Chemical Physics Impact 8 (2024) 100484

Contents lists available at ScienceDirect

Chemical Physics Impact

journal homepage: www.sciencedirect.com/journal/chemical-physics-impact

Full Length Article

QSAR and docking based lead optimization of nitrogen heterocycles for enhanced prostaglandin EP2 receptor agonistic potency

Rahul D Jawarkar [a,*], Magdi E.A. Zaki [b,*], Sami A. Al-Hussain [b], Abdul Samad [c], Long Chiau Ming [d,e,f], Summya Rashid [g], Gehan M. Elossaily [h], Susmita Yadav [i], Suraj Mali [i,*]

[a] Department of Medicinal Chemistry, Dr. Rajendra Gode Institute of Pharmacy, University-Mardi Road, Amravati 444901, India
[b] Department of Chemistry, Faculty of Science, Imam Mohammad Ibn Saud Islamic University, Riyadh 13318, Saudi Arabia
[c] Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Tishk International University, Erbil, Kurdistan Region, Iraq
[d] Department of Pharmacy Practice, Faculty of Pharmacy, Universitas Airlangga, Surabaya, Indonesia
[e] School of Medical and Life Sciences, Sunway City 47500, Sunway University, Malaysia
[f] Pengiran Anak Puteri Rashidah Sa'adatul Bolkiah Institute of Health Sciences, Universiti Brunei Darussalam, BE1410 Gadong, Brunei Darussalam
[g] Department of Pharmacology & Toxicology, College of Pharmacy, Prince Sattam Bin Abdulaziz University, P.O. Box 173, Al-Kharj 11942, Saudi Arabia
[h] Department of Basic Medical Sciences, College of Medicine, AlMaarefa University, P.O. Box 71666, Riyadh 11597, Saudi Arabia
[i] Department of Pharmaceutical Sciences and Technology, Birla Institute of Technology, Ranchi 835215, India

ARTICLE INFO

Keywords:
Prostaglandin EP2 Receptor
QSAR
Pharmacophoric features
Crystal structures
Lead compounds

ABSTRACT

In the existing effort, a dataset of 309 experimentally screened molecules for in vitro (Ki) agonist potential for Prostaglandin E2 (PGE2) receptor 2 subtype (EP2), which is a metabolite of arachidonic acid that binds with and regulates cellular responses to PGE2, was investigated in the QSAR (Quantitative structure–activity relationship) study. A six-parameter QSAR model was developed that meets the specified values for internal and external validation as well as random parameters such as R^2_{tr} = 0.808, Q^2_{LMO} = 0.794, R^2_{ex} = 0.781. Insightful and quantitative opinion reveals several underappreciated and distinct structural features that are responsible for the agonist potency of these molecules on Prostaglandin EP2 receptor such as; the hydrogen atom is correct 2 bonds from the donor atom, the sp2 hybridized carbon atom is correct 2 bonds from the cyclic nitrogen atom, and so on. The developed QSAR model captures the narrative as well as the novel pharmacophoric features. The QSAR effect was further demonstrated using the reported crystalline buildings of CP533536 with the Prostaglandin EP2 receptor activity. The evaluation led to the identification of valuable new pharmacophoric properties that will be used to optimize lead compounds in the future.

List of Abbreviations

CADD	Computer Aided Drug Designing
SMILES	Simplified Molecular-Input Line-Entry System
GA	Genetic Algorithm
MLR	Multiple Linear Regression
QSAR	Quantitative Structure-Activity Relationship
QSARINS	QSAR Insubria
OECD	Organization for Economic Co-operation and Development
OFS	Objective Feature Selection
EVA	Evatanepag
PGE2	Prostaglandin EP2 receptor
TAP	Taprenepag

Introduction

The EP2 subtype of the Prostaglandin E2 (PGE2) receptor, a G protein-coupled plasma membrane receptor, plays a crucial role in regulating diverse physiological functions, including tumor-related processes such as occurrence, invasion, metastasis, angiogenesis, chronic inflammation, tumor immunity, and cell apoptosis. Recent research has focused on delineating specific EP2 receptors and their associated signaling pathways within the cyclooxygenase-2 (COX-2)/PGE2/EP2 pathway.

COX-2 and its prostaglandin products have gained attention for their involvement in tumor progression in various organs over the past decade. However, the inhibition of COX-2 using non-steroidal anti-inflammatory drugs (NSAIDs) and specific inhibitors poses side effects,

* Corresponding authors.
E-mail addresses: rahuljawarkar@gmail.com (R.D. Jawarkar), mezaki@imamu.edu.sa (M.E.A. Zaki), mali.suraj1695@gmail.com (S. Mali).

https://doi.org/10.1016/j.chphi.2024.100484
Received 27 December 2023; Received in revised form 17 January 2024; Accepted 17 January 2024
Available online 18 January 2024

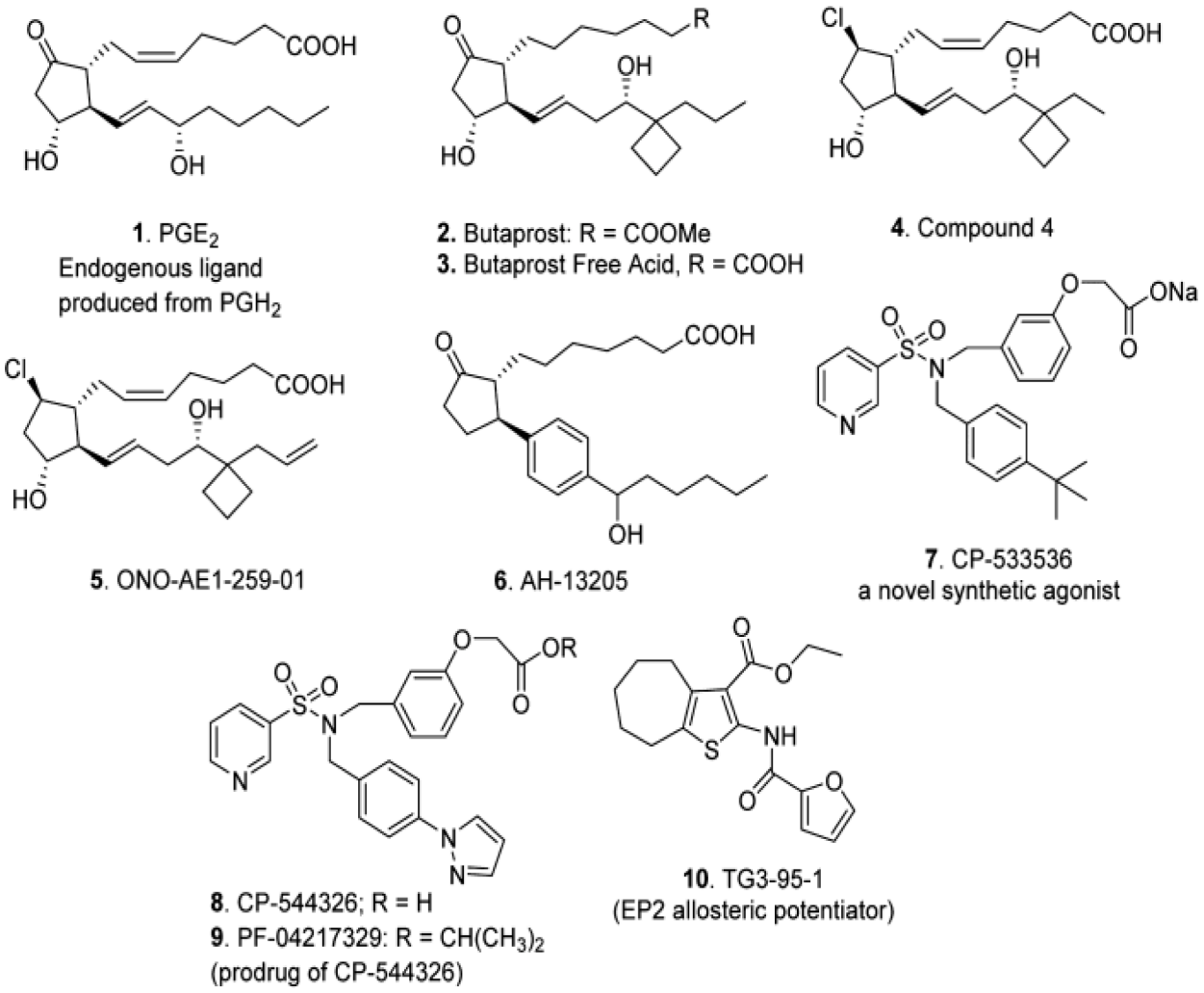


Fig. 1. Depiction of various prostaglandin receptor PGE2 agonists.

limiting their utility. As a primary prostanoid derived from COX-2, PGE2 promotes tumor cell activities. Targeting the EP2 receptor among the four G protein-coupled plasma membrane receptors for PGE2 emerges as a promising strategy in anticancer treatment. EP2 is expressed in multiple human tissues, indicating its significant role in biological development. Studies in animals and human tissues lacking EP2 or treated with EP2 antagonists have shown a downregulation of key signaling molecules, suggesting that inhibiting EP2 may mitigate the proliferation and invasion of cancer cells. Overall, understanding the intricate interactions within the COX-2/PGE2/EP2 pathway offers potential insights for developing targeted therapies with reduced side effects [1–8].

In the case of additional prostanoid receptors, mimetic, negligible effort was followed by researchers about the discovery and improvement of the short-molecule agonist on PGE2 [9–11], is now available for clinical practice. Typical molecules out of which actual were clinically practiced and optimized are illustrated in Fig. 1.

These classes of agonists are structural analogues of endogenous ligand1, as well as nonselective PGE2 agonists including PGE2 Ki = 38 nM. In addition, a synthetic agonist 2 (butaprost) has been described with a 63-fold lower efficacy than 1, including a PGE2 Ki selectivity = 2400 nM [12] −4-fold due to PGE2 on the EP3 or EP4 or selectively closed closer to the IP receptor [13] than reduced acid 3 (Fig. 4) [14]. In addition, compound 457 appearing as is extremely selective because the PGE2 receptor is similar to the prostate-specific receptor [15]. Thus, compound IV is endowed with deficiency stability as unrestricted acid, but its binding is increased significantly upon conversion to the lysine salt [16]. These results suggest the need for a stronger PGE2-induced agonist.

The EP2 receptor is present in various body and brain components, contributing to both essential and undesirable functions. It is crucial to determine if a small molecule adequately describes a natural target, such as a PGE2 receptor, recognizing its affinity akin to a drug. Recent studies using a mouse model for ischemic stroke and inflammatory neurodegenerative diseases align with the proposed role of EP2 in epilepsy, supporting potential developments in treatment.

While accessible EP2 agonists and antagonists precisely bind to PGE2 in vitro assays, in vivo evaluation is necessary to confirm their effects on the receptor. Quantitative Structure-Activity Relationship Analysis (QSAR) methods, relying on mathematical models, have garnered interest for predicting the biological activity of molecules. QSAR studies, promoted for their time and resource efficiency, involve the combination of molecular descriptors for accurate predictions [16–20].

