Molecular Diversity (2024) 28:2099–2118
https://doi.org/10.1007/s11030-023-10782-9

ORIGINAL ARTICLE

Explainable artificial intelligence-assisted virtual screening and bioinformatics approaches for effective bioactivity prediction of phenolic cyclooxygenase-2 (COX-2) inhibitors using PubChem molecular fingerprints

Mithun Rudrapal[1] · Kevser Kübra Kirboga[2,3] · Mohnad Abdalla[4] · Siddhartha Maji[5]

Received: 11 July 2023 / Accepted: 22 November 2023 / Published online: 10 January 2024

Abstract

Cyclooxygenase-2 (COX-2) inhibitors are nonsteroidal anti-inflammatory drugs that treat inflammation, pain and fever. This study determined the interaction mechanisms of COX-2 inhibitors and the molecular properties needed to design new drug candidates. Using machine learning and explainable AI methods, the inhibition activity of 1488 molecules was modelled, and essential properties were identified. These properties included aromatic rings, nitrogen-containing functional groups and aliphatic hydrocarbons. They affected the water solubility, hydrophobicity and binding affinity of COX-2 inhibitors. The binding mode, stability and ADME properties of 16 ligands bound to the Cyclooxygenase active site of COX-2 were investigated by molecular docking, molecular dynamics simulation and MM-GBSA analysis. The results showed that ligand 339,222 was the most stable and effective COX-2 inhibitor. It inhibited prostaglandin synthesis by disrupting the protein conformation of COX-2. It had good ADME properties and high clinical potential. This study demonstrated the potential of machine learning and bioinformatics methods in discovering COX-2 inhibitors.

Graphical abstract

This study uses machine learning, bioinformatics and explainable artificial intelligence (XAI) methods to discover and design new drugs that can reduce inflammation by inhibiting COX-2. The activity and properties of various molecules are

✉ Mithun Rudrapal
rsmrpal@gmail.com

✉ Kevser Kübra Kirboga
kubra.kirboga@bilecik.edu.tr

Mohnad Abdalla
mohnadabdalla200@gmail.com

Siddhartha Maji
smaji@okstate.edu

[1] Department of Pharmaceutical Sciences, School of Biotechnology and Pharmaceutical Sciences, Vignan's Foundation for Science, Technology & Research (Deemed to Be University), Guntur 522213, India

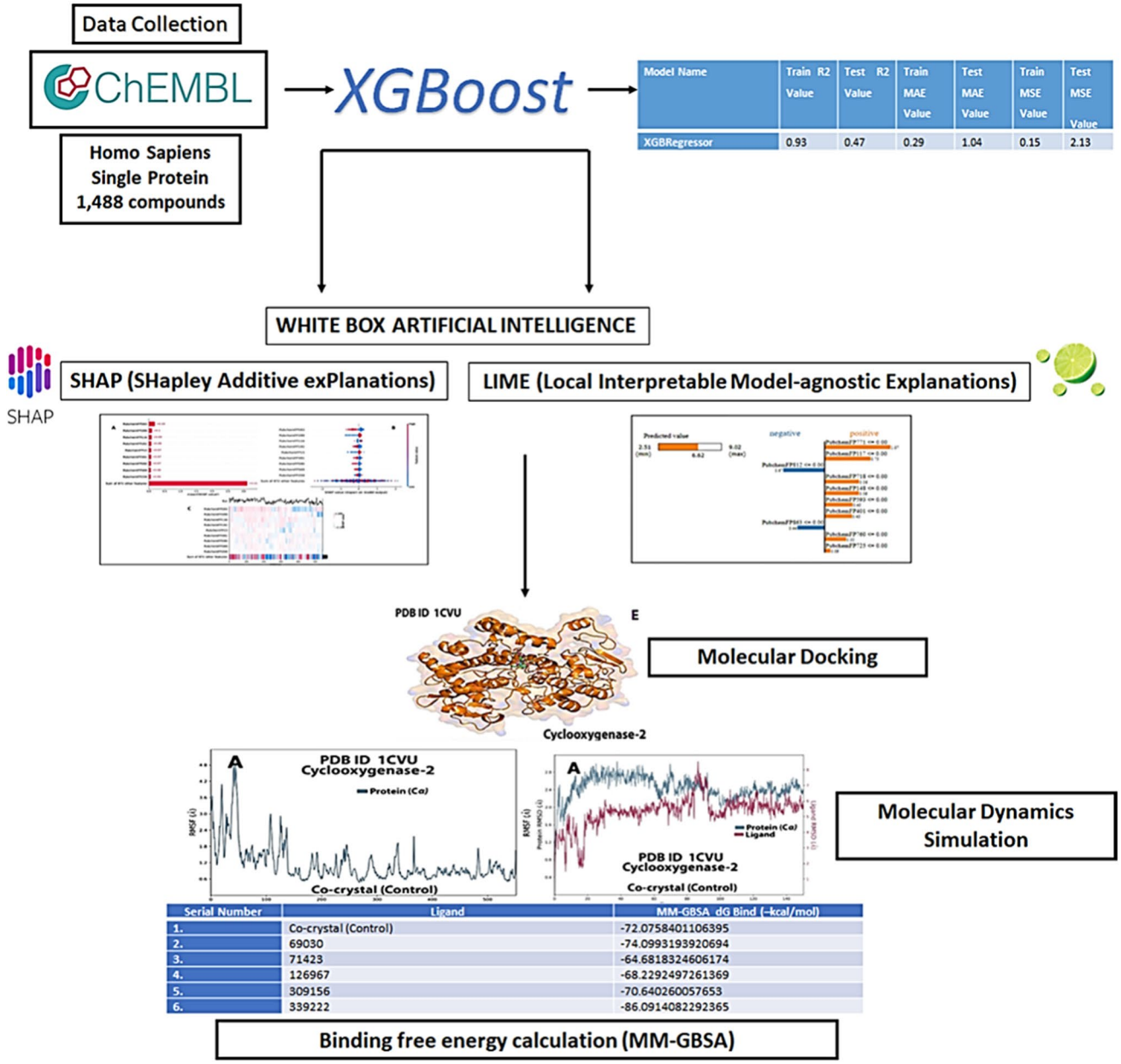
[2] Informatics Institute, Istanbul Technical University, 34469 Maslak, Istanbul, Turkey

[3] Bioengineering Department, BilecikSeyhEdebali University, 11230 Bilecik, Turkey

[4] Pediatric Research Institute, Children's Hospital Affiliated to Shandong University, Jinan 250022, Shandong, People's Republic of China

[5] Department of Chemistry, Oklahoma State University, Stillwater, OK, USA

modelled and analysed. The best molecule is selected, and its interaction with the enzyme is investigated. The results show how this molecule can block the enzyme and prevent inflammation. XAI methods are used to explain the molecular features and mechanisms involved.



Keywords Cyclooxygenase-2 · Explainable artificial intelligence · Shapley explanations · Molecular dynamics · COX-2 inhibitors

Introduction

Cyclooxygenase-2 (COX-2) inhibitors, a member of the group of nonsteroidal anti-inflammatory drugs (NSAIDs) used to treat conditions such as inflammation, mild to moderate pain, and fever, directly target the COX-2 enzyme that creates prostaglandins [1]. COX-2 inhibitors reduce prostaglandin E2 (PGE2) synthesis by inhibiting the COX-2 enzyme. PGE2 is a prostaglandin that protects the gastric mucosa and reduces acid secretion [2]. Therefore, COX-2 inhibitors reduce the risk of peptic ulceration. However, COX-2 inhibitors also reduce prostacyclin (PGI2) synthesis.

