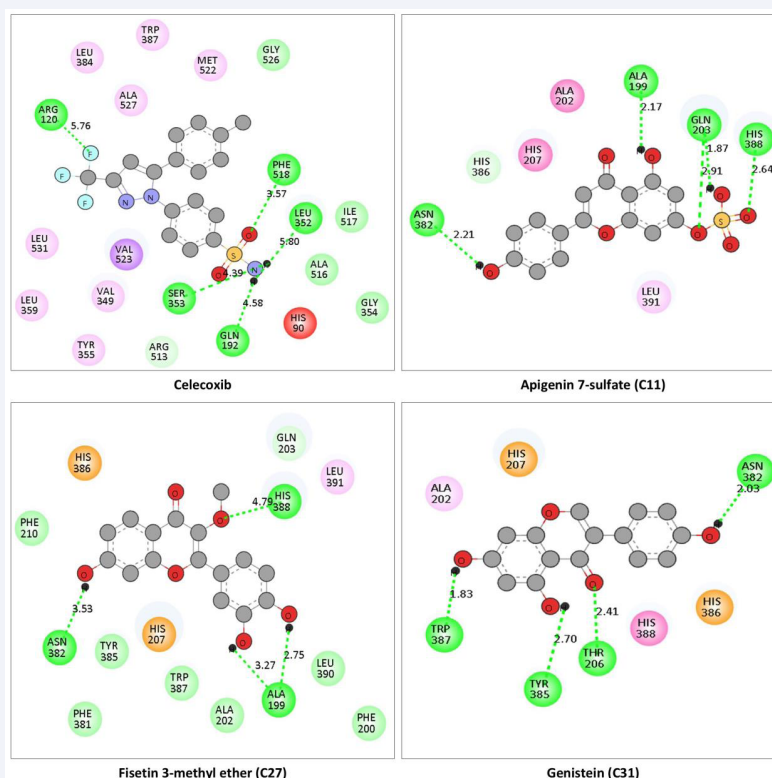
Binding Stability of the Prioritized Protein-Ligand Complexes:

To investigate the dynamic behavior, conformational stability, and binding efficiency of the top two protein-ligand complexes, apigenin-7-sulfate (complex-1) and genistein (complex-2), 50 ns molecular dynamics simulations were performed. As shown in Figure 13, the protein backbone RMSD remained relatively stable in both complexes. The complex-1 exhibited an average protein RMSD of 2.69 Å (range: 1.20-3.20 Å), while complex-2 showed a slightly lower average RMSD of 2.36 Å (range: 1.07-2.84 Å), suggesting comparatively less structural deviation in the genistein-bound system. The ligand RMSD was higher in both complexes than their respective protein RMSDs, with apigenin-7-sulfate showing an average of 3.79 Å (range: 1.89-5.30 Å) and genistein at 3.08 Å (range: 1.06-3.52 Å), indicating moderate mobility within the binding pocket while maintaining the binding pose throughout the simulation. The RMSF analysis results, as shown in Figure 14, further supported these findings. The average residue-level fluctuation in the protein backbone was 1.13 Å (0.46-4.10 Å) for complex-1 and 0.99 Å (0.42-2.91 Å) for complex-2, with the highest fluctuations localized to loop regions. The ligand RMSF values were lower in both systems, particularly for genistein (0.64 Å, range: 0.56-0.79 Å) compared to apigenin-7-sulfate (0.98 Å, range: 0.80-1.30 Å), indicating that genistein maintained a more rigid and consistent interaction within the active site.

The radius of gyration (Rg) values were consistent across both simulations (Figure 15), reflecting the protein's overall compactness and folding stability. The average Rg for the complex-1 was 24.39 Å (range: 24.00-28.86 Å), remaining stable up to 40 ns before showing slight fluctuations. In comparison, the genistein complex had an average Rg of 24.40 Å (range: 23.98-29.19 Å), indicating comparable structural compactness throughout the trajectory. The MM/GBSA binding free energy calculations revealed a notable difference in binding affinities. Notably, apigenin-7-sulfate demonstrated a significantly more favorable binding energy of -94.19 kcal.mol⁻¹, compared to -76.43 kcal.mol⁻¹ for genistein. This suggests that, despite its slightly higher structural mobility, apigenin-7-sulfate forms stronger intermolecular interactions with key residues in the COX-2 binding site. Collectively, these findings indicate that both ligands form stable

Table 13: Molecular docking results of the selected ligands against COX-2 (PDB ID: 6COX). The table shows binding energy (ΔG , kcal/mol), predicted inhibition constant (K_i , nM), hydrogen-bond-forming residues, and weakly interacting residues (hydrophobic and van der Waals contacts).