The FDA's efforts in creating a diverse chemistry repository/database, incorporating trial, proficiency, and safety data, further support the development of computational algorithms and predictable QSAR models. Despite numerous studies on the quantitative structure-activity relationship of EP2 agonists in various conditions, such as ocular hypertension and inflammatory diseases, no QSAR assessment has been implemented to date. In this context, our work involves QSAR analysis on a modestly sized data set of polyphasic EP2 agonists, aiming to provide valuable insights for lead optimization. The results from this analysis can significantly contribute to the development of new

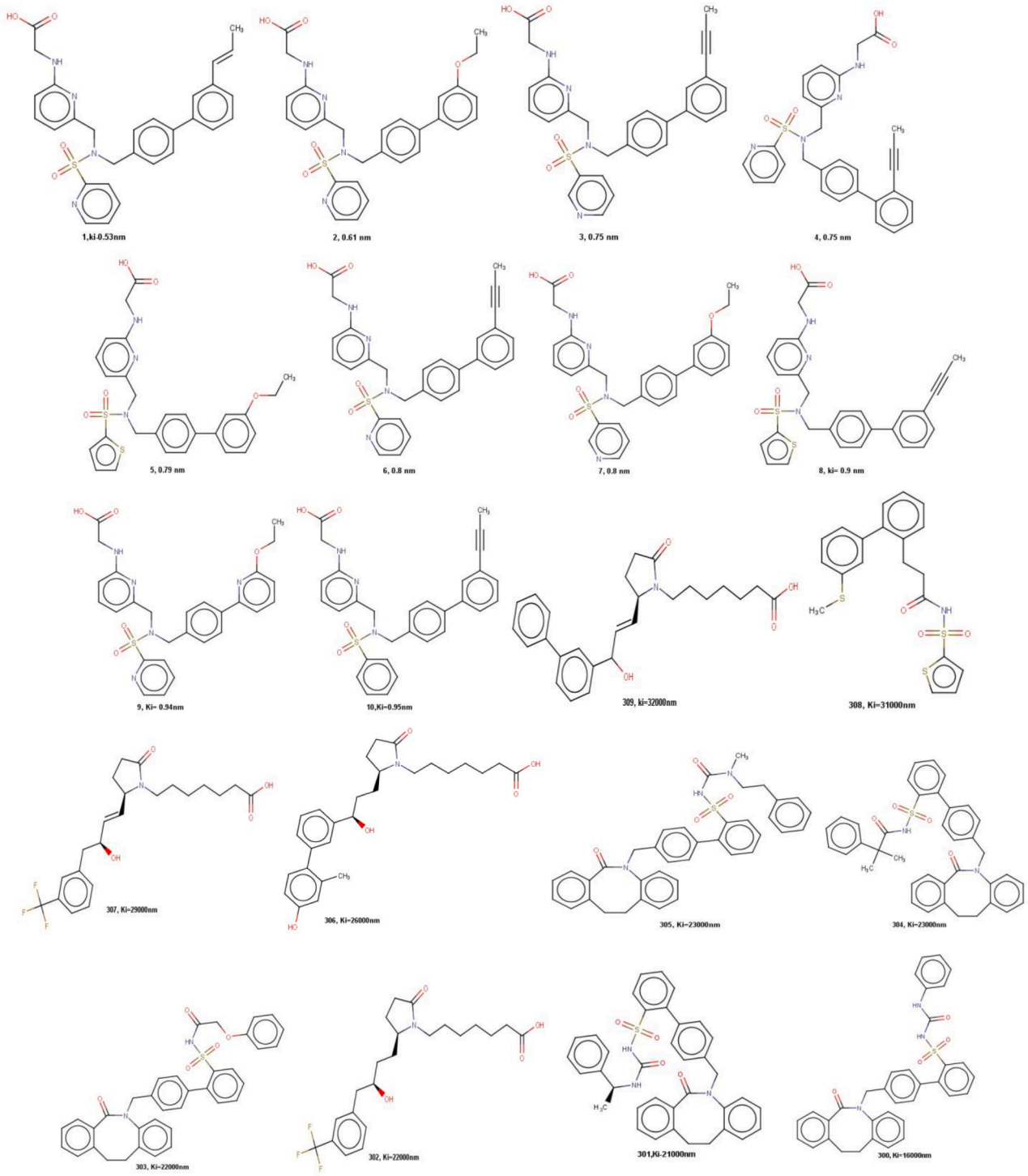


Fig. 2. Representative examples from selected datasets with ki values (10 most active 1–10 molecules and 10 least active 300–309 molecules).

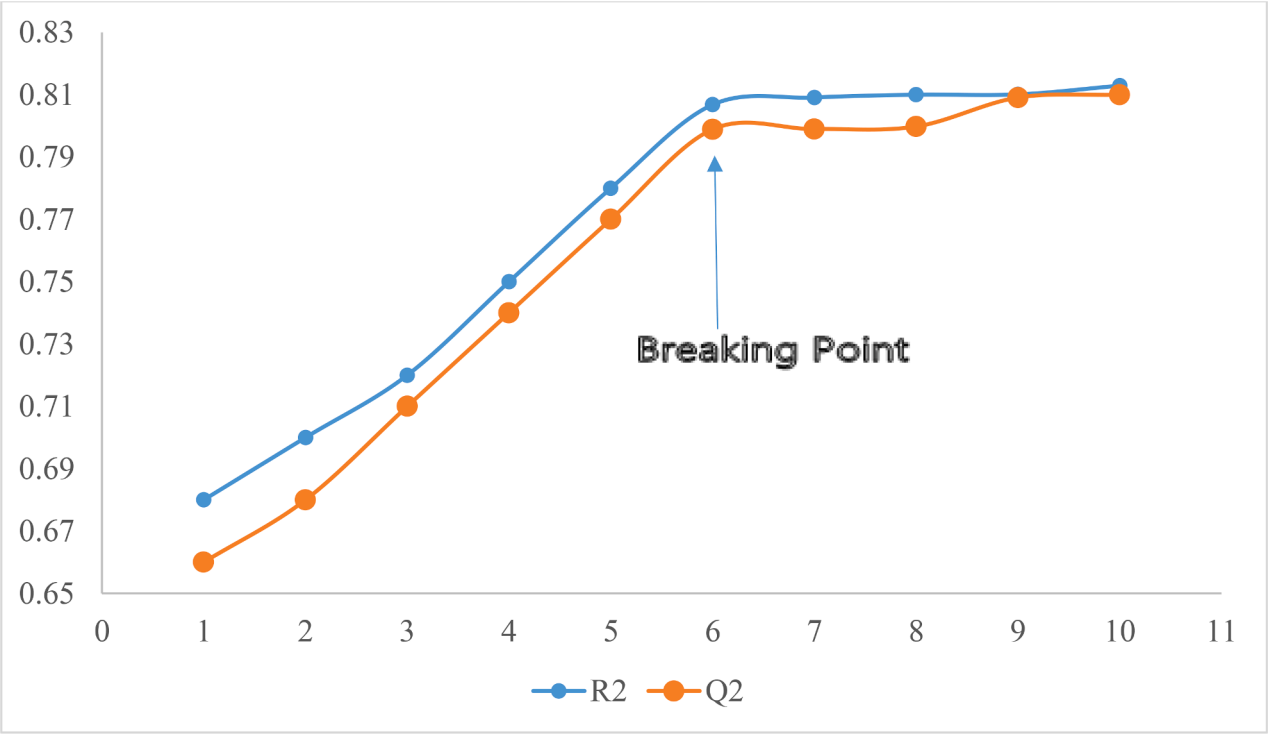


Fig. 3. Depiction of Plot for the number of descriptors against the Coefficient of Determination R^2 and Leave-One out Coefficient of Determination Q^2 to identify the optimum number of descriptors.

compounds, particularly prostaglandin EP2 receptor agonists.

Methods

To construct a robust QSAR model with substantial applicability in predicting agonist activity of EP2 receptors, several sequential steps are undertaken. These include the collection and curation of series information, application of structure technology, and sequential acquisition of estimated molecular descriptors. The subsequent stages involve Objective Feature Selection (OFS), division of the dataset into training and external validation sets, Subjective Feature Selection (SFS), and the development and validation of the QSAR model. With the QSAR model for EP2 receptor agonist activity now established, today's allocation based on agonist activity supersedes conventional OECD guidelines.

Data collection & curation

The data utilized to construct, train, and validate the QSAR model for EP2 receptor agonists is sourced from the binding database (https://www.bindingdb.org/bind/chemsearch), a publicly accessible database (Date of accession: 12–01–2023). This dataset comprises structurally distinct molecules that have undergone empirical studies for agonist activity (Ki) on the EP2 receptor. Subsequently, molecules featuring questionable Ki values (enzyme inhibition constants), duplications, salts, metal-based inhibitors, etc., were meticulously excluded, as detailed in the statistical maintenance section [21–26,17,27,28]. As a concluding note, the dataset comprises 309 distinct molecules with various structural configurations, each experimentally assessed for efficacy in terms of Ki (nM) (refer to the supplementary material, available in the Excel file for the "final supplementary material"). The empirical Ki values range consistently up to 85,000 nM [29], demonstrating a diversity of values beyond 0.3. Subsequently, the Ki values undergo conversion based on the negative logarithmic scale ($pKi = -\log_{10}Ki$), simplifying value assignments. Fig. 2 illustrates examples of the 10 most active molecules (1–10) and the 10 least active molecules (300–309) for reference purposes only.

Calculation of molecular descriptors and objective feature selection (OFS)

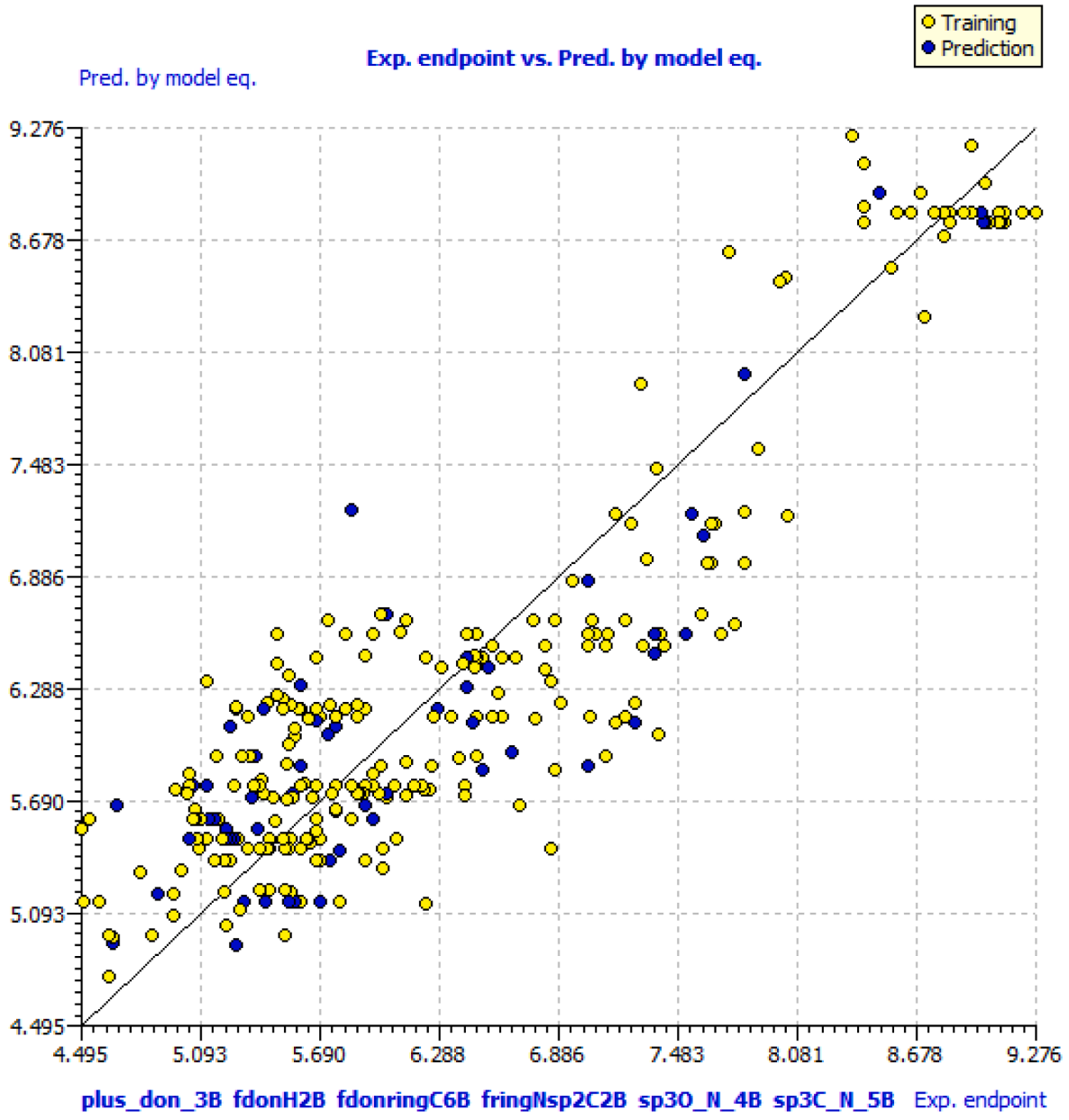
The SMILES notation was initially converted into a 3D optimized structure using Open Babel 3.1 to compute molecular descriptors [29]. To enhance the interpretability of the QSAR analysis mechanism, it is crucial to accurately compute various molecular descriptors and prune them to prevent the risk of overfitting with noisy redundant descriptors. For this purpose, PyDescriptor, capable of calculating over 30,000 molecular descriptors, was employed. This extensive array encompasses 1D to 3D molecular descriptors. Subsequently, Objective Feature Selection (OFS) was implemented using QSARINS 2.2.4 [17], excluding molecular descriptors that were near-constant, constant, or strongly cross-correlated ($|R| > 0.90$). The final set comprises 1260 molecular descriptors, offering a diverse range that spans various chemical spaces (Refer to the supplemental material for formulas).