PGI2 is a prostaglandin that is a vasodilator and prevents platelet aggregation [3, 4]. Therefore, COX-2 inhibitors increase the risk of cardiovascular side effects. Therefore, elucidating the molecular interaction mechanisms between COX-2 inhibitors and COX-2 is vital for evaluating the efficacy and safety of COX-2 inhibitors and designing new and better drug candidates [5]. COX-2 inhibitors prevent the oxygenation of arachidonic acid (AA) by binding to the active site of the COX-2 enzyme [6]. AA is the precursor of prostaglandins such as PGE2 and PGI2. COX-2 inhibitors differ by their binding affinity and selectivity with COX-2 [7]. In this study, the binding affinities and selectivities of phenolic compounds with COX-2 were calculated by molecular docking, molecular dynamics simulation and MM-GBSA analyses. These analyses reveal phenolic compounds' potential as COX-2 inhibitors and provide important information for designing new drug candidates. It should be noted that in this study, phenolic compounds were chosen as COX-2 inhibitors, and it was hypothesised that these compounds could overcome peptic ulceration and cardiovascular side effects. However, it should be emphasised that this assumption needs to be confirmed by in vitro or clinical experiments because the study aims to find inhibitors that overcome these side effects. Phenolic compounds play an important role in slowing inflammation because they can prevent or treat various chronic diseases that occur due to inflammation [8]. Phenolic compounds show anti-inflammatory capacity by inhibiting the production or action of pro-inflammatory mediators involved in the inflammation process [9]. These mediators include cyclooxygenase (COX), tumour necrosis factor-α (TNF-α), nuclear factor-kappa B (NF-κB) and nitric oxide (NO) [10]. In 2021, three phenolic acids (p-coumaric acid, caffeic acid and gallic acid) were isolated from the resin of Eucalyptus maculata. These compounds and their semi-synthetic derivatives were designed as potential anticancer agents with dual inhibitory activity against EGFR and COX-2. Comprehensive biological evaluations and in silico analyses of these phenolic compounds and their relation to COX-2 inhibition were conducted [11]. In 2022, the design, biological evaluation and silico analysis of phenolic compounds as potential anticancer agents that dual inhibit EGFR and COX-2 were performed [1].In 2023, four phenolic compounds isolated from Eucalyptus maculata resin, their anti-inflammatory effects, and their capacity to inhibit COX-1, COX-2, TNF-α, NF-κB and NO were investigated [12]. Other mechanisms, target selectivity, molecular structures and novel interactions of compounds in the studies related to COX-2 inhibition have not yet been fully elucidated. Therefore, it is clear that more studies are needed for these compounds to be safer and more effective as anti-inflammatory agents. In this study, molecular docking, molecular dynamics simulation and MM-GBSA analyses of COX-2 inhibitors and their interactions with the Cyclooxygenase active site and binding energies of COX-2 were investigated using bioinformatics methods. In addition, inhibition activities of COX-2 inhibitors were modelled, and important molecular properties were determined using machine learning and explainable artificial intelligence methods. Machine learning and explainable artificial intelligence (XAI) methods have been used to model the inhibitory activities of COX-2 inhibitors and identify important molecular features. These methods reveal complex relationships regarding the efficacy and selectivity of COX-2 inhibitors and explain these relationships understandably. This way, new strategies can be developed to design and optimise COX-2 inhibitors. This study demonstrates the potential of bioinformatics and machine learning methods in discovering COX-2 inhibitors.

Materials and methods

Explainable artificial intelligence

The XAI part of the project involved data collection, molecular fingerprinting, machine learning and XAI application steps. Python 3.11 was used for the research, along with the following libraries: Matplotlib 3.7.1, pip 22.0.4, Sklearn 1.2.2, Pandas 2.0.1, RDkit 2023.3.1, Shap 0.41.0, eli5 0.13.0, scikit-plot 0.3.7 and NumPy 1.24.3. The application ran on a computer with Intel® Core™ i5-8300H CPU 2.30GHz, 64-bit OS, × 64 processor, and 32 GB RAM.

Data compilation and curation

Datasets were collected from the version 25 ChEMBL database for COX-2 inhibition [13]. Data were "Homo sapiens" and "Single Protein" filtered. The IC50 selection for the bioactivity unit was attained through a data creation approach. pIC_{50} may be estimated for substances whose IC_{50} values were determined using allopurinol as a positive control. As a result, 3643 compounds were collected for COX-2. We developed a regression model for COX-2 inhibition. We set threshold values < 1 and > 10 μM to distinguish active and inactive substances. We classified the compounds according to their cell-killing potency: highly active (IC_{50} < 1 μM), moderately active (1 μM < IC_{50} < 10 μM), weakly active (IC_{50} > 10 μM) [14] and inactive (IC_{50} > 50 μg/mL) [15, 16]. After these procedures, a non-redundant and enhanced final dataset for COX-2 inhibition with 1,488 chemicals was produced.

Descriptors of molecular properties

PaDEL-descriptor software was used to determine the molecular properties. PaDEL-descriptor software is used for calculating molecular identifiers and fingerprints. The software

calculates 797 identifiers (663 1D, 2D identifiers and 134 3D identifiers) and 10 types of fingerprints. These identifiers and fingerprints are primarily calculated using The Chemistry Development Kit [17]. PaDEL-descriptor can also calculate PubChem fingerprints. PubChem fingerprints are a type of fingerprint that represents chemical structures as binary sequences and are used to measure chemical similarities [17]. PubChem fingerprints are created based on the atomic properties of chemical structures, bond types, ring systems and other properties [17]. When calculating PubChem fingerprints, the PaDEL-descriptor creates an 881-bit vector specified by PubChem [17]. Each bit of this vector indicates whether the chemical structure has a particular property.

Data split and model development & evaluation

Data split

After obtaining a data collection machine learning models can handle, we split our dataset into two subgroups. One of these subgroups was the training set, which had a higher data percentage and was used to develop the model. The other subgroup was the more limited data set used to test the model. We split our dataset into training and test sets at a ratio of 80%–20%.

Model development

The prediction model in the current work incorporates COX-2 inhibition, and the predictors are PubChem chemical fingerprints. Three statistical measures, Coefficient of Determination (R2) (formula 1), Mean Square Error (MSE) (formula 2) and Mean Absolute Error (MAE) (formula 3), are used to assess the model's predictive power (Table 1).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}_i\right)^2} \quad (1)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \quad (2)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \quad (3)$$

Model evaluation

After training our model, we used cross-validation (CV) methods to measure its performance. CV methods also include performance prediction for unknown data and generalisation degree measurement. We used the ten-fold CV method. In this method, the original dataset is divided into two subgroups (training set and validation set) in each experiment iteration. We selected the parameters that evaluate the performance of each model and determined the model with the best performance. Then, we completed the final validation of the best model. If the validation results are statistically significant, we can say that we have created a new drug prediction model [18, 19].

Mechanistic interpretations of permutation importance

Any acceptable estimator can be employed when the data are tabulated using a permutation feature significance model assessment technique. This is especially helpful for nonlinear or opaque estimators. The definition of permutation feature significance is the decrease in model score brought on by the random rearranging of a single feature value [20]. The technique breaks the connection between the feature and the goal, and the model score decline shows how much the feature is believed to be reliable. This technique has the benefit of being model agnostic and may be computed more than once with different feature permutations. The significance of permutation indicates how significant a feature is to a specific model rather than the intrinsic predictive value of a feature on its own.