Compound	Binding Energy, (Kcal.mol ⁻¹)	Inhibition Constant (pK _i), nM	Hydrogen Bonding Pattern		Other weakly interacting residues
			No. of H-bonds	Residues involved	
Celecoxib	-9.95	0.05	5	Arg120, Phe518, Ser353, Gln192, Leu352	Leu384, Trp387, Met522, Ala527, Gly526, Leu531, Val523, Leu359, Val349, Tyr355, Arg513, Ile517, Ala516, Gly354
C11	-7.92	1.56	5	Ala199, Gln203, Asn382, His388	His386, His207, Ala202, Leu391
C27	-6.86	9.44	4	Asn382, Ala199, His388	His386, Phe210, Gln203, Leu391, Leu390, Phe200, Ala202, Trp387, His207, Tyr385, Phe381
C31	-7.08	6.51	4	Trp387, Tyr385, Thr206, Asn382	Ala202, His207, His388, His386

**Figure 11** 2D interaction plots of Celecoxib (control) and selected phytochemicals (C11, C27, and C31) within the active site of COX-2 (PDB ID: 6COX). The figure illustrates the Hydrogen Bonding and other non-covalent interactions between the ligands and key residues of the COX-2 binding pocket.

complexes with COX-2 under physiological conditions. The MD simulation revealed that genistein exhibits slightly enhanced structural stability, whereas apigenin-7-sulfate shows superior binding affinity. These insights may support the further development of these compounds as potent cyclooxygenase-2 (COX-2) inhibitors.

DISCUSSION

Cyclooxygenase-2 is a well-established molecular target in the pathophysiology of inflammatory disorders, including arthritis, cancer, neuroinflammation, and other chronic diseases. Unlike the constitutively expressed

COX-1, COX-2 is an inducible isoform upregulated at sites of inflammation, making it an attractive target for therapeutic intervention. However, the long-term use of synthetic COX-2 inhibitors has been associated with adverse cardiovascular and gastrointestinal side effects, necessitating a continuous search for safer and more selective alternatives. In this context, plant-derived phytochemicals represent a promising reservoir of bioactive compounds with multi-target potential and improved safety profiles. Additionally, the emergence of computational methods, such as QSAR modelling, molecular docking, and in-silico pharmacokinetics/toxicity