Splitting of the data set molecules into training and external sets and subjective feature selection

Before engaging in subjective feature selection, it is prudent to partition the entire dataset into training and prediction sets (also referred to as external or test sets) with suitable configurations and sizes to prevent information leaks. To ensure impartiality, the entire dataset was randomly divided into a training set (80 % = 247 molecules) and a prediction or external set (20 % = 62 molecules). The primary objective of the training set was to determine the optimal number of molecular descriptors, while the predictive/external set was exclusively utilized for external validation of the model (predictive QSAR).

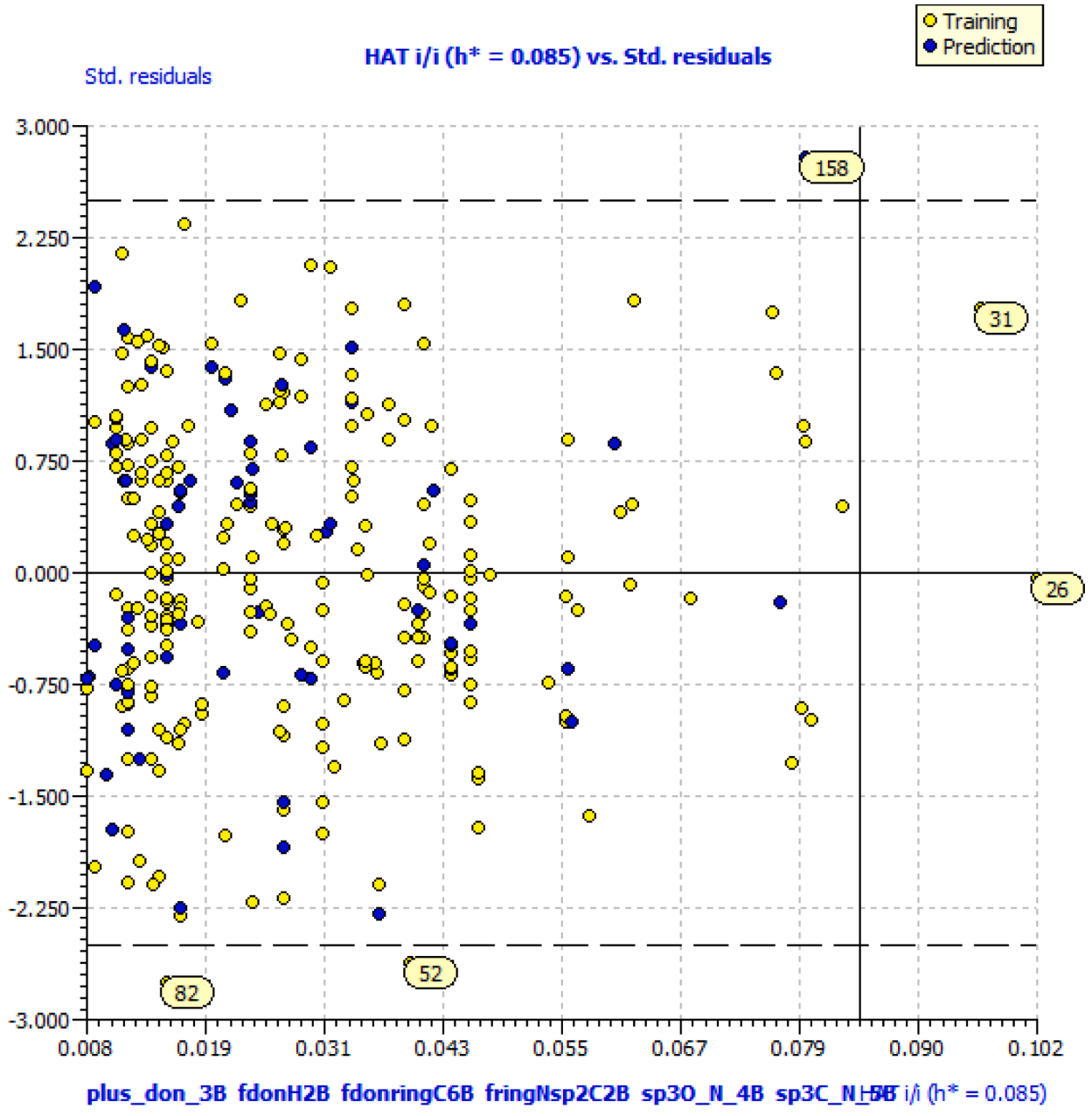
For subjective feature selection, the Genetic Algorithm Fusion Multiple Linear Regression (GA-MLR) method, as introduced in QSARINS 2.2.4, was employed to select the appropriate descriptors, utilizing Q^2_{LOO} as fitness parameters.

To promote fast-growing QSAR model, it is much more important to



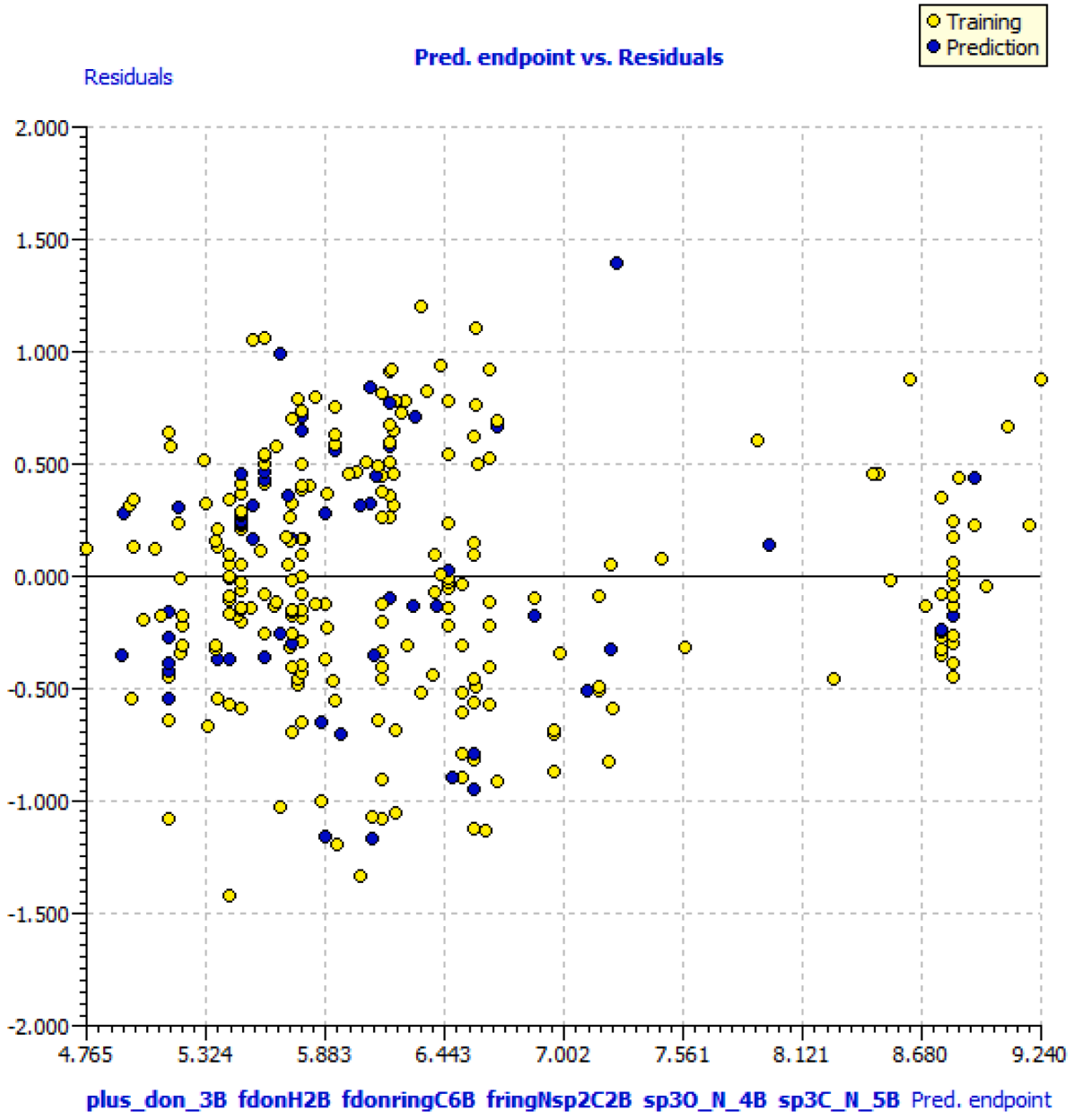
(A)

Fig. 4. (A) Graph of experimental vs. Predicted pIC_{50} values for a model (B) Williams plot for a model (C) Graph of Residual vs. Predicted pIC_{50} values for a model (D) Insubria Plot for a model .