Permutation feature importance is the reduction in model score that occurs by randomly rearranging a single feature value [21]. This method breaks the connection between the feature and the target, and the decrease in the model score

Table 1 Statistical values of the evaluated four models

Model name	Train R2 value	Test R2 value	Train MAE value	Test MAE value	Train MSE value	Test MSE value
RandomForestRegressor	0.81	0.03	0.42	1.06	0.31	1.74
XGBRegressor	**0.93**	**0.47**	**0.29**	**1.04**	**0.15**	**2.13**
DecisionTreeRegressor	0.91	0.37	0.10	1.14	0.24	2.85
MLPRegressor	0.63	0.12	0.60	1.15	0.63	2.02

indicates how reliable the feature is. This method has the advantages of being model-independent and calculated more than once with different feature permutations. Permutation importance indicates how important a feature is for a particular model. In this study, we applied PFI to the dataset containing all PubChem identifiers. After determining the most effective descriptors on the model thanks to PFI (Permutation Feature Importance), we selected the 500 most essential descriptors. We created a new dataset with these top 500 descriptors and applied XAI to this dataset with the XGBoost model (bold in Table 1). In this way, we could see more clearly the impact of the most important descriptors on the model [20, 22].

Explainability with Shapley additive explanations (ShAP)

Among the different explain ability strategies, ShAP (Shapley additive explanations) selection benefits from several qualities [23, 24]. It will also be suitable for the XGBoost-model we chose for COX-2 inhibition. Additionally, it provides ShAP values for completeness, correctness and consistency [24–26]. Moreover, ShAP applications are simple and easily documented with pictures. Initially, Shapley's values were presented as a game theory to adequately compensate participants according to their contribution to their overall gains [27]. An estimated model's quantitative significance is determined based on contributions. The average marginal contribution of the feature value to each possible combination of all the features in your model can be considered the ShAP value in our investigation. The discrepancy between the actual and average expectations for the whole set of molecular fingerprints may be used to explain a ShAP value for a specific molecular fingerprint value. When dealing with ShAP values, we should note that the projected value remains the same when a related feature is removed [28]. The Shapley values of each feature are determined using the ShAP method and are then represented as a linear model of all components. ShAP values employ a specific index for each feature across all binary features. By doing so, ShAP values make it simple to represent interactions that might otherwise go unnoticed. This feature is crucial because it helps to understand the variables in the model and how they relate to one another [25]. We estimated the TreeShAP values of the best-performing ML-based model using the COX-2 inhibition data with Python as a training and test set [29]. To illustrate the significance of molecular fingerprints for the models, we used ShAP values. To assess how these chemical fingerprints contribute to the output of both models and their relevance in the models, we next constructed ShAP value plots for both models. The crucial interactions between these molecular fingerprints and their targets were examined [30]. Finally, we created individual (local) and global plots to comprehend how different factors influence the effectiveness of COX-2 inhibition. We then decided on the ShAP results and the impact of COX-2 inhibition on the drug discovery process.

Local interpretable model-agnostic explanations (LIME)

LIME (Local Interpretable Model-agnostic Explanations) aims to improve local proxy models for specific expectations. Black-box machine learning techniques, local proxy models, or understandable models are utilised to explain predictions. Redundant models are built to closely match the underlying black box model's predictions. LIME focuses on developing local proxy models rather than global ones. LIME is model agnostic; therefore, it may be used with any machine learning model. The technique attempts to comprehend the model by dissecting the input of the data samples and seeing how the predictions vary [24, 31]. The explainer-LIME approach was utilised to construct estimation values in this study.

Docking methods

Protein preparation

The molecular docking method integrated the ligand orientations in the Cyclooxygenase active site of COX-2 to provide more exact and realistic initial ligand-bound coordinates. The crystal structure of the Cyclooxygenase active site of COX-2 [32] was extracted from the website https://www.rcsb.org (PBD ID: 1CVU) and used to build initial 3D coordinates. At first, Co-crystallised waters were removed, then polar hydrogen was added and hydrogen bonds were optimised. The OPLS force field [33] was used to minimise energy. pH was maintained at 7 during the preparation of the entire protein preparation.

Grid box generation of receptor

Grid box generation at the Cyclooxygenase active site of COX-2 was the most challenging part, and it was accomplished by the Glide module, which enables the Grid box in addition to van der Waals radius for scaling factor (1.0) and partial charge cut-off (0.25) without the need of force [34]. It implied the binding site by selecting the pre-docked ligand and placing a cubical-shaped box on the pre-dock ligand's centroid with the grid box spacing of 20 Å [35].

Ligand preparation and docking

Marvin Sketch 20.10 [36] was used to draw the 2D structure of ligands and converted to 3D using Maestro12.3. The ligands were developed using the "LigPrep" module by

minimising using the OPLS_2005 force field; pH was kept constant at 7 [37]. Maestro 12.3 (Schrödinger) was employed to run the docking simulations of sixteen selected molecules to the Cyclooxygenase active site of COX-2 with the "Extra Precision (XP)" model, and the rest of the parameters were kept as default [14]. Binding interactions were analysed using Maestro 12.3 (Schrödinger) [35, 38–56].

Predictive ADME studies

Swiss-ADME (http://www.swissadme.ch/): a free web tool offered by the Swiss Institute of Bioinformatics, was used to determine ADME characteristics [57, 58]. The test compounds were created using the molecular sketcher on the server website (http://www.swissadme.ch). This structure was translated to SMILES format (the program's real input). The physicochemical characteristics, lipophilicity, drug likeliness and bioavailability radar, among other things, were studied, which provide a pharmacokinetic profile of the drug [59].

Molecular dynamic simulation

Although molecular docking investigations have provided binding mode interaction of protein inhibitors in complex, molecular dynamics simulation can detect the tiniest disagreement [60–65]. The chosen compounds were studied for possible atomic features inside the solvent system using the best interactions and energy-docked conformation acquired from docking study findings. So, to do so, MD simulation was employed using the Schrodinger Desmond module [61]. The system was immersed in a periodic TIP3P water bath, which extended approximately 10.0 Å in each direction using the Desmond System Builder tool. A suitable number of counter-ions were used to neutralise the complexes. 0.15 M NaCl was added to the simulation panel to maintain the isosmotic condition in the system. Equilibration was carried out until the system reached a stationary state before the simulation started. A production MD run for 150 ns was carried out with a target pressure of 1 bar and a temperature of 310 K, respectively. A simulation interaction diagram was used to investigate the MD simulation data comprehensively. The root means square deviation (RMSD) of the protein and selected compound complex, the protein's root means square fluctuation (RMSF), the protein–ligand interaction diagram, the interacting amino acid residues with the ligand in each trajectory frame and the trajectory of different ligand properties were all derived.

Binding free energy calculation

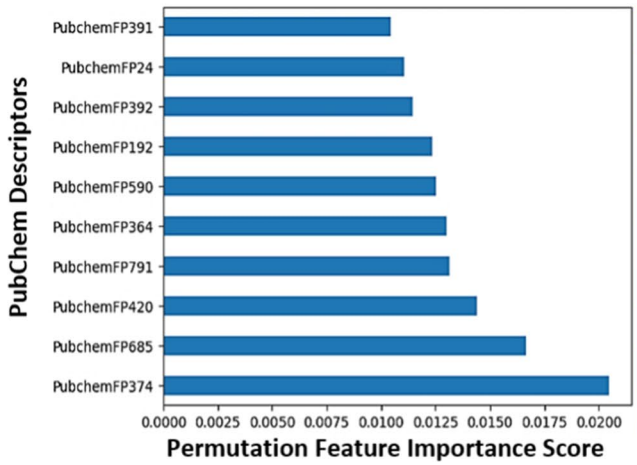
Maestro 12.3 was used to calculate the relative binding free energies (△ Gbind) of the protein-selected compound complex [66] using the Prime Molecular Mechanics Generalised Born Surface Area (MM-GBSA) module. The OPLS-2005 force field was used to calculate the binding free energies (kcal/mole) with the novel solvent model system in VSGB 2.0, and the rest of the parameters were kept as default [67].