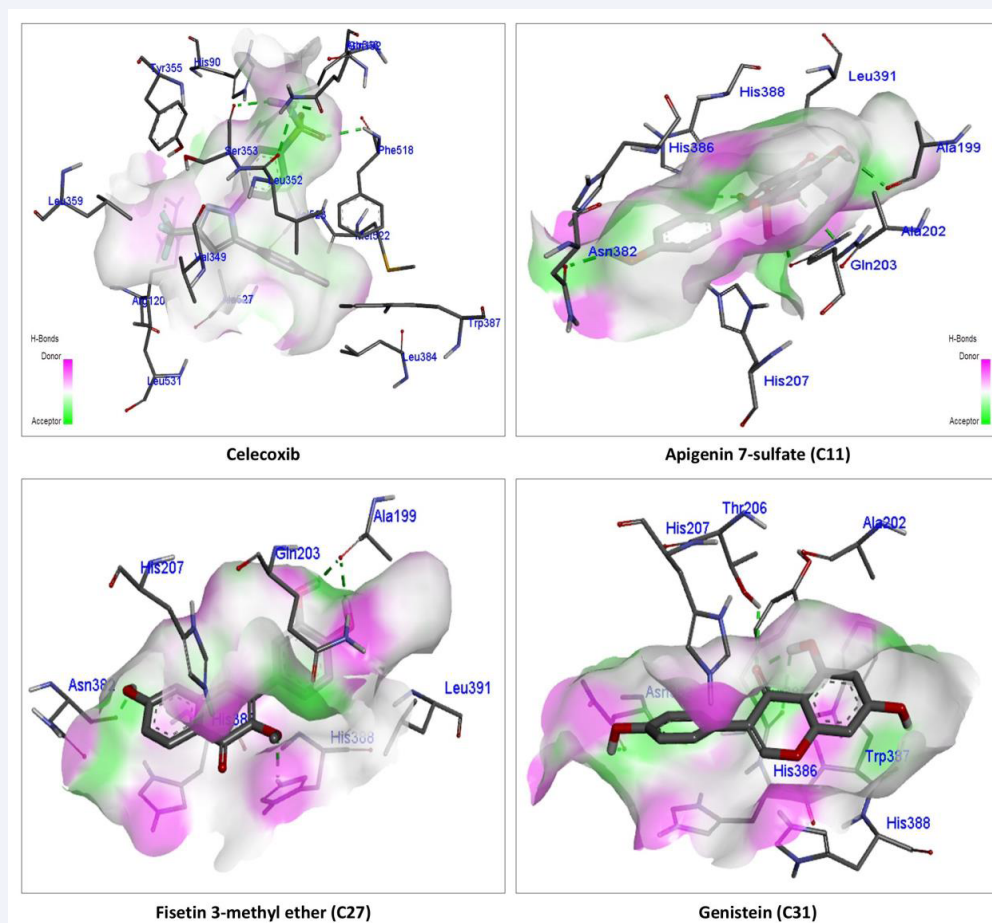


Figure 12 3D binding conformations of Celecoxib (control) and selected phytochemicals (C11, C27, and C31) within the active site of COX-2 (PDB ID: 6COX). The figure displays the spatial orientation of the control drug Celecoxib and three top-ranked phytochemicals docked into the COX-2 active site.

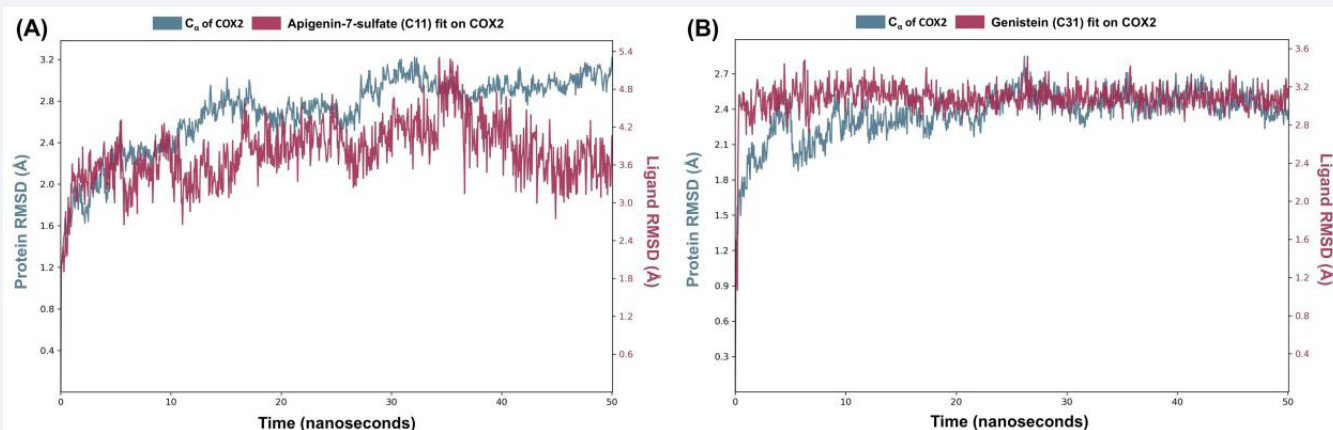


Figure 13 RMSD plots of the protein C_{α} backbone and the ligands over a 50ns molecular dynamics simulation. **(A)** 6COX-Apigenin-7-sulfate (C11) complex showing stable protein structure with moderate ligand mobility. **(B)** 6COX-Genistein (C31) complex exhibiting overall lower deviations and enhanced structural stability.