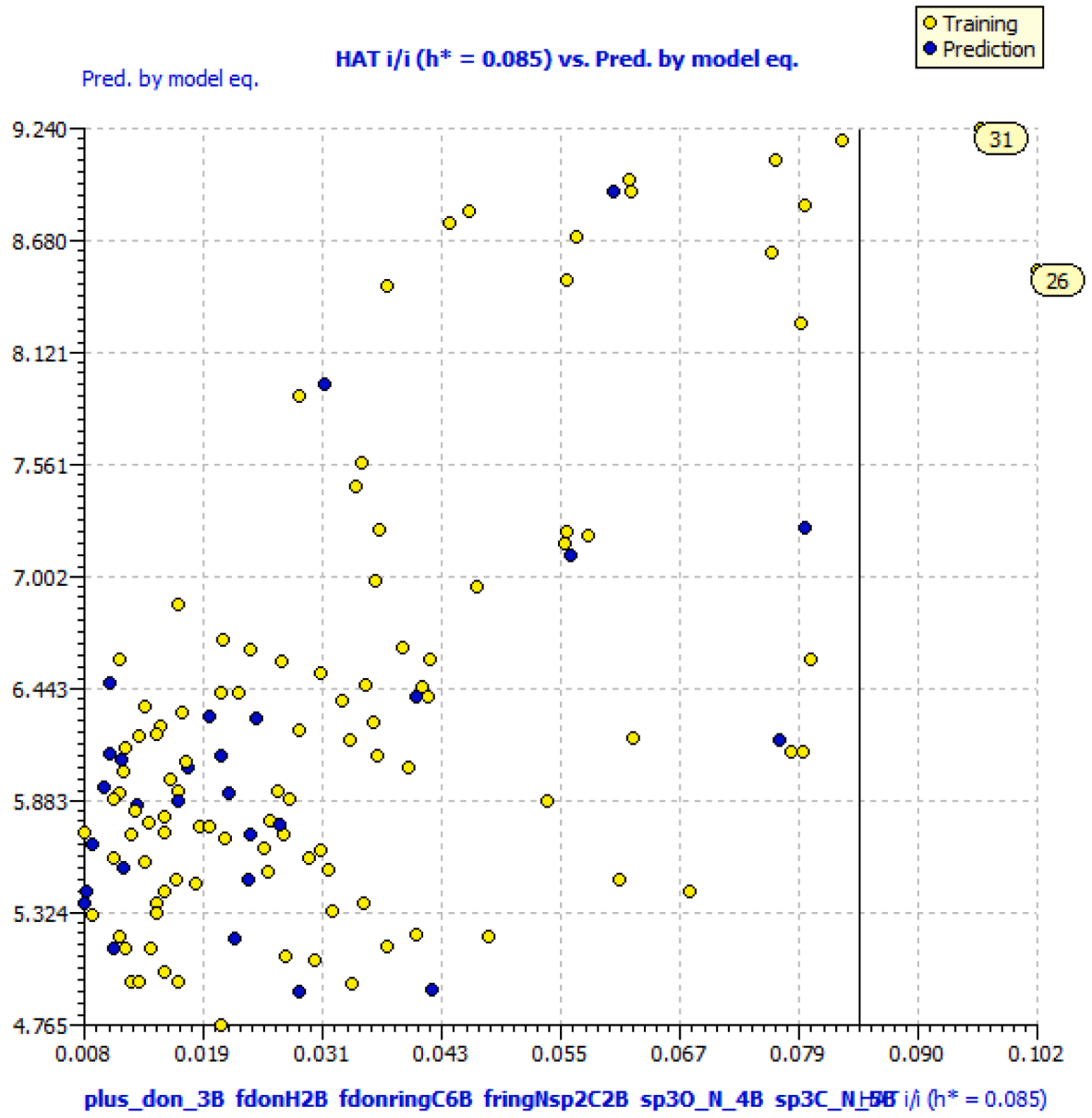
(B)

Fig. 4. (*continued*).



(C)

Fig. 4. (*continued*).



(D)

Fig. 4. (*continued*).

avoid overfitting and the need for a sufficient range of molecular descriptors to achieve acceptable interpretability. Therefore, in this study, we performed a break factor by drawing a diagram in a series of molecular descriptors with complex versions and R^2tr and Q^2_{LOO} values [30]. Therefore, the graph provides a concept that roughly represents the range of molecular descriptors similar to the decay factor as the gold standard range of QSAR versioning descriptors. From the comment plot, we obtained the fractional factors of the six variables. Therefore, more than 6 descriptors were truncated throughout the QSAR model building (Figs. 8-12).

Model development and validation-

We employed various validation criteria, as outlined in the literature, to assess the robustness of the model construction. To accomplish this, we utilized the coefficient of determination (r^2), Leave-one-out cross-validation (Q^2_{LOO}), and Leave-many-out mutual validations (Q^2_{LMO}). Additionally, an estimated standard error was defined for each model developed, incorporating RMSE (Root Mean Square Error) for both the Training (RMSETR) and External Prediction Set ($RMSE_{ext}$). These metrics collectively represent the overall error in the model, serving as an integral aspect of the accuracy of the QSAR model described for a specific dataset [31]. Cross-correlation between descriptors underwent testing through the QUIK rule (Q under the influence of K). The QUICK rule was set at 0.05 to mitigate cross-correlation between descriptors. The reliability of the developed QSAR model was established by conducting Y randomization, involving 2000 iterations to assess the appropriateness of randomly arranged Y data. Randomization of the built QSAR model entailed shuffling the dependent variables (PIC_{50} values) in the training set and recalculating a new coefficient of determination. A significantly lower coefficient of determination in the new model implies that the reported model was not derived through random correlation in the current QSAR analysis [32,33]. In essence, the predictive prowess of the developed QSAR model hinges on the proximity of

Table 1
The statistical parameters associated with fitting, double validation and Y-scrambling for QSAR model.

Statistical Parameters	Model
Fitting	
R^2	0.8086
R^2_{adj}	0.8038
R^2-R^2_{adj}	0.0048
LOF	0.2903
Kxx	0.3451
Delta K	0.0527
RMSE tr	0.5128
MAE tr	0.4163
RSS tr	65.2034
CCC tr	0.8942
S	0.5201
F	169.6780
Internal Validation	
Q^2_{LOO}	0.7975
R^2-Q^2_{LOO}	0.0111
$RMSE_{cv}$	0.5274
MAE_{cv}	0.4283
$PRESS_{cv}$	68.9823
CCC_{cv}	0.8882
Q^2_{LMO}	0.7943
R^2_{Yscr}	0.0244
RMSE AV $_{Yscr}$	1.1576
Q^2_{Yscr}	−0.0334
External Validation	
$RMSE_{ext}$	0.5294
MAE_{ext}	0.4380
$PRESS_{ext}$	17.0981
R^2_{ext}	0.7817
Q^2-F^1	0.7855
Q^2-F^2	0.7810
Q^2-F^3	0.7959
CCC_{ext}	0.8761
r^2m aver.	0.6758
r^2m delta	0.1926
k'	0.9970
K	0.9958
Clos'	0.0856
Clos	0.0002

predicted values to observed values (experimental bioactivity). The presence of even a single outlier diminishes the predictive power of the QSAR model. Consequently, we sought to highlight outliers based on compounds exhibiting markedly higher residues in the GA-MLR QSAR model. Moreover, we identified outlier connections by comparing predicted values with standardized residual values. Likewise, structural changes in database connections were discerned through the leverage effect on the Williams diagram. The extent of the developed QSAR model is determined by amalgamating leverage and standard residuals.

Results

Despite the fact that the studies considered are based on reasonable data sets, the occupancy of diverse molecular scaffolds, functional groups, substituents, branched rings, i.e. Non-aromatics, homoaromatic, heteroaromatic, fused rings; unusually, spirocompounds and the like covered the vast chemical space. Therefore, the developed QSAR model is based on a shared dataset. Fig. 3

R^2 (the coefficient of determination), R^2_{adj}- adjusted coefficient of determination, CCC_{tr}, Leave-One-Out (LOO), Leave-Many-Out (LMO) or bootstrap and calculating the corresponding cross-validated correlation coefficients (Q^2_{LOO}, Q^2_{LMO}) etc. have values well above the approved conformance parameter thresholds, confirming that the QSAR model with the required number of molecular descriptors is statistically acceptable. Internal validation parameters such as Q^2_{LOO}, Q^2_{LMO}. The values indicate the statistical robustness of the QSAR model in descending order. External predictability of both models is reflected in high values such as external validation aspects R^2_{ex}, Q^2_{Fn}, etc. The Williams diagram of the model (each Fig. 4) underpins the model applicability domain (AD). Compliance with the approved range of many parameters and low correlation between molecular descriptors hinders the feasibility of random development of QSAR models (supplementary information). These reasons support the statistical robustness and excellent external predictability of these models [34,35]. (See Table 1)

GA-MLR QSAR model

Model (divided set: training set-80 % and prediction set-20 %)

$_{p}$KI=6.565 (± 0.222) + -0.057 (± 0.018) * plus_don_3B + 0.58 (± 0.062) * fdonH2B + -0.28 (± 0.061) * fdonringC6B + 0.544 (± 0.08) * fringNsp2C2B + 0.268 (± 0.121) * sp3O_N_4B + -0.056 (± 0.02) * sp3C_N_5B

The aforementioned statistical validation parameters are recommended and carry standard implications for evaluating both internal and external robustness (refer to the supplementary material for clear explanations and formulas). Elevated values for exclusive parameters such as R^2_{tr} (scale of determination) and R2adj, along with low LOF values (indicating better performance) such as $R^2_{adj.}$ Low LOF values (bad) such as (adjusted R-squared) and R^2_{cv} (Q^2_{loo}) (leave one out cross validated R-squared), R^2_{ex} (external R-squared), Q^2-Fn and CCC_{ex} (match correlation coefficient), $RMSE_{tr}$ (Mean Squared Error), MAE_{tr} (Mean Absolute Error), R^2_{Yscr} (R^2 of Y-scrambling), etc., suggest that the model possesses the capability to provide accurate external predictions without random correlation. Furthermore, the Williams plot illustrates the statistical applicability of the model (refer to Fig. 4). Consequently, the model adheres to all the guidelines for constructing a reliable QSAR model as recommended by the Organization for Economic Co-operation and Development (OECD).

Discussion

Mechanistic interpretation

fdonH2B, fringNsp2C2B, sp3O_N_4B: All these molecular descriptors have positive values for the coefficients of the developed QSAR model, and increasing the values of these molecular descriptors increases the agonist activity of the Prostaglandin EP2 receptor. fdonH2B (frequency of occurrence of hydrogen atom exactly at 2 bonds from donar atom) have positive correlation with the agonist activity of Prostaglandin EP2 receptor. If the same hydrogen atom is present together in one or three bonds of another donor atom, it will be hidden during the calculation of the molecular descriptor fdonH2B. This finding is evident when comparing molecule 1 (pKi = 9.27, fdonH2B = 6) with molecule 170 (pKi = 5.74, fdonH2B = 0) (see Fig. 5). Shifting the descriptor value from 0 to 6 on the molecule 170 increases the pKi value of the agonist activity of Prostaglandin EP2 receptor on the molecule 170 by 3.53 units (about a 30-fold increase in Prostaglandin EP2 receptor antagonistic action). The significance of this molecular descriptor is that all 10 active molecules in this descriptor (Molecules 1–10, pKi range 9.276–9.022, fdonH2B = 6) take a value of 6 and all 10 are the least active. It can be explained by the facts. For numerator (molecule 294–300, 303–305, pKi range 5–4.638, fdonH2B = 0), the value of this descriptor is 0 (see Fig. 5).

The positive correlation of current descriptors is noteworthy, as an increase in the value of a particular descriptor corresponds to a further enhancement in the agonist activity of the Prostaglandin EP2 receptor dataset molecule. Specifically, the presence of exactly two hydrogen atoms in the second bond from the donor atom seems to play a crucial role in the agonist activity of the Prostaglandin EP2 receptor, as indicated by the descriptor fdonH2B. However, it is noteworthy that replacing fdonH2B with the descriptor fdonH3B, where hydrogen atoms

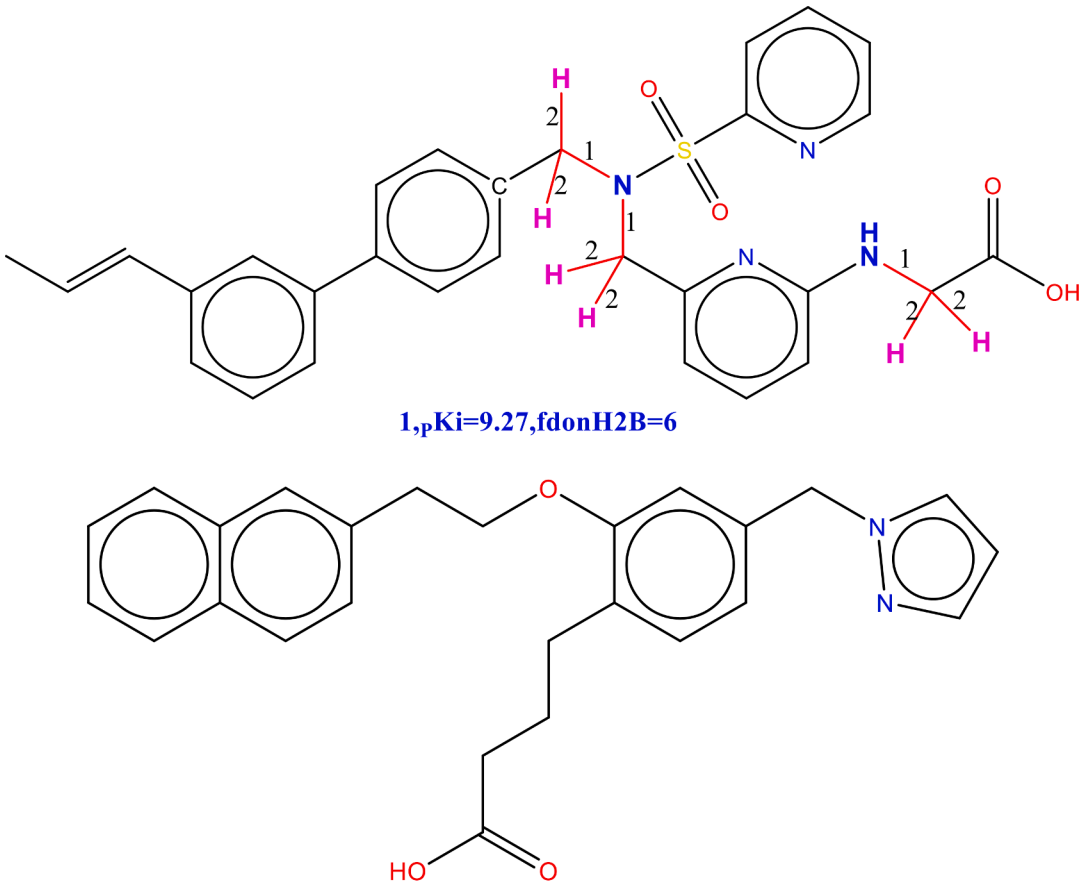


Fig. 5. Pictorial illustration of the molecular descriptor fdonH2B for the molecule 1 and 170 only.