Results and discussion

This study presents an innovative computer-aided drug design approach that integrates bioinformatics and machine learning methods to evaluate the potential of phenolic compounds as COX-2 inhibitors. The presented method can be accepted as a prototype in research on COX-2 inhibitors and also has the potential to be adapted for other drug design problems. The details we present in our study are fully included in the article, from the model used to the algorithms to the data set information. These details ensure that the study is both reproducible and buildable by other researchers. In this study, we established QSAR models for COX-2 inhibition. For model development, performance and validation, we applied PFI to 1488 molecules obtained from the ChEMBL database and used 500 molecules. 400 molecules were allocated as the training set and 100 as the test set.

To build our model, we used the PubChemFP fingerprint, which consists of 881 bits representing molecules' chemical structure and binding sites. We used the XGBoost regression method to evaluate our model's performance. This method is a machine learning method consisting of a large number of decision trees. This method provides high accuracy in both classification and regression problems. The R2 value in the training set of our model is 0.93, and the R2 value in the test set is 0.47 (Table 1). In this study, a model was developed using different ML algorithms to predict the drug potential of COX-2 inhibitors based on bioactivity data. In addition, these models were trained to predict the inhibitory potential of the inhibitors based on the chemical properties of the COX-2 inhibitors in the dataset. Table 1 shows the performance of four different algorithms: RandomForestRegressor, XGBRegressor, DecisionTreeRegressor and MLPRegressor. The performance of these algorithms was evaluated by criteria such as R2 values, mean absolute error (MAE) and mean square error (MSE). The results revealed that the XGBRegressor algorithm performed best.

The PFI method is a method used to determine feature selection. This method determines the essential features by measuring how much the model's accuracy decreases when a feature (molecular fingerprint) is removed. Therefore, this method is useful in feature selection and has helped you identify the three most critical molecular fingerprints (Fig. 1). PubchemFP374 represents aromatic rings



PubchemFP391	N(~C)(~C)(~C)
PubchemFP24	>= 2 F
PubchemFP392	N(~C)(~C)(~H)
PubchemFP192	>= 3 any ring size 6
PubchemFP590	C-C:C-O-[#1]
PubchemFP364	C(~F)(:C)
PubchemFP791	NC1CCC(N)CC1
PubchemFP420	C=O
PubchemFP685	O=C-C-C-C-N
PubchemFP374	C(~H)(~H)(~H)

Fig. 1 Feature importance plot for COX-2 inhibition models

and hydrogen bonds of molecules, while PubchemFP685 contains hydrogen bonds and double bonds of molecules. PubchemFP420, on the other hand, represents the properties of molecules associated with hydrogen bonds, double bonds and carboxylic acids. These properties are essential chemical properties that affect the activities of COX-2 inhibitors and can, therefore, be used in molecular design and exploration studies (Figs. 2, 3, and 4).

The XAI methods SHAP and LIME applied to the dataset are a set of algorithms used to determine the order of importance of molecular fingerprints. The SHAP method measures the contribution of each feature to the model prediction by considering every combination of a feature. As a result, it was determined that PubchemFP593, PubchemFP899 and PubchemFP116 are important for COX-2 inhibitors. Higher values of these properties cause the model to predict higher COX-2 inhibition activity, while lower values give lower activity estimates. The LIME method is explicitly used to describe the predicted result. This method measures the effect of features in a given sample and helps explain why a prediction is made. As a result, PubchemFP771, PubchemFP117 and PubchemFP718 were determined to be essential for COX-2 inhibitors.

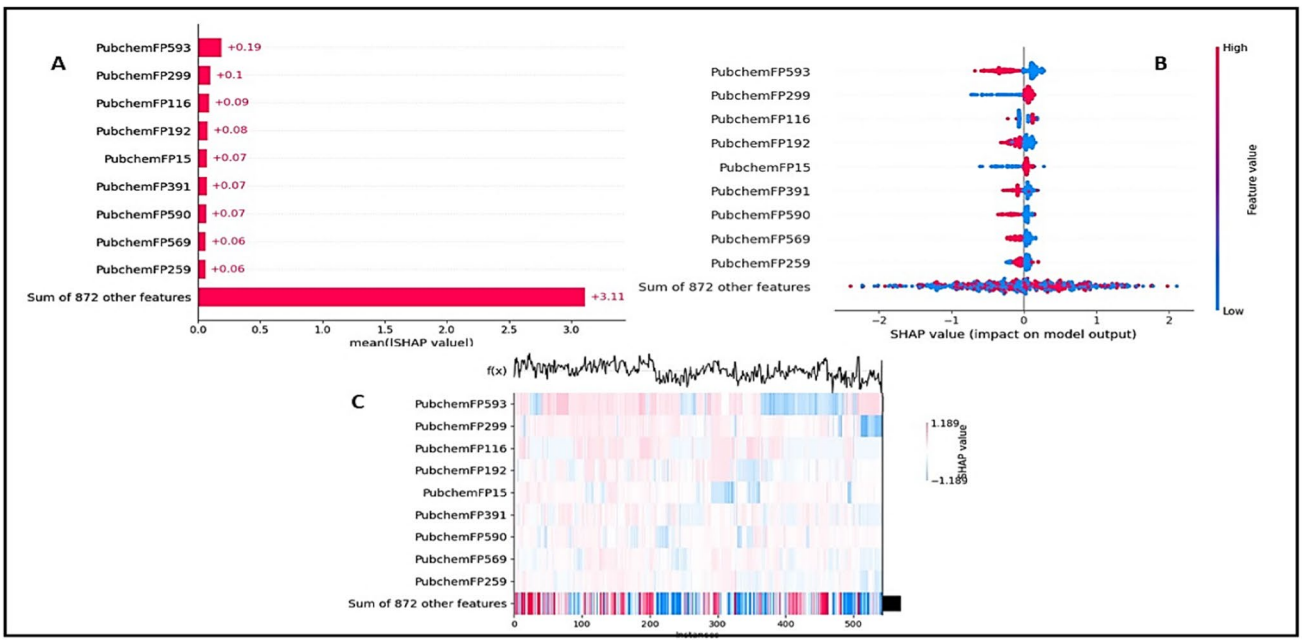


Fig. 2 Bar graphs (**A**), Beeswarm (**B**), and Heatmap (**C**) were created according to ShAP values for COX-2 inhibition

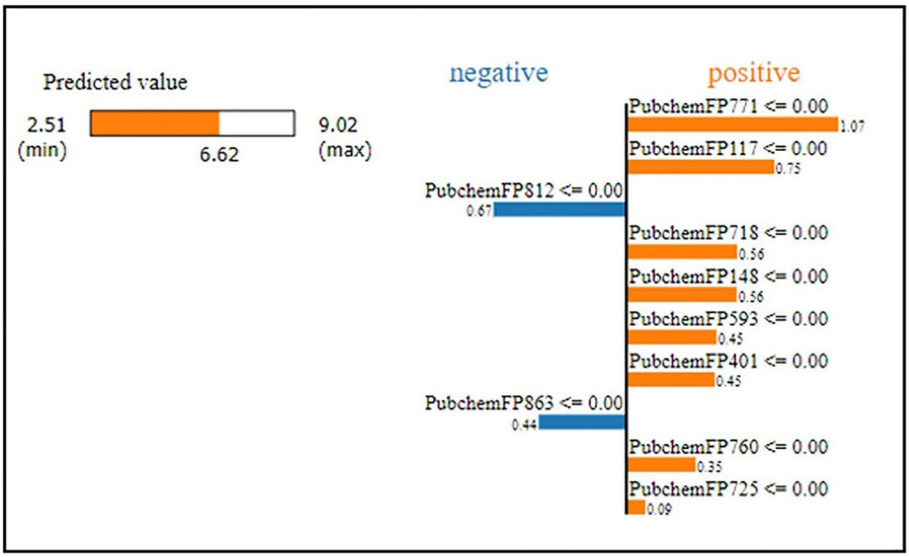


Fig. 3 The LIME graph created for COX-2 inhibition is shown. This chart colour -codes the contribution of each feature to the model prediction. Orange-coloured features are features that increase COX-2 inhibition. Blue-coloured features are features that reduce COX-2 inhibition. The length of the features expresses the magnitude of their impact on the model prediction