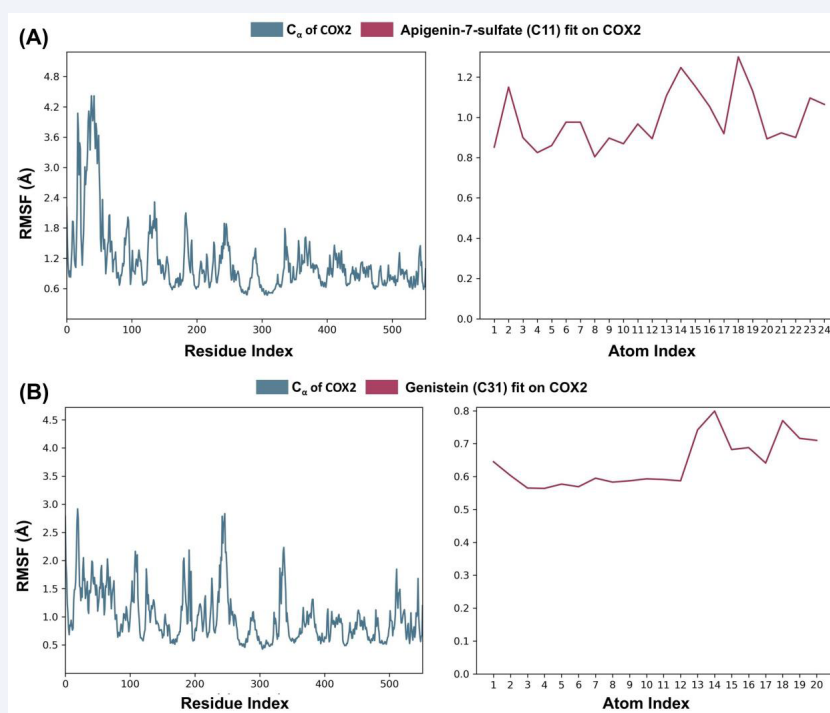


Figure 14 RMSF plots representing residue-level fluctuations of 6COX and ligand atomic mobility over a 50 ns MD simulation. **(A)** Showing the RMSF of 6COX residues (left) and apigenin-7-sulfate (C11) atoms relative to the protein (right). **(B)** Showing the RMSF of 6COX residues (left) and genistein (C31) atoms relative to the protein (right). Lower ligand RMSF values indicate stable interactions within the binding site in both complexes.

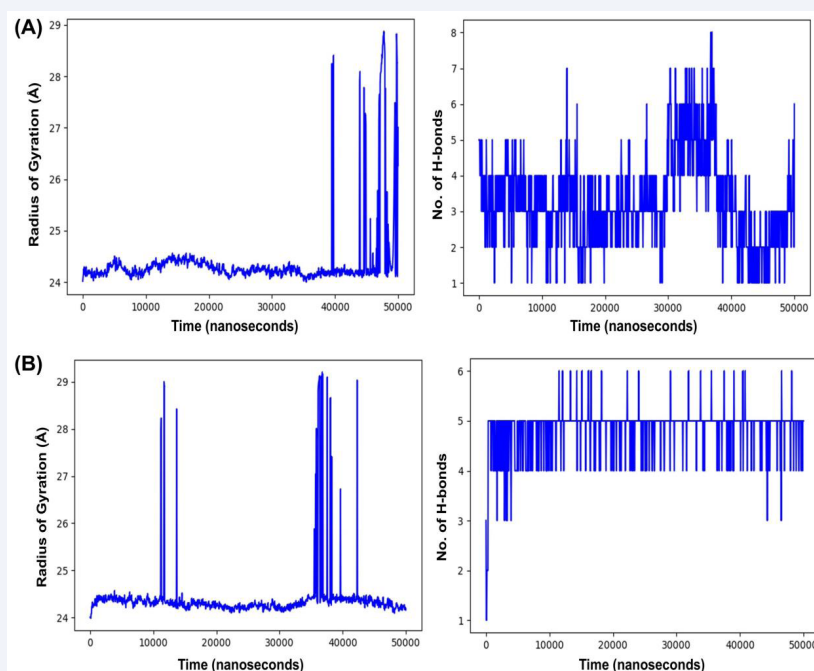


Figure 15 Radius of gyration (Rg) and hydrogen bond profiles of 6COX complexes over 50 ns MD simulation. **(A)** 6COX–Apigenin-7-sulfate (C11): Rg plot (left) shows stable protein compactness; hydrogen bond plot (right) indicates persistent ligand–protein interactions. **(B)** 6COX–Genistein (C31): Rg plot (left) demonstrates consistent structural compactness; hydrogen bond plot (right) shows sustained but slightly fewer H-bond interactions compared to Apigenin-7-sulfate.

profiling, has revolutionized early-stage drug discovery. These tools enable rapid screening of chemical libraries, prediction of molecular interactions, and prioritization of lead compounds based on mechanistic and safety considerations. As these predictive methods continue to evolve with better algorithms and curated datasets, they are increasingly recognized as reliable substitutes for reducing the burden of empirical screening and for guiding rational drug design with enhanced precision.