are generated from the donor atom in exactly three bonds, leads to a reduction in the statistical representation of the model ($R^2 = 0.52$, $Q^2 = 50$). This reduction suggests that, given hydrogen's minimal mass, the smallest mass must be located in close proximity to the donor atom for optimal agonist activity of the Prostaglandin EP2 receptor. Similar trends are observed in the ligand molecule EVA, exemplified by (CP533536), reinforcing the importance of having minimal mass near the donor atom for improved agonist activity (refer to Fig. 12). fringNsp2C2B (frequency of occurrence of sp2 hybridized carbon atom exactly at 2 bonds from ring nitrogen atom). This molecular descriptor has a positive number in the developed QSAR model, so increasing its value may further increase the agonist activity of Prostaglandin EP2 receptor. The most active series of molecules are 14 (pKi = 8.95, fringNsp2C2B = 4), 22 (pKi = 8.72, fringNsp2C2B = 4), 26 (pKi = 8.55, fringNsp2C2B = 4), 28 (pKi = 8.42, fringNsp2C2B =). Four)).), 29 (pKi = 8.42, fringNsp2C2B = 4) and 31 (pKi = 8.35, fringNsp2C2B = 4), sp2 hybridized carbon atoms are precisely located in the two bonds from the ring nitrogen atom, but are the most active. The lower molecules do not have 280 (pKi = 5.097, fringNsp2C2B = 0) 287 (pKi = 5.06, fringNsp2C2B = 0), 289 (pKi = 5.036, fringNsp2C2B = 0), and 290 (pKi = 5.036, fringNsp2C2B = 0). (See Fig. 6)

Increasing the descriptor value from 0 to 4 in molecule 280 results in a substantial rise of approximately 3.86 units in the pki value for the agonist activity of the Prostaglandin EP2 receptor compared to molecule 14 (an approximately 30-fold increase). This indicates an elevated agonist efficacy for the Prostaglandin EP2 receptor. Consequently, replacing the descriptor fringNsp2C2B with fringNsp2C3B (frequency of sp2 hybrid carbon atoms in exactly three bonds from the ring nitrogen atom) diminishes the statistical power of the model ($R^2 = 0.67$, $Q^2 = 65$). This underscores the significance of the fringNsp2C2B descriptor.

Upon closer observation, the presence of a simple sp2-hybridized carbon atom adjacent to the ring nitrogen atom appears crucial for enhancing the agonist efficacy of the Prostaglandin EP2 receptor. Further insights reveal that the optimal distance between the ring nitrogen and the sp2 hybrid carbon atom is two bonds. This finding gains support from the presence of fringNsp2C3B in the clinical trial molecule CP544326 (refer to Fig. 6).

sp3O_N_4B (occurrence of nitrogen atoms within 4 bonds from the sp3 oxygen atom) This descriptor has a positive number in the developed QSAR model, so increasing its value can further increase the agonist activity of Prostaglandin EP2 receptor. This observation Supported when have compared molecules 14, 26, 31, 19, 42, 44, 46, 48, 61, and 66 (pKi = 8.95–7.17, sp3O_N_4B = 2) with molecules 221, 222 and 225, 227, 228., 229, 230, 231, 233, 234, and 235 (pKi = 5.50–5.40, sp3O_N_4B = 0).(See Fig. 7)

This observation suggests that the molecular descriptor sp3O_N_4B plays an important role in determining the agonistic activity of Prostaglandin EP2 receptor. This may be the reason that may be due to fluctuations in the biological activity of these molecules. Subsequent use of the molecular descriptor sp3O_N_7B (the number of sp3 hybrid oxygen atoms in the nitrogen atom or in 7 bonds from the nitrogen atom) instead of sp3O_N_4B slightly reduces the statistical detectability of the model ($R^2 = 0.795$)., $Q^2 = 0.785$). Similar opinions are highlighted and reinforced by the presence of the sp3O_N_4Bin molecule TG3951 in clinical trials. (See Fig. 7). Therefore, the optimal distance value between sp3 hybrid oxygen and nitrogen atoms is 4 bonds. plus_don_3B, fdonringC6B, sp3C_N_5B: These three molecular descriptors have negative coefficients, so reducing their values increases the agonist

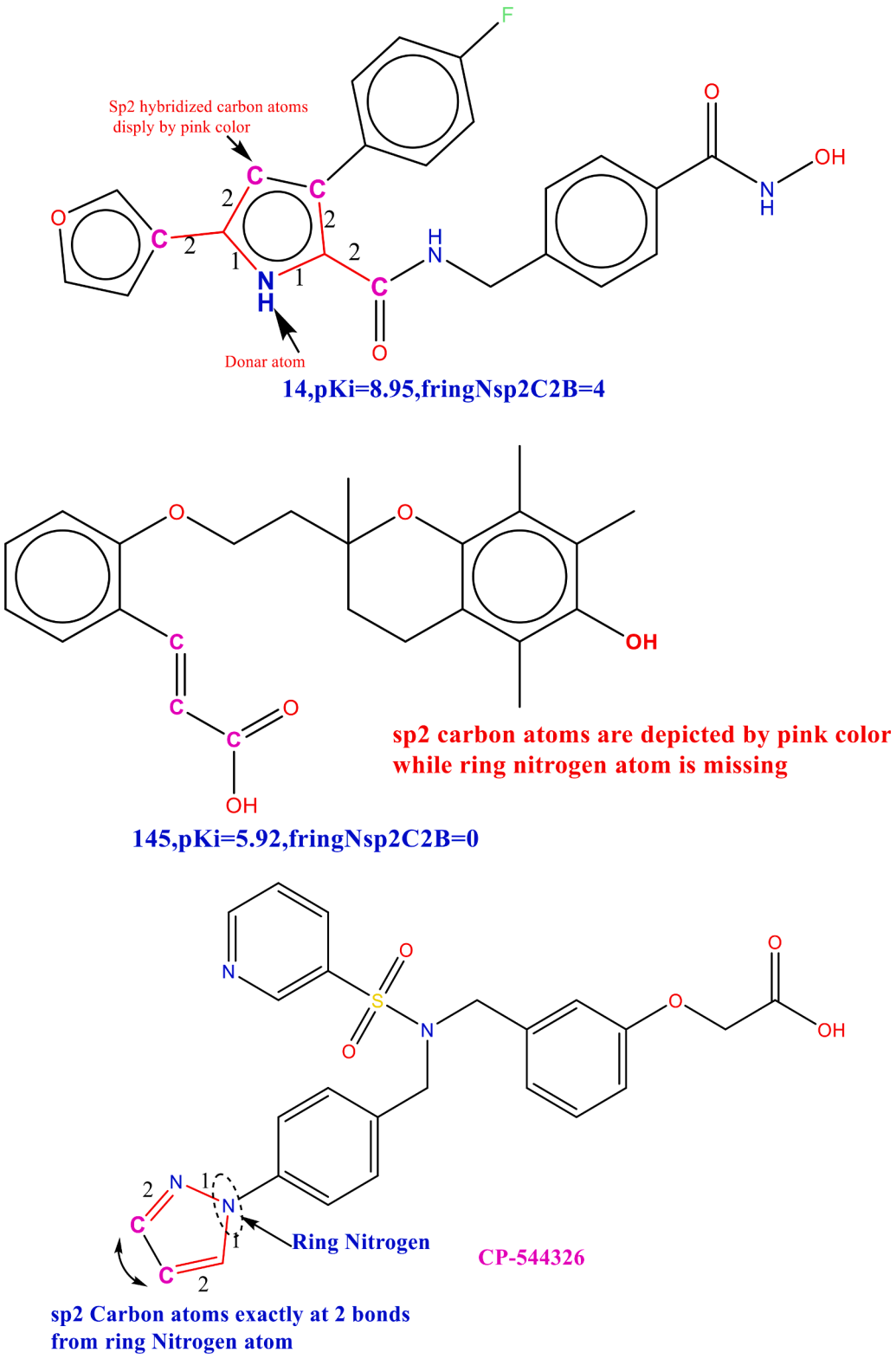


Fig. 6. Pictorial presentation of the molecular descriptor fringNsp2C2B for the molecule 14,145and CP-544326.

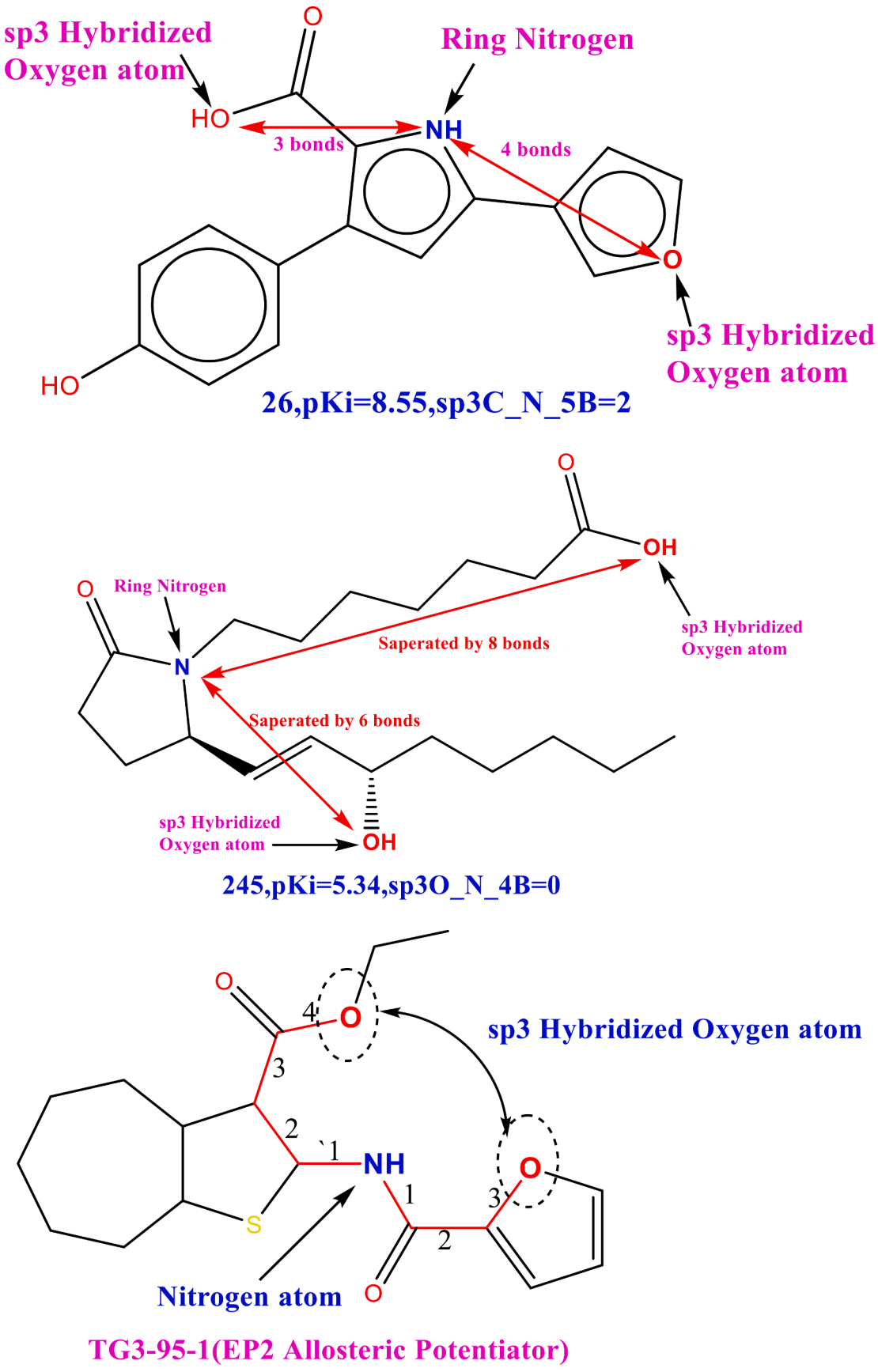


Fig. 7. Presentation of the molecular descriptor sp3O_N_4B for the molecules 26, 245 and TG3-95-1(PGE2 Allosteric Potentiator) only.