Shapley additive explanations (ShAP)

PubchemFP593	N-C-C-C-N
PubchemFP299	N-H
PubchemFP116	>= 1 saturated or aromatic carbon-only ring size 3
PubchemFP192	>= 3 any ring size 6
PubchemFP15	>= 2 N
PubchemFP391	N(~C)(~C)(~C)
PubchemFP590	C-C:C-O-[#1]
PubchemFP259	>= 3 aromatic rings

Local Interpretable Model agnostic Explanations (LIME)

PubchemFP771	Nc1c(Cl)cccc1
PubchemFP117	>= 1 saturated or aromatic nitrogen-containing ring size 3
PubchemFP812	NC1CC(N)CCC1
PubchemFP148	>= 1 unsaturated non-aromatic nitrogen-containing ring size 5
PubchemFP401	N(~O)(~O)
PubchemFP760	Cc1c(Br)cccc1
PubchemFP725	Sc1ccc(N)cc1
PubchemFP718	Cc1ccc(Br)cc1

Fig. 4 Important fingerprints and bit substructure information were revealed by SHAP and LIME methods

The PubchemFP593 property specifies whether the molecule contains an aromatic nitro group [68]. The PubchemFP899 property specifies whether the molecule contains an aromatic sulfonamide group [69]. The PubchemFP116 property specifies whether the molecule contains an aromatic nitro group [69]. The PubchemFP771 feature indicates an aromatic sulfone group in the molecule, and the PubchemFP117 feature indicates an aromatic hydroxyl group. PubchemFP718, on the other hand, means whether there is an aromatic sulfonyl group in the molecule. These properties can be used to predict biological activities such as COX-2 inhibition. These properties affect the compatibility and selectivity of the molecules to the COX-2 enzyme [1].

Post-docking analysis

All the selected molecules (16 ligands) were subjected to a molecular docking study in the cyclooxygenase active site of COX-2 (PDB ID: 1CVU) [32]. All the binding energies with

best scores (bold) and docking scores aregiven in Table 2. After going through various literature reviews [38, 70–72], it has been observed that the cyclooxygenase active site of COX-2 consists of different amino acids such as Met113, Val116, Phe205, Phe209, Val228, Val344, Ile345, Tyr348, Val349, Leu352, Leu359, Leu384, Tyr385, Trp387, Phe518, Ala527, Val523, Ser530 and Leu534. Among which most of them are aromatic and form Hydrophobic pocket into the active site, and residue 530 to 587 plays a crucial role in catalysing the reaction, which converts the Arachidonic acid to Prostaglandin H_2 followed by Prostaglandin G_2 [32, 37, 67, 73–80]. 16 molecules were selected and subjected to docking simulation, but only the best-scoring molecules were analysed to reduce the cost and time.

Table 2 List of selected molecules with their binding affinity in cyclooxygenase active site of COX-2 (PDB ID: 1CVU)

Ligand	Binding energy	Size	Torsional DOF	AutoDock elements
Co-crystal	–8.4	22	14	C OA
69,030	**–9.6**	25	3	A C Cl OA N S HD
69,683	–9.1	26	4	A C S OA Cl
71,423	**–9.7**	25	3	A C S OA Cl
98,492	–7.4	21	4	A C SA
126,967	**–9.9**	25	3	A C F I OA N S HD
144,281	–6.1	13	4	A C SA OA Cl
145,626	–8.0	21	8	A C SA OA
172,194	–8.5	23	3	A C S OA N
173,169	–8.9	24	4	A C HD S OA
185,639	–8.7	25	4	A C Cl NA OA N S HD
306,639	–8.7	24	3	A C S OA F
309,156	**–10.4**	25	3	A C F Cl OA S
319,099	–9.0	25	4	A C S OA SA
322,305	–9.3	31	6	A C F NA OA N S HD
339,222	**–9.8**	30	5	A C F OA N S Br
342,317	–6.2	13	4	A C SA OA

In the present study, the Arachidonic acid (co-crystal) was redocked in PDB ID: 1CVU to test the reliability and reproducibility of the docking protocol for our study (Figs. 5a and 6a). The Carboxylic Acid group of Arachidonic acid has formed only one Hydrogen bond with Ser530. Ligand 69,030 has formed four hydrogens with four different amino acids (Figs. 5b and 6b). The sulfonamide group has formed a hydrogen bond with the amino group of Phe518 (S=O–NH_2). The heteroatom (Nitrogen) of dihydropyridine moiety has formed an H-bond with the carboxylic acid of Met522 (NH–OH) and entirely occupied the hydrophobic pocket.

In contrast, the 4-Pyrone group has formed an H-bond with Ser530, like our native ligand. The methyl group in the sulfonamide moiety has formed a covalent bond with the terminal Amino group of Arg513. The sulfonamide group of Ligand 71,423 has formed a hydrogen bond (Figs. 5c and 6c) with the terminal amino group of Arg120 (S=O–NH_2) and inhibits the translation of the prostaglandin [81, 82]. In the Ligand 126,967, the two oxygen atoms of the Sulfonamide