In the present study, a multi-model QSAR-based screening was employed to prioritize plant-derived COX-2 inhibitors, as drug leads, that not only demonstrate strong binding affinity but also fulfil drug-likeness and safety criteria. Following QSAR-guided prioritization of active molecules, a panel of 20 phytochemicals was subjected to ADMET screening using the SwissADME and PreADME servers. The comprehensive profiling included absorption, metabolism (CYP inhibition profile), distribution properties (e.g., BBB permeability and P-gp substrate), excretion potential (log K_p), and toxicity filters (e.g., PAINS, BRENK, and lead-likeness violations). Among the screened compounds, only C11, C27, C31, and C37 fulfilled most or all of the critical drug-likeness criteria and exhibited acceptable safety profiles, warranting further investigation. Additional parameters, including human intestinal absorption (HIA), Caco-2 and MDCK cell permeability, plasma protein binding (PPB), and hERG inhibition, were also assessed using PreADME. Notably, all four top candidates demonstrated favorable HIA and moderate to high membrane permeability. Although moderate hERG-inhibition risk was observed in most compounds, this is a common alert in early-phase drug discovery and can be addressed through structural refinements. These results underscore the pharmacokinetic feasibility of the selected candidates and support their progression toward preclinical development. All four prioritized compounds were categorized as Toxicity Class V, with predicted LD₅₀ values ranging from 2500 to 5000 mg/kg (oral, rodent). None of the compounds were flagged for severe organ toxicity, mutagenicity, or carcinogenicity, although some showed moderate interactions in nuclear receptor and stress response pathways. These results provide an early indication of safety, although further *in vitro* and *in vivo* toxicological evaluations are needed for confirmation.

The molecular docking results further validated the selection of these compounds. Using AutoDock, detailed molecular interaction analyses were performed at the COX-2 binding site. As expected, the reference molecule, celecoxib, exhibited the highest binding affinity (-9.95 kcal/mol) and the most extensive interaction network, forming five hydrogen bonds with key residues, including

Arg120, Phe518, and Ser353. Among the phytochemicals, C11 emerged as the strongest binder (-7.92 kcal/mol; pK_i 1.56 nM), forming five hydrogen bonds, notably with Ala199 and His388, while also showing minimal off-target toxicity. The compounds C31 and C27 also displayed meaningful interactions within the COX-2 binding site, engaging residues Tyr385, Asn382, and Trp387, which are essential for selective inhibition. The involvement of such catalytically relevant residues reinforces the specificity and potential efficacy of these phytochemicals as COX-2 inhibitors. These findings are in strong agreement with earlier QSAR modeling results, which established statistically robust, mechanistically interpretable models for predicting anti-inflammatory activity. The UniColumn statistical analysis confirmed that the dataset partitioning strategy used in the QSAR modeling was reliable, maintaining internal consistency between training and test sets. The developed QSAR models adhered to OECD validation principles, including a well-defined endpoint, transparent algorithm, defined applicability domain (AD), and appropriate performance metrics. The Williams plot analysis ensured that predictions were made within a valid chemical space, reinforcing the models' reliability in guiding compound prioritization. Collectively, this multi-tiered computational workflow has identified C11 and C31 as promising lead candidates, demonstrating a balance of target affinity, favorable pharmacokinetics, and acceptable toxicity profiles.

The identifier C11 in our study corresponded to Apigenin 7-sulfate, one of the most promising phytochemicals for COX-2 inhibition, supported by strong docking interactions and favorable ADMET properties. These computational insights align well with prior experimental evidence reported by Kumar et al. [32], who isolated and characterized apigenin from *Justicia gendarussa* and demonstrated its potent anti-inflammatory activity both *in vitro* and *in vivo*. In their study, apigenin treatment significantly suppressed COX-2 expression, along with other key inflammatory mediators, including NF-κB, TNF-α, IL-1β, and PGE2, in ox-LDL-induced human peripheral blood mononuclear cells (hPBMCs) [32]. Interestingly, further evidence supporting the biological relevance of apigenin and its analogues comes from the work of Okla et al. [33], who explored the compound's role in adipocyte physiology under inflammatory stress. While primarily focusing on adipocyte browning, their findings revealed a dual modulatory role of apigenin on the inflammatory pathway and the COX-2/PGE2 axis. Apigenin not only suppressed IL-1β-induced NF-κB activation but also induced COX-2 and PGE2 expression, which, in turn, facilitated energy expenditure pathways via EP4 signaling [33]. This dual effect highlights apigenin's

context-dependent regulatory role on COX-2, potentially enhancing adaptive signaling, i.e., browning of adipose tissue, while restraining pathological inflammation. Such complex pharmacological behavior further supports our docking-based observation that Apigenin 7-sulfate binds effectively to the COX-2 active site, potentially modulating its function in different biological contexts. These data collectively point to apigenin-based scaffolds as versatile modulators with therapeutic potential not just in inflammation but also in metabolic-inflammatory disorders such as obesity. Further reinforcing the pharmacological relevance of apigenin-based scaffolds, Adham et al. [34], demonstrated that apigenin exerts potent anti-inflammatory and anticancer effects by modulating the STAT1/COX-2/iNOS signaling axis. Their study showed that apigenin significantly downregulated COX-2 expression in STAT1-overexpressing HEK293 cells, thereby inhibiting inflammatory cascades and enhancing apoptosis in multiple myeloma cell lines. This was accompanied by induction of ferroptosis and cell cycle arrest, indicating multimodal cytotoxicity [34]. Although the compound identified in our study is Apigenin 7-sulfate, a sulphated analogue of apigenin, it shares the core flavone scaffold and key functional groups responsible for COX-2 interaction. These converging lines of evidence from *in silico* predictions and experimental models strongly support continued exploration of this flavonoid derivative as a viable lead for anti-inflammatory drug development.