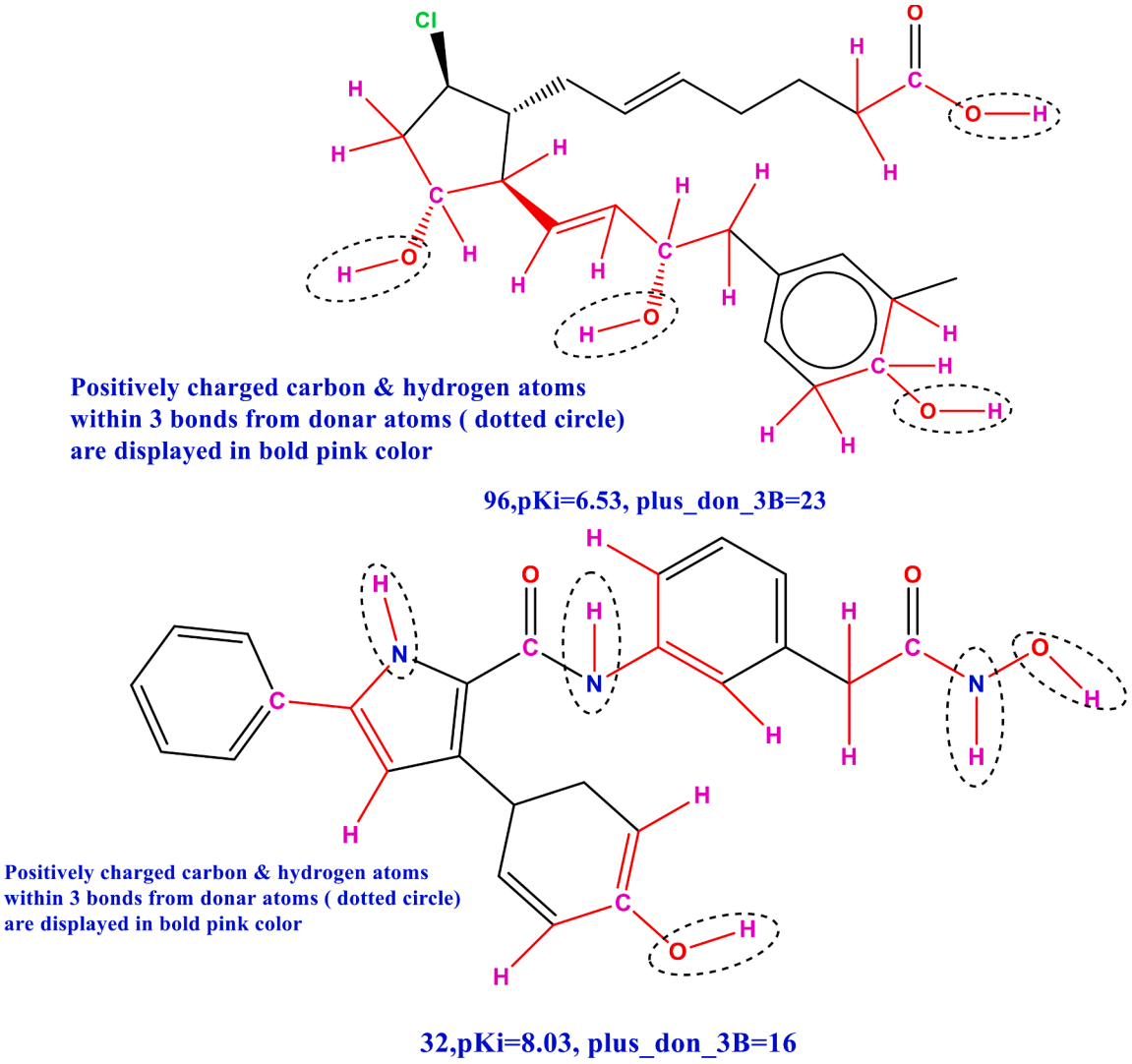
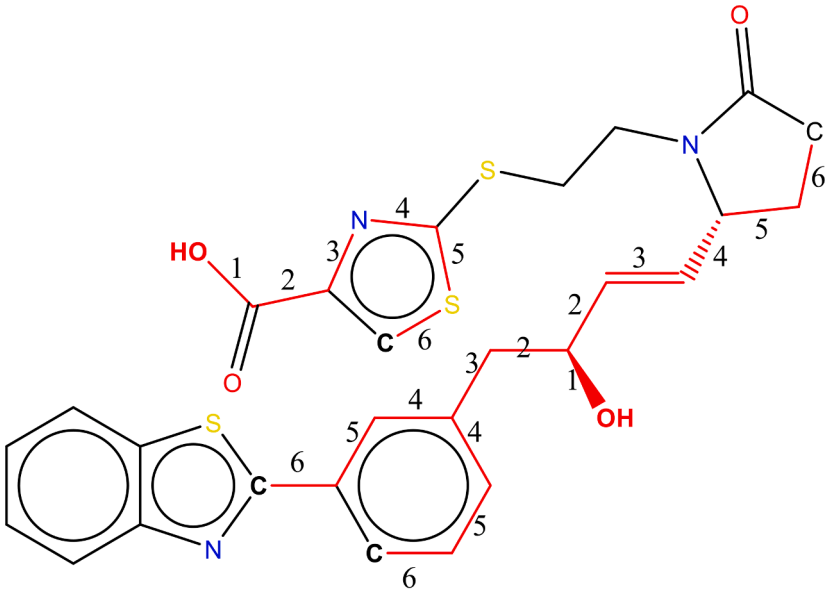


Fig. 8. Pictorial explanation of the molecular descriptor plus_don_3B for the molecules 96 and 32 only.



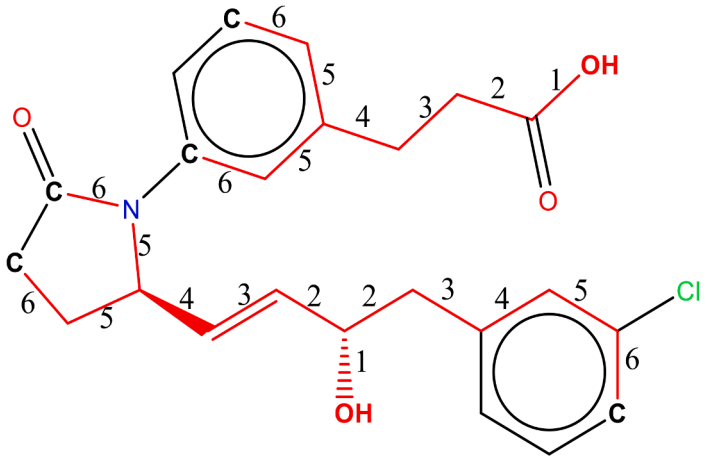


Fig. 9. Pictorial explanation of the molecular descriptor fdonringC6B for the molecules 36 and 267 only.

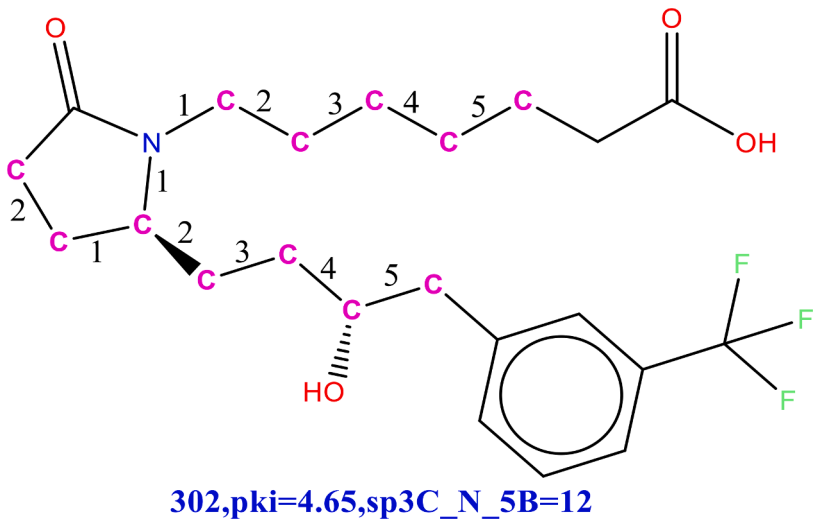


Fig. 10. Illustration of molecular descriptor sp3C_N_5B for the molecule 302.

activity of Prostaglandin EP2 receptor. plus_don_3B (number of positively charge atom from donor atom within 3 bonds) Negative numbers in the molecular descriptor plus_don_3B of the QSAR model indicate that increasing the value of a particular descriptor may reduce the agonist activity of Prostaglandin EP2 receptor. (See Fig. 8)

If the same positively charged atom is present in two or less bonds from the donor atom at the same time, the calculation of this molecular descriptor is ignored. Furthermore, reducing the value of the descriptor from 23 to 16 in molecule 96 increases activity by about 1.5 units (the agonist efficacy of Prostaglandin EP2 receptor increases about 15-fold). In addition, this observation is supported by a comparison of the following molecular pairs: 106, 147, 148, 160, 182, and 193 (pki = 6.46 to 5.27, plus_don_3B = 22), molecules 18, 20, 21, 23, 24, and 25 (pki = 8.5 to 8.8, plus_don_3B = 22). These observations show the importance of this molecular descriptor in the QSAR.

fdonringC6B (frequency of occurrence of ring carbon atom exactly at 6 bonds from donor atom) If an equivalent donor atom also occurs in a 4 or 7 bond from a ring carbon atom, it is omitted from the calculation of this descriptor. Replacing the fdonringC6B molecular descriptor with ringC_don_6B (frequency of ring carbon atoms in 6 bonds from the donor atom) significantly reduces the performance of the model (R^2 = 0.742, Q^2 = 0.735). Therefore, the fdonring_C6B molecular descriptor is a better choice for predicting agonists. Anterior prostaglandin-E2 receptor PGE2 activity. (See Fig 9)

Therefore, all observations in the QSAR model and their negative numbers suggest reducing the value of this descriptor in order to expect better agonistic action on Prostaglandin EP2 receptor. This observation applies to the next pair of molecules in the dataset: molecule 27 (pki = 8.49, fdonringC6B = 1) and molecule 267 (pki = 5.21, fdonringC6B = 5). Furthermore, reducing the value of the descriptor fdonringC6B from 5 to 1 in molecule 267 increases the pki value by about 3.28 units (the agonist activity of Prostaglandin EP2 receptor increases about 30-fold for Prostaglandin EP2 receptor).

sp3C_N_5B (number of sp3 hybridized carbon atoms within 5 bonds from nitrogen atom) If a similar nitrogen atom is present in the 3 or 6 bond of another sp3 hybrid carbon atom, it will be removed during the descriptor calculation. Therefore, this molecular descriptor is a negative number in the developed QSAR model, so its value should be kept as low as possible in order to expect better agonist activity of Prostaglandin EP2 receptor. Then reducing the value of the molecular descriptor sp3C_N_5B from 12 (pKi = 4.65, sp3C_N_5B = 12) in the molecule 302 to 3 increases the pki value by about 4.39 units (about 40 times more than the prostaglandin E2 (Receptor PGE2 agonist efficacy of molecule 302 compared to molecule 8) Therefore, this analysis confirmed that sp3 hybridized carbon atoms play an important role in determining agonist activity.