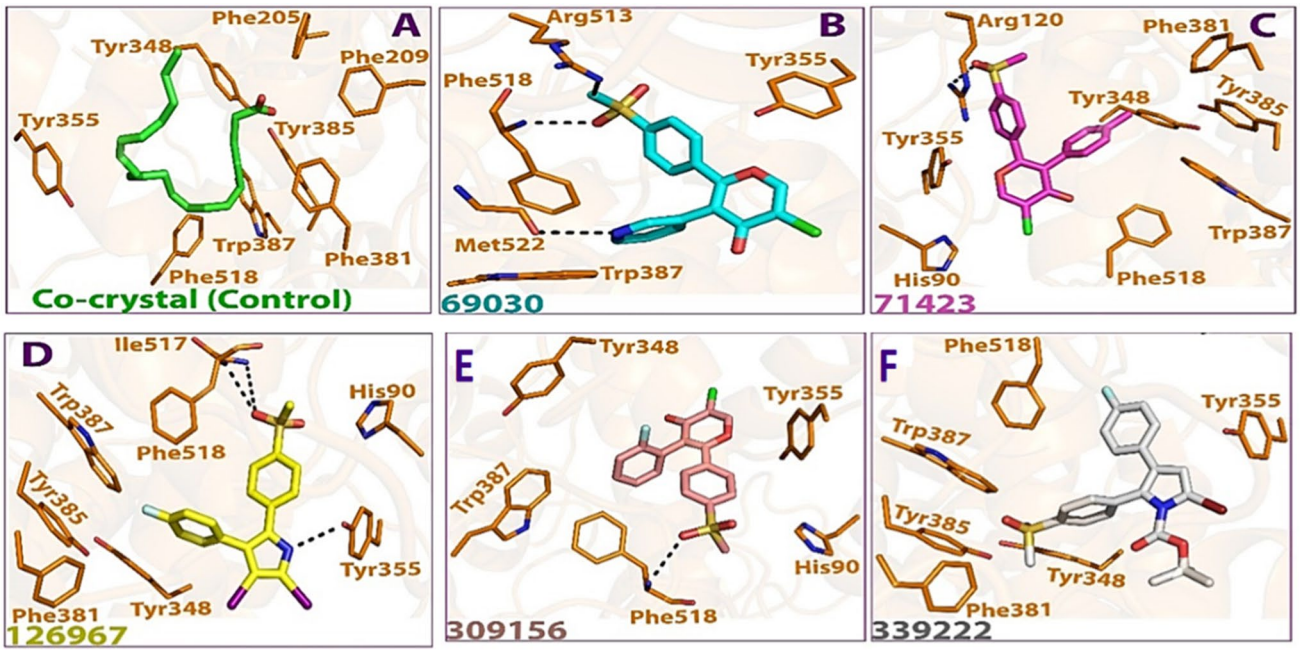


Fig. 5 3D representation of Binding interaction of the selected ligands in cyclooxygenase active site of COX-2 (PDB ID: 1CVU). The receptor is shown in saffron colour, and H-bond interactions are shown in black dashed lines

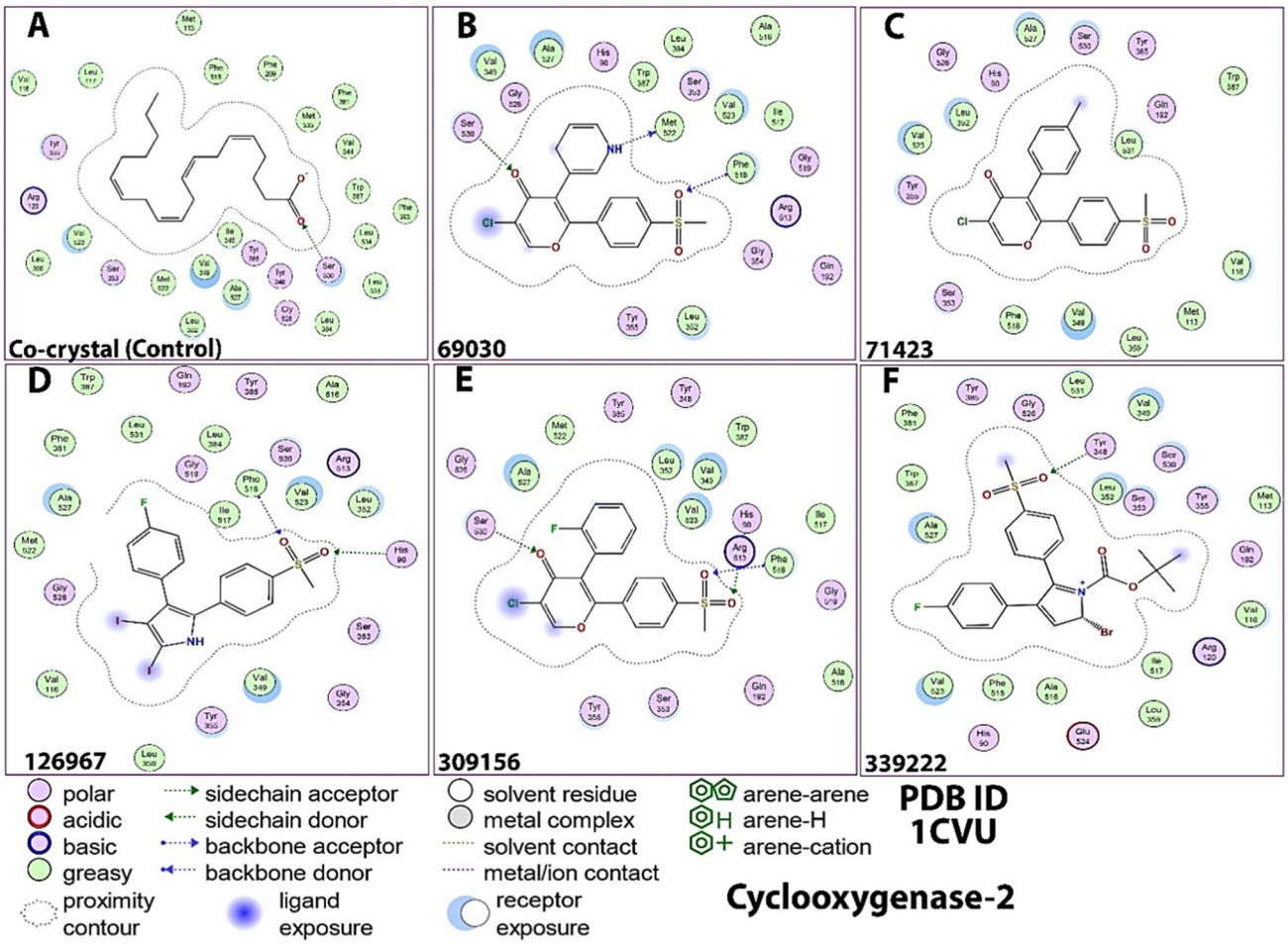


Fig. 6 2D representation of binding interaction of the selected ligands in cyclooxygenase active site of COX-2 (PDB ID: 1CVU). H-bonds are shown in Blue dashed arrows. Hydrophobic amino acids are shown in the green circle (indicated by green dashed arrows), and Polar amino acids are in purple

moiety have formed two different hydrogen bonds with Phe518 and His90, respectively (Figs. 5d and 6d), blocking the Hydrophobic Aromatic pocket. The pyrrole moiety of the Ligand has formed another hydrogen bond with Tyr355 and blocks the catalytic binding site [83, 84], which forms the penta-dienyl radical centre on C11 of Arachidonic Acid [85]. Ligand 309,156 has formed three hydrogen bonds with three different amino acids (Figs. 5e and 6e), among which the two oxygen atoms of Sulfonamide moiety have included two different hydrogen bonds with Phe518 and His90 by entering the top channel of the active site [86]. The carbonyl oxygen of 4-Pyrone moiety has formed an H-Bond with Ser530 and blocks the helix 17 pockets of COX-2, which leads to the inhibitions of cyclopentane cyclisation step of Prostaglandin G_2 from the Arachidonic Acid [32]. The carbonyl carbon of Ligand 339,222 has formed a strong H-bond with Tyr348 (Figs. 5f and 6f), which is near to Tyr385 and Trp387, helps to destabilise the protein conformation of COX-2 and inhibits the synthesis of Prostaglandin G_2 to Prostaglandin H_2 [84]. After analysing the binding orientation of all the best-scoring molecules in the Cyclooxygenase active site of COX-2, it can be concluded that our selected molecules can be suitable competitive inhibitors of Arachidonic Acid due to the better binding interaction with the receptor than the co-crystal molecule.

ADME parameters prediction

The in silico ADME parameters of the newly selected compounds were calculated using the SwissADME (http://www.swissadme.ch) server. Lipinski's rule of five is essential to ensure a drug-like pharmacokinetic profile (Table 3) when using rational drug design [57, 87].