The second most promising natural compound in our computational work was genistein (C31), with high binding affinity and favorable pharmacokinetic profiles targeting COX-2. This computational evidence is also strongly supported by multiple experimental studies demonstrating genistein's potent anti-inflammatory activity through COX-2 modulation in diverse biological contexts. For instance, Mirahmadi et al. [35] provided compelling *in vivo* data showing that genistein administration effectively attenuated lipopolysaccharide (LPS)-induced neuroinflammation and cognitive dysfunction in a rodent model. Their findings highlighted significant downregulation of COX-2 expression, accompanied by reductions in pro-inflammatory cytokines such as IL-6 and TNF- α , corroborating the role of genistein in suppressing COX-2-mediated inflammatory pathways [35]. Furthermore, Venza et al. [36], demonstrated genistein's ability to inhibit the proliferation of melanoma cells by interfering with PGE2-induced IL-8 production, mediated via downregulation of the EP3 receptor and subsequent reduction in COX-2 signaling [36]. This study emphasizes genistein's broader anti-cancer and anti-inflammatory mechanisms linked to prostaglandin biosynthesis

pathways, supporting the therapeutic relevance of COX-2 inhibition suggested by our computational predictions. Further, Tian et al. [37], explored the combinatorial effects of genistein with celecoxib, revealing synergistic cytotoxicity against prostate cancer cells. Their work underscored that inhibition of COX-2 synthesis is a critical mechanism contributing to the anticancer efficacy of genistein [37]. These findings further substantiate the potential of genistein as an effective COX-2 inhibitor, consistent with its high docking score and ADMET properties observed in our study. Lastly, Du et al. [38], provided mechanistic insights into genistein's neuroprotective and anti-inflammatory effects in a Parkinson's disease model, showing that genistein suppresses microglial activation and pro-inflammatory mediator production, including COX-2, through estrogen receptor and IGF-1R signaling pathways [38,39]. The ability of genistein to modulate COX-2 expression and inflammatory signaling in neurodegeneration aligns well with our molecular docking data, indicating its potential therapeutic benefit in inflammatory disorders. Collectively, these converging lines of evidence, including computational predictions and experimental validations, emphasize the efficacy of apigenin and genistein as natural COX-2 inhibitors. The molecular docking and ADMET profiling presented here not only rationalize the observed biological effects but also provide a strong foundation for further *in vitro* and *in vivo* studies. This integrative approach highlights the promise of these compounds as leads in the development of novel anti-inflammatory agents with potential applications ranging from cancer to neuroinflammatory diseases.

CONCLUSION

This study reports a comprehensive *in silico* strategy for prioritizing phytochemicals as COX-2 inhibitors using multi-model QSAR, supported by computational ADMET and protein-ligand interaction analyses. This systematic multi-parameter screening not only streamlined the hit prioritization process and reduced experimental burden but also highlighted the value of natural product libraries as a resource for developing safer anti-inflammatory therapeutics. However, the observations are based entirely on computational predictions, which, although supported by validated models and well-established methods, may not fully capture the complexity of biological systems. Thus, future research should therefore prioritize experimental validation through enzyme-based inhibition assays to confirm COX-2 selectivity, followed by cell-based anti-inflammatory studies to evaluate functional efficacy. In addition, expanding the phytochemical dataset and incorporating advanced machine learning algorithms may also improve the predictive robustness of QSAR models in

future investigations. Also, an extended molecular dynamics simulation over longer timescales would provide deeper insight into binding stability, conformational adaptability, and interaction persistence under physiologically relevant conditions. Finally, these computational findings, supported by prior experimental evidence, indicate the potential COX-2 inhibitory activity and favorable drug-likeness of the two prioritized compounds, warranting their further evaluation in preclinical and clinical settings for the development of next-generation COX-2 inhibitors.

ACKNOWLEDGEMENT

The authors gratefully acknowledge VLife Technologies, a division of NovaLead Pharma Pvt. Ltd., Maharashtra, India, for providing the licensed access to VLifeMDS software, which was instrumental in the development and validation of our QSAR models. The first author gratefully acknowledges the financial support provided by the Anusandhan National Research Foundation (ANRF), Government of India, New Delhi, under the National Postdoctoral Fellowship (N-PDF) scheme (File No. PDF/2025/000361).