Comparison of QSAR results with the reported crystal structures

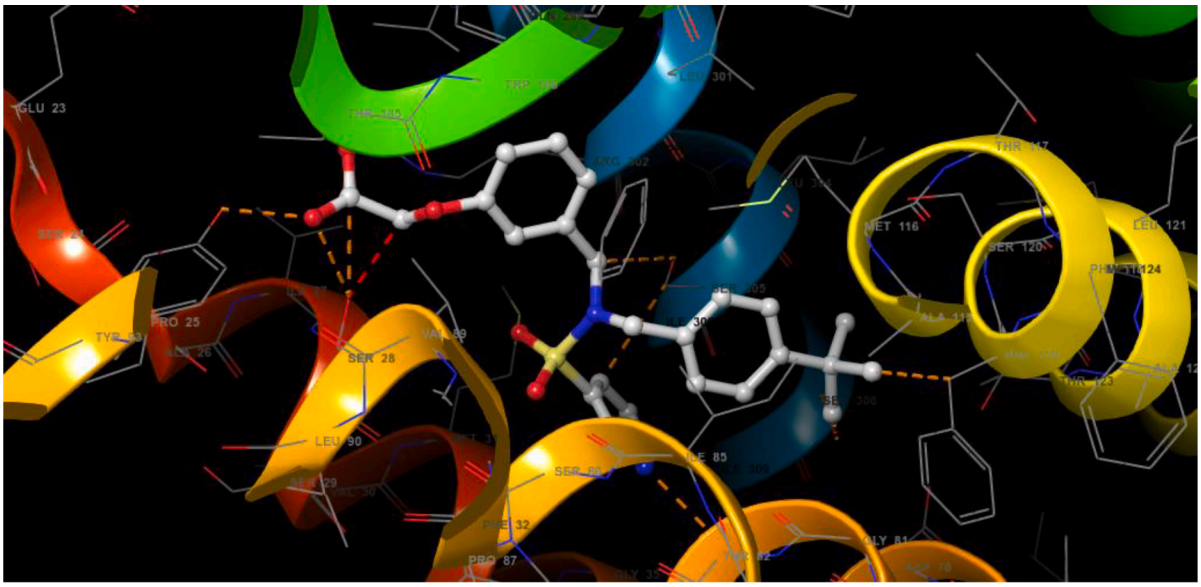
The cryo-electron microscope (cryoEM) structure of the PGE2G (PDB7CX4) complex, in conjunction with the highly selective agonist EVA (CP533536), was resolved at global resolutions of 2.8 Å, 2.8 Å, and 2.9 Å, respectively, for the purpose of comparison. A cursory analysis reveals that EVA binds to PGE2 through hydrophobic interactions across three sub-pockets. Within sub-pocket A, the carboxyl terminus of EVA establishes an H-binding interaction with Arg302 and Ser28 residues. Simultaneously, the phenyl linker extends towards the solvent-accessible surface area, engaging in interactions with hydrophobic residues.

In the region B sub pocket, the pyridine ring and sulfonic acid form stronger hydrogen bonds with Thr82 and Ser86 of the EVA-PGE2 structure. The tert-butyl group fits well in the C sub pocket and forms a complex with residues such as Ile 85, Met116, Ser 305, Ser 308. Comparison of QSAR results with X-ray resolution pose 1 (pdb 1MQ6) in Fig. 11 successfully identified consensus and complementary pharmacophore properties that dominate the agonist activity of lead molecules [36–44].

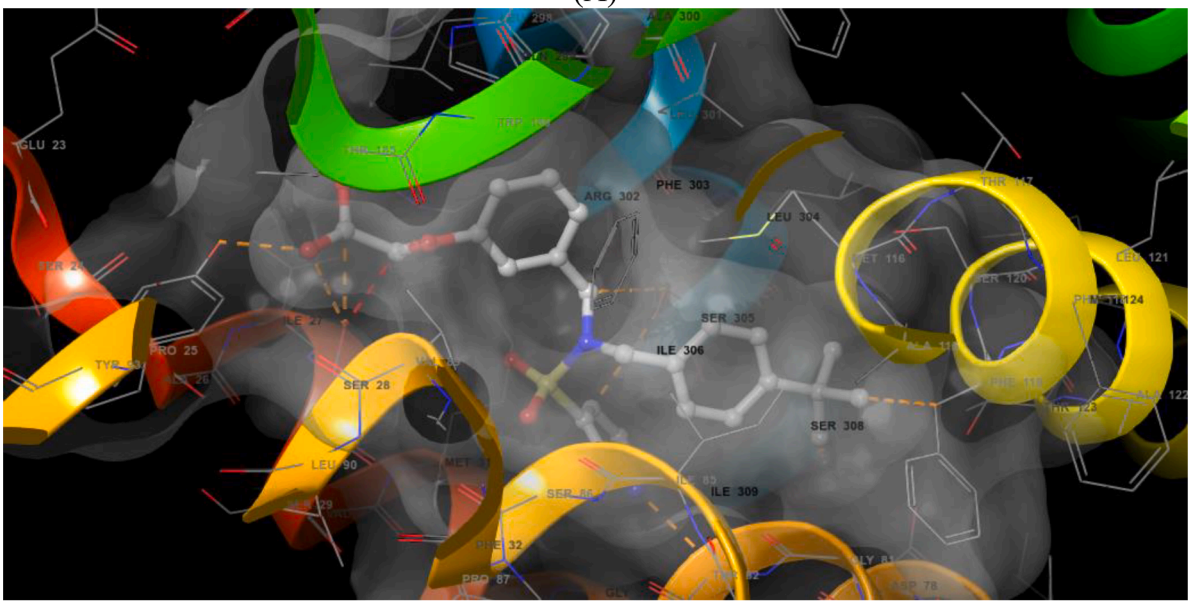
The descriptor fdonH2B emphasizes the importance of the hydrogen atom in the two bonds from the donor atom. A similar function exists in the ligand molecule EVA (CP533536), indicating that: There should be a minimum mass near the donor atom (see Fig. 12). Donor atoms do not interact, but reduce steric collisions between the drug and the receptor. Therefore, the QSAR results are consistent and complement the reported X-ray data.

Conclusions

The current study is to clarify the important pharmacophore functions that dominate the potential agonists activity of the prostaglandin E2 receptor (PGE2): R^2tr = 0.808, Q^2LMO = 0.794, R^2ex = 0.781. A thoroughly validated GAMLR-QSAR model, focusing on six descriptors, has been successfully developed. The QSAR outcomes seamlessly integrate both reported pharmacophore properties and novel pharmacophore properties. This model encapsulates part of the pharmacophore function, potentially amplifying the agonistic activity of the prostaglandin E2 receptor. Noteworthy examples include a hydrogen atom precisely positioned 2 bonds away from a donor atom, a sp2-hybridized



(A)



(B)

Fig. 11. X-ray resolved pose for CP-533536 in the active site of Prostaglandin receptor PGE2 (pdb −7CX4) (A) without surface (B) with surface (Glide, 2023, Schrodinger, LLC, NY).

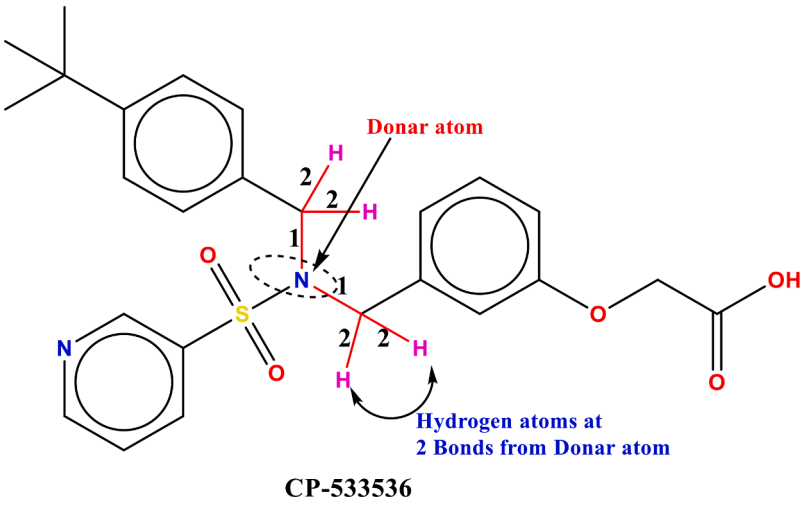


Fig. 12. Display of Molecular Descriptor fdonH2B in selective EP (2) agonist EVA(CP-533536).

carbon atom precisely 2 bonds away from a ring nitrogen atom, and a nitrogen atom within 4 bonds from a sp3 oxygen atom. This facilitates future optimization with ease and creativity. The QSAR model exhibits an admirable balance between predictive power and mechanistic association, further reinforced by published crystal structure data for CP533536 at the active site of the prostaglandin receptor PGE2. These models play a pivotal role in optimizing existing compounds into more potent agonist lead molecules, thereby mitigating inflammatory diseases such as ocular hypertension, bone fractures, and asthma.

Ethics approval and consent to participate

NA.

Consent for publication

Not Applicable.

Availability of data and material

dataset of structurally diverse 309 nitrogen ring containing heterocycles experimentally tested (Ki) for PGE2 agonistic potential has been carefully chosen for QSAR investigation from renown and publicly accessible Binding database (https://www.ebi.ac.uk/chembl/).

Funding

Not applicable.

CRediT authorship contribution statement

Rahul D Jawarkar: Writing – review & editing, Writing – original draft, Visualization, Conceptualization. **Magdi E.A. Zaki:** Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Sami A. Al-Hussain:** Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation. **Abdul Samad:** Conceptualization. **Long Chiau Ming:** Data curation, Conceptualization. **Summya Rashid:** Data curation, Conceptualization. **Gehan M. Elossaily:** Conceptualization. **Susmita Yadav:** Formal analysis, Writing – original draft. **Suraj Mali:** Writing – review & editing, Writing – original draft, Conceptualization.

Declaration of competing interest

The authors whose names are listed in the manuscript certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Data availability

Data will be made available on request.

Acknowledgements

We are thankful to Shri Yogendraji R. Gode, President, Dr. Rajendra Gode Institute of Pharmacy, University Mardi Road, Amravati, for providing extended support and Research Amenities during entire course of Research work. Authors of this manuscript are thankful to Dr. Paola Gramatica for providing a copy of QSARINS 2.24 software. The Glide, 2023 was availed from Schrodinger Drug Discovery package available at Dept. of Pharmacy, Birla Institute of Technology, Mesra, India.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.chphi.2024.100484.