Bioavailability Radar is provided for a quick assessment of drug-likeness. Lipophilicity [57](Log P:-0.7 and +5.0), size (Mol. Wt. between 150 and 500 g/mol), polarity (TPSA between 20 and 130 Å2), solubility (log S not higher than 6), flexibility (no more than 9 rotatable bonds) and saturation

Table 3 In silico predicted ADME properties of the selected compound

Molecule	Mol wt. (gm/mol)	HBA	HBD	MR	TPSA (Å^2)	MLog P (O/W)	Solubility (Mg/ml)	Lipinski rule
69,030	361.80	5	0	91.75	85.62	1.34	4.04e-02	Yes
71,423	374.84	4	0	98.92	72.73	2.63	4.53e-03	Yes
126,967	567.16	3	1	110.15	58.31	4.46	1.35e-04	2 Violation (M.W and Log P)
309,156	378.80	5	0	93.91	72.73	2.79	6.40e-03	Yes
339,222	494.37	5	0	118.26	73.75	4.46	1.78e-04	1 Violation (MLog P)
AA (Co-crystal ligand)	303.46	2	0	96.19	40.13	-	-	-

(fraction of carbons in the sp3 hybridisation not less than 0.25) are the six qualities considered as physicochemical properties [57, 58, 88–91]. Descriptors defined a physicochemical range on each axis, shown as a pink region (Fig. 7) in which the molecule's radar plot had to fall totally to be declared drug-like.

Molecular dynamic simulation analysis

Although molecular docking investigations have revealed insights regarding complicated binding modes (protein-inhibitor), modelling molecular dynamics can detect the tiniest disagreement [53, 61, 92, 93]. The selected ligands were investigated for possible atomic features inside the solvent system for 150 ns using the best interactions and energy-docked conformation derived from docking data to analyse the conformational stability of ICVU-ligand complexes and ICVU–co-crystal ligand complex in the physiological parameters, such as temperature, pressure, and were embedded with water molecules. The MD simulations of 1CVU–ligand complexes analysis were compared with that of the co-crystal complex for 150 ns.

The RMSD was derived after MD simulations to evaluate changes in individual atom states using their initial state as the reference. This means that the initial docked posture of the ligands at the target protein's binding site will be used as the reference frame, and the mobility of this original frame during the MD simulation will be analysed in terms of time. Figure 8 shows the RMSD values for the protein on the Y-axis (left side). An RMSD value

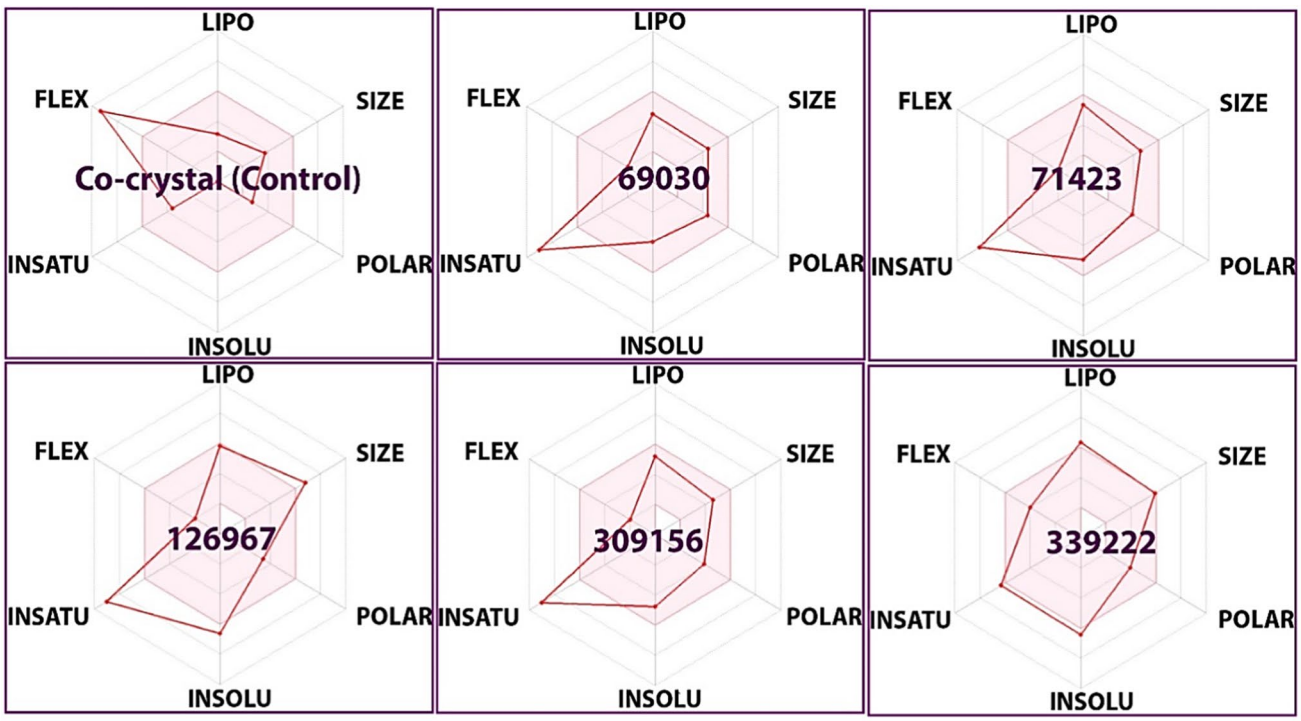


Fig. 7 Bioavailability radar of all the compounds. The pink colour region represents the optimal range of each physicochemical value