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability Statement

Additional supplementary data supporting the findings of this study may be requested from the corresponding author upon reasonable request.

Author Contributions

YBL: Conceptualization, experimental investigation, software, data analysis and interpretation, manuscript drafting, review, and correspondence. **PBM** and **NSK:** Supervision and critical review of the manuscript. All authors reviewed and approved the final version of the manuscript.

Use of AI tools Statement

No generative AI models were used in the preparation, writing, or content development of this manuscript. All analyses, interpretations, and written materials are solely the result of the authors' original work.

REFERENCES

- Smith WL, Urade Y, Jakobsson PJ. Enzymes of the cyclooxygenase pathways of prostanoid biosynthesis. *Chem Rev.* 2011; 111: 5821-5865.
- Kaur B, Singh P. Inflammation: Biochemistry, cellular targets, anti-inflammatory agents and challenges with special emphasis on cyclooxygenase-2. *Bioorg Chem.* 2022; 121: 105663.
- Hashemi Goradel N, Najafi M, Salehi E, Farhood B, Mortezaee K. Cyclooxygenase-2 in cancer: A review. *J Cell Physiol.* 2019; 234: 5683-5699.
- Jin K, Qian C, Lin J, Liu B. Cyclooxygenase-2-Prostaglandin E2 pathway: A key player in tumor-associated immune cells. *Front Oncol.* 2023; 13: 1099811.
- Guan PP, Wang P. Integrated communications between cyclooxygenase-2 and Alzheimer's disease. *FASEB J.* 2019; 33: 13-33.
- Teismann P. COX-2 in the neurodegenerative process of Parkinson's disease. *BioFactors.* 2012; 38: 395-397.
- Rojas A, Jiang J, Ganesh T, Yang MS, Lelutiu N, Gueorguieva P, et al. Cyclooxygenase-2 in epilepsy. *Epilepsia.* 2014; 55: 17-25.
- Rumzhum NN, Ammit AJ. Cyclooxygenase 2: its regulation, role and impact in airway inflammation. *Clin Exp Allergy.* 2016; 46: 397-410.
- Pu D, Yin L, Huang L, Qin C, Zhou Y, Wu Q, et al. Cyclooxygenase-2 Inhibitor: A Potential Combination Strategy With Immunotherapy in Cancer. *Front Oncol.* 2021; 11: 637504.
- Ferrer MD, Busquets-Cortés C, Capó X, Tejada S, Tur JA, Pons A, et al. Cyclooxygenase-2 Inhibitors as a Therapeutic Target in Inflammatory Diseases. *Curr Med Chem.* 2019; 26: 3225-3241.
- Patrignani P, Patrono C. Cyclooxygenase inhibitors: From pharmacology to clinical read-outs. *Biochim Biophys Acta.* 2015; 1851: 422-432.
- Patrono C. Cardiovascular effects of cyclooxygenase-2 inhibitors: a mechanistic and clinical perspective. *Br J Clin Pharmacol.* 2016; 82: 957-964.
- Rizzo MT. Cyclooxygenase-2 in oncogenesis. *Clinica Chimica Acta.* 2011; 412: 671-687.
- Bernardini S, Tiezzi A, Laghezza Masci V, Ovidi E. Natural products for human health: an historical overview of the drug discovery approaches. *Nat Prod Res.* 2018; 32: 1926-1950.
- Aware CB, Patil DN, Suryawanshi SS, Mali PR, Rane MR, Gurav RG, et al. Natural bioactive products as promising therapeutics: A review of natural product-based drug development. *South African Journal of Botany.* 2022; 151: 512-528.
- Chaachouay N, Zidane L. Plant-Derived Natural Products: A Source for Drug Discovery and Development. *Drugs and Drug Candidates.* 2024; 3: 184-207.
- Liu T, Hwang L, Burley SK, Nitsche CI, Southan C, Walters WP, et al. BindingDB in 2024: a FAIR knowledgebase of protein-small molecule binding data. *Nucleic Acids Res.* 2025; 53: D1633-D1644.
- Laskar YB, Mazumder PB, Talukdar AD. Hibiscus sabdariffa anthocyanins are potential modulators of estrogen receptor alpha activity with favourable toxicology: a computational analysis using molecular docking, ADME/Tox prediction, 2D/3D QSAR and molecular dynamics simulation. *J Biomol Struct Dyn.* 2023; 41: 611-633.
- Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J Cheminform.* 2015; 7: 20.