References

[1] A.N. Hata, R.M. Breyer, Pharmacology and signaling of prostaglandin receptors: multiple roles in inflammation and immune modulation, Pharmacol. Ther. 103 (2) (2004) 147–166, https://doi.org/10.1021/cr200010h.
[2] K. Andreasson, Emerging roles of PGE2 receptors in models of neurological disease, Prostaglandins Other Lipid Mediat. 91 (3–4) (2010) 104–112, https://doi.org/10.1016/j.pharmthera.2004.06.003.
[3] A.N. Hata, R.M. Breyer, Pharmacology and signaling of prostaglandin receptors: multiple roles in inflammation and immune modulation, Pharmacol. Ther. 103 (2) (2004) 147–166, https://doi.org/10.1016/j.pharmthera.2004.06.003.
[4] R.L. Jones, M.A. Giembycz, D.F. Woodward, Prostanoid receptor antagonists: development strategies and therapeutic applications, Br. J. Pharmacol. 158 (1) (2009) 104–145, https://doi.org/10.1111/j.1476-5381.2009.00317.x.
[5] R. Iwamura, M. Tanaka, E. Okanari, T. Kirihara, N. Odani-Kawabata, N. Shams, K. Yoneda, Identification of a selective, non-prostanoid PGE2 receptor agonist for the treatment of glaucoma: omidenepag and its prodrug omidenepag isopropyl, J. Med. Chem. 61 (15) (2018) 6869–6891, https://doi.org/10.1021/acs.jmedchem.8b00808.

[6] K. Tani, A. Naganawa, A. Ishida, H. Egashira, K. Sagawa, H. Harada, M. Toda, Design and synthesis of a highly selective PGE2-receptor agonist, Bioorg. Med. Chem. Lett. 11 (15) (2001) 2025–2028, https://doi.org/10.1016/S0960-894X(01)00359-6.

[7] K.O. Cameron, B.A. Lefker, H.Z. Ke, M. Li, M.P. Zawistoski, C.M. Tjoa, D. D. Thompson, Discovery of CP-533536: an PGE2 receptor selective prostaglandin E2 (PGE2) agonist that induces local bone formation, Bioorg. Med. Chem. Lett. 19 (7) (2009) 2075–2078, https://doi.org/10.1016/j.bmcl.2009.01.059.

[8] C. Qu, C. Mao, P. Xiao, Q. Shen, Y.N. Zhong, F. Yang, Y. Zhang, Ligand recognition, unconventional activation, and G protein coupling of the prostaglandin E2 receptor PGE2 subtype, Sci. Adv. 7 (14) (2021) eabf1268, https://doi.org/10.1126/sciadv.abf1268.

[9] M. Abramovitz, M. Adam, Y. Boie, M.C. Carrière, D. Denis, C. Godbout, K. M. Metters, The utilization of recombinant prostanoid receptors to determine the affinities and selectivities of prostaglandins and related analogs, Biochim. Biophys. Acta (BBA)-Molecular Cell Biol. Lipids 1483 (2) (2000) 285–293, https://doi.org/10.1016/S1388-1981(99)00164-X.

[10] K. Tani, A. Naganawa, A. Ishida, H. Egashira, K. Sagawa, H. Harada, M. Toda, Development of a highly selective PGE2-receptor agonist. Part 2: identification of 16-Hydroxy-17, 17-trimethylene 9β-chloro PGF derivatives, Bioorg. Med. Chem. 10 (4) (2002) 1107–1114, https://doi.org/10.1016/S0968-0896(01)00370-4.

[11] M.E. Bunnage, E.L.P. Chekler, L.H. Jones, Target validation using chemical probes, Nat. Chem. Biol. 9 (4) (2013) 195–199.

[12] G.M. Simon, M.J. Niphakis, B.F. Cravatt, Determining target engagement in living systems, Nat. Chem. Biol. 9 (4) (2013) 200–205.

[13] K. Roy, R. Narayan Das, A review on principles, theory and practices of 2D-QSAR, Curr. Drug Metab. 15 (4) (2014) 346–379.

[14] A. Gupta, V. Kumar, P. Aparoy, Role of topological, electronic, geometrical, constitutional and quantum chemical based descriptors in QSAR: mPGES-1 as a case study, Curr. Topics Med. Chem. 18 (13) (2018) 1075–1090, https://doi.org/10.2174/1568026618666180719164149.

[15] D. Usmanov, B. Rasulev, V. Syrov, U. Yusupova, N. Ramazonov, Structure-hepatoprotective activity relationship study of iridoids: a QSAR analysis, Int.J. Quant. Struct. Property Relationships (IJQSPR) 5 (3) (2020) 108–118, https://doi.org/10.4018/IJQSPR.20200701.oa3.

[16] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, A. Tropsha, QSAR modeling: where have you been? Where are you going to? J. Med. Chem. 57 (12) (2014) 4977–5010, https://doi.org/10.1021/jm4004285.

[17] Gramatica, P., Chirico, N., Papa, E., Cassani, S., & Kovarich, S. (2013). QSARINS: a new software for the development, analysis, and validation of QSAR MLR models. https://doi.org/10.1002/jcc.23361.

[18] J. Huang, X. Fan, Why QSAR fails: an empirical evaluation using conventional computational approach, Mol. Pharm. 8 (2) (2011) 600–608, https://doi.org/10.1021/mp100423u.

[19] M.E. Zaki, S.A. Al-Hussain, V.H. Masand, S. Akasapu, I. Lewaa, QSAR and Pharmacophore Modeling of Nitrogen Heterocycles as Potent Human N-Myristoyltransferase (Hs-NMT) Inhibitors, Molecules. 26 (7) (2021) 1834, https://doi.org/10.3390/molecules26071834.

[20] Fourches, D., Muratov, E., & Tropsha, A. (2010). Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J. Chem. Inf. Model.*, 50(7), 1189.10.1021/ci100176x.

[21] P. Gramatica, S. Cassani, P.P. Roy, S. Kovarich, C.W. Yap, E. Papa, QSAR modeling is not "push a button and find a correlation": a case study of toxicity of (benzo-) triazoles on algae, Mol. Inform. 31 (11–12) (2012) 817–835, https://doi.org/10.1002/minf.201200075.

[22] J.C. Dearden, M.T. Cronin, K.L. Kaiser, How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR), SAR. QSAR. Environ. Res. 20 (3–4) (2009) 241–266, https://doi.org/10.1080/10629360902949567.

[23] V. Consonni, D. Ballabio, R. Todeschini, Comments on the definition of the Q 2 parameter for QSAR validation, J. Chem. Inf. Model. 49 (7) (2009) 1669–1678, https://doi.org/10.1021/ci900115y.

[24] M.K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, J. Chong, BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology, Nucleic Acids Res. 44 (D1) (2016) D1045–D1053, https://doi.org/10.1093/nar/gkv1072.

[25] N.A. Hummell, A.V. Revtovich, N.V. Kirienko, Novel immune modulators enhance Caenorhabditis elegans resistance to multiple pathogens, mSphere 6 (1) (2021), https://doi.org/10.1128/mSphere.00950-20 e00950-20.

[26] V.H. Masand, V. Rastija, PyDescriptor: a new PyMOL plugin for calculating thousands of easily understandable molecular descriptors, Chemom. Intell. Lab. Syst. 169 (2017) 12–18, https://doi.org/10.1016/j.chemolab.2017.08.003.

[27] V.H. Masand, D.T. Mahajan, G.M. Nazeruddin, T.B. Hadda, V. Rastija, A.M. Alfeefy, Effect of information leakage and method of splitting (rational and random) on external predictive ability and behavior of different statistical parameters of QSAR model, Med. Chem. Res. 24 (3) (2015) 1241–1264.

[28] V. Consonni, R. Todeschini, D. Ballabio, F. Grisoni, On the misleading use of for QSAR model comparison, Mol. Inform. 38 (1–2) (2019) 1800029, https://doi.org/10.1002/minf.201800029.

[29] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, Open Babel: an open chemical toolbox, J. Cheminform. 3 (1) (2011) 1–14.

[30] R.D. Jawarkar, R.L. Bakal, M.E. Zaki, S. Al-Hussain, A. Ghosh, A. Gandhi, I. Lewaa, QSAR Based Virtual screening derived Identification of a Novel Hit as a SARS CoV-229E 3CLpro inhibitor: GA-MLR QSAR modeling supported by molecular docking, molecular dynamics simulation and MMGBSA calculation approaches, Arab. J. Chem. (2021) 103499.

[31] D. Krstajic, L.J. Buturovic, D.E. Leahy, S. Thomas, Cross-validation pitfalls when selecting and assessing regression and classification models, J. Cheminform. 6 (1) (2014) 1–15.

[32] T.M. Martin, P. Harten, D.M. Young, E.N. Muratov, A. Golbraikh, H. Zhu, A. Tropsha, Does rational selection of training and test sets improve the outcome of QSAR modeling? J. Chem. Inf. Model. 52 (10) (2012) 2570–2578, https://doi.org/10.1021/ci300338w.

[33] N. Chirico, P. Gramatica, Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection, J. Chem. Inf. Model. 52 (8) (2012) 2044–2058, https://doi.org/10.1021/ci300084j.

[34] P.P. Roy, S. Kovarich, P. Gramatica, QSAR model reproducibility and applicability: a case study of rate constants of hydroxyl radical reaction models applied to polybrominated diphenyl ethers and (benzo-) triazoles, J. Comput. Chem. 32 (11) (2011) 2386–2396. https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.21820.

[35] N. Chirico, P. Gramatica, Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient, J. Chem. Inf. Model. 51 (9) (2011) 2320–2335, https://doi.org/10.1021/ci200211n.

[36] S.N. Mali, A. Pandey, Synthesis of new hydrazones using a biodegradable catalyst, their biological evaluations and molecular modeling studies (Part-II), J. Comput. Biophys. Chem. 21 (07) (2022) 857–882.

[37] S. Ghosh, S.N. Mali, D.N. Bhowmick, A.P. Pratap, Neem oil as natural pesticide: pseudo ternary diagram and computational study, J. Indian Chem. Soc. 98 (7) (2021) 100088.

[38] B.S. Jadhav, R.S. Yamgar, R.S. Kenny, S.N. Mali, H.K. Chaudhari, M.C. Mandewale, Synthesis, in silico and biological studies of thiazolyl-2h-chromen-2-one derivatives as potent antitubercular agents, Curr. Comput. Aided. Drug Des. 16 (5) (2020) 511–522.

[39] R. Kshatriya, D. Kambale, S. Mali, V.P. Jejurkar, P. Lokhande, H.K. Chaudhari, S. Saha, Brønsted acid catalyzed domino synthesis of functionalized 4H-chromens and their ADMET, molecular docking and antibacterial studies, ChemistrySelect. 4 (27) (2019) 7943–7948.

[40] E.S. Aazam, R. Thomas, Understanding the behavior of a potential anticancer lamotrigine in explicit solvent (water and DMSO) using quantum mechanical tools and abinitio molecular dynamics, Chem. Phys. Impact 8 (2024) 100404.

[41] K.J. Rajimon, N. Sreelakshmi, D.S.R. Nair, N.F. Begum, R. Thomas, An in-depth study of the synthesis, electronic framework, and pharmacological profiling of 1-(anthracen-9-yl)-N-(4-nitrophenyl) methanimine: in vitro and in silico Investigations on Molecular docking, dynamics simulation, BSA/DNA binding and toxicity studies, Chem. Phys. Impact (2024) 100462.

[42] E.S. Aazam, R. Thomas, Solution stage fluorescence and anticancer properties of azomethine compounds from sulpha drugs: synthesis, experimental and theoretical insights, J. Mol. Struct. 1295 (2024) 136669.

[43] G. Ashok, F. Thomas, R. Thomas, Understanding the hydrogen bonding preferences and dynamics of Prontosil in water and methanol, Chem. Phys. Impact (2024) 100453.

[44] R.A. Kumar, S.J. Al-Otaibi, Y.S. Mary, Y.S. Mary, N. Acharjee, R. Thomas, R. R. Pillai, T.L. Leena, Surface adsorption of adenine on pristine and B/N/O/P-doped coronene as a biosensing substrate for DNA detection-DFT study, J. Mol. Liq. 393 (2024) 123546.