20. Zambre VP, Hambarde VA, Petkar NN, Patel CN, Sawant SD. Structural investigations by in silico modeling for designing NR2B subunit selective NMDA receptor antagonists. *RSC Advances*. 2015; 5: 23922-23940.
21. Kar S, Roy K, Leszczynski J. Applicability Domain: A Step Toward Confident Predictions and Decidability for QSAR Modeling. *Methods Mol Biol*. 2018; 1800: 141-169.
22. Rakhimbekova A, Madzhidov TI, Nugmanov RI, Gimadiev TR, Baskin II, Varnek A. Comprehensive Analysis of Applicability Domains of QSPR Models for Chemical Reactions. *Int J Mol Sci*. 2020; 21: 5542.
23. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep*. 2017; 7: 42717.
24. Banerjee P, Kemmler E, Dunkel M, Preissner R. ProTox 3.0: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res*. 2024; 52: W513-W520.
25. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000; 28: 235-242.
26. Ayman R, Radwan AM, Elmetwally AM, Ammar YA, Ragab A. Discovery of novel pyrazole and pyrazolo[1,5-a]pyrimidine derivatives as cyclooxygenase inhibitors (COX-1 and COX-2) using molecular modeling simulation. *Arch Pharm (Weinheim)*. 2023; 356: e2200395.
27. Mathada BS, Basha NJ, Javeed M, Karunakar P, Venkatesulu A, Erappa K, et al. Novel pyrimidines as COX-2 selective inhibitors: synthesis, DFT analysis, molecular docking and dynamic simulation studies. *Journal of Biomolecular Structure and Dynamics*. 2024; 42: 1751-1764.
28. Jisha SP, Nagesh GY, Karunakar P, Nidhi G, Sridhar BT, Basavarajaiah SM. Exploration of novel isoniazid embedded 1,3,4-oxadiazole hybrids as anti-TB, antioxidant, and COX inhibitors: synthesis, spectral analysis, and molecular modeling studies. *Journal of the Iranian Chemical Society*. 2025; 22: 717-731.
29. Dassault Systèmes. BIOVIA Discovery Studio Visualizer (25.1.0.24284). 2025.
30. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015; 1-2: 19-25.
31. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*. 2009; 30: 2785-2791.
32. Kumar KS, Sabu V, Sindhu G, Rauf AA, Helen A. Isolation, identification and characterization of apigenin from *Justicia gendarussa* and its anti-inflammatory activity. *Int Immunopharmacol*. 2018; 59: 157-167.
33. Okla M, Al Madani JO, Chung S, Alfayez M. Apigenin Reverses Interleukin-1 β -Induced Suppression of Adipocyte Browning via COX2/PGE2 Signaling Pathway in Human Adipocytes. *Mol Nutr Food Res*. 2020; 64: e1900925.
34. Adham AN, Abdelfatah S, Naqishbandi AM, Mahmoud N, Efferth T. Cytotoxicity of apigenin toward multiple myeloma cell lines and suppression of iNOS and COX-2 expression in STAT1-transfected HEK293 cells. *Phytomedicine*. 2021; 80: 153371.
35. Mirahmadi SM, Shahmohammadi A, Rousta AM, Azadi MR, Fahanik-Babaei J, Baluchnejadmojarad T, et al. Soy isoflavone genistein attenuates lipopolysaccharide-induced cognitive impairments in the rat via exerting anti-oxidative and anti-inflammatory effects. *Cytokine*. 2018; 104: 151-159.
36. Venza I, Visalli M, Oteri R, Beninati C, Teti D, Venza M. Genistein reduces proliferation of EP3-expressing melanoma cells through inhibition of PGE2-induced IL-8 expression. *Int Immunopharmacol*. 2018; 62: 86-95.
37. Tian J, Guo F, Chen Y, Li Y, Yu B, Li Y. Nanoliposomal formulation encapsulating celecoxib and genistein inhibiting COX-2 pathway and Glut-1 receptors to prevent prostate cancer cell proliferation. *Cancer Lett*. 2019; 448: 1-10.
38. Du ZR, Gu Y, Xie XM, Zhang M, Jiang GY, Chen WF. GPER and IGF-1R mediate the anti-inflammatory effect of genistein against lipopolysaccharide (LPS)-induced nigrostriatal injury in rats. *J Steroid Biochem Mol Biol*. 2021; 214: 105989.
39. Knox C, Wilson M, Klinger CM, Franklin M, Oler E, Wilson A, et al. DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Res*. 2024; 52: D1265-D1275.