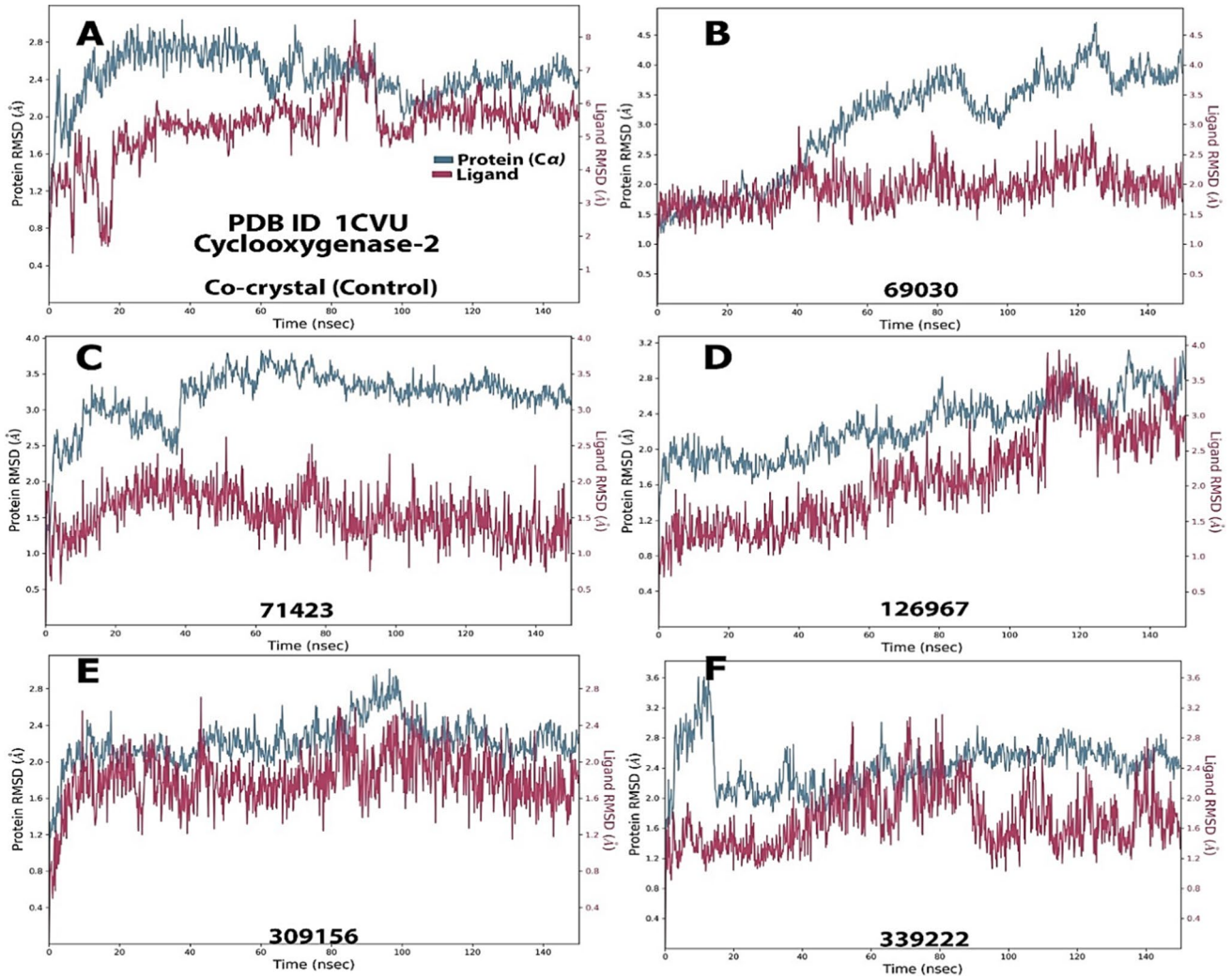


Fig. 8 The root means square deviation (RMSD) of ligand interaction during 150 ns MD trajectory for **A** COX 2–co-crystal ligand complex **B** COX 2–69,030 ligand complex **C** COX 2–71,423 ligand complex **D** COX 2–126,967 ligand complex **E** COX 2–309,156 ligand complex **F** COX 2–339,222 ligand complex

of 1–4 Å is acceptable for small proteins [62, 63]. The range of this RMSD value can increase as a protein's size increases [14]. Throughout the simulation, the RMSD of the protein backbone of the COX 2-co-crystal ligand complex (Fig. 8A) remained below 2.4 Å. The RMSD value of the COX-2–69,030 complex protein backbone (Fig. 8B) was determined to be about 4.0 Å. The RMSD of the protein backbone of the COX 2-71,423 complex (Fig. 8C) is approximately 3.2 Å. The RMSD of the protein backbone of the COX 2–126,967 complex (Fig. 8D) was determined to be less than 3.0 Å throughout the simulation. The RMSD of the protein backbone of the COX 2–309,156 ligand complex (Fig. 8E) is about 2.2 Å. Throughout the simulation, the RMSD of the protein backbone of the COX 2–339,222 ligand complex (Fig. 8F) remained less than 2.4 Å.

The protein's RMSF "root mean square Fluctuation" produced a stable structure during the simulation, providing an appropriate basis for subsequent investigation. The areas of the protein that differ the most from the reference can be observed from the RMSF plot [38, 60]. The RMSF plot shows the portions of the protein that differ the most from the reference. Figure 9 depicts the graphical RMSF data for the protein–ligand combination obtained by the MD simulation. Figure 9 clearly illustrates that the RMSF value for protein backbone amino acids fluctuates significantly in the N- and C-terminal regions compared to other protein regions. All of the protein–ligand complexes (Fig. 9B–F) should have similar RMSF values to the COX 2-co-crystal ligand complex (Fig. 9A) [94].

Most of the RMSF of COX 2–co-crystal ligand complex (Fig. 9A) fluctuates between 0.6 and 2.4 Å. The majority of

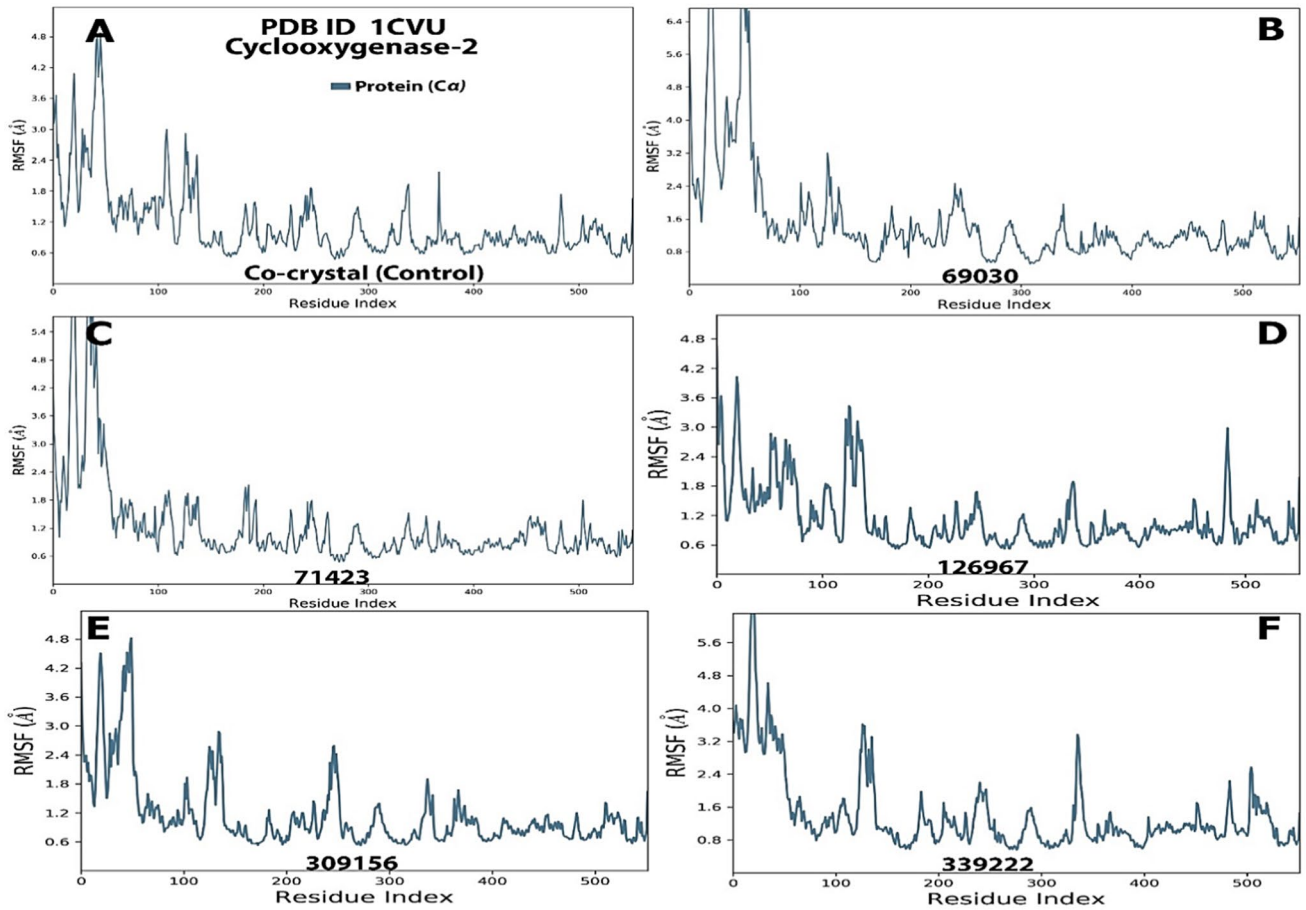


Fig. 9 The root means square fluctuation (RMSF) of 1CVU protein during 150 ns MD, representing local changes along the protein chain for **A** COX 2–co-crystal ligand complex **B** COX 2–69,030 ligand complex **C** COX 2–71,423 ligand complex **D** COX 2–126,967 ligand complex **E** COX 2–309,156 ligand complex **F** COX 2–339,222 ligand complex

the RMSF of COX 2–69,030 ligand complex (Fig. 9B) was found to differ within the range of 0.8 and 2.5 Å. Most of the RMSF of COX 2–71,423 ligand (Fig. 9C) varies between 0.6 and 2.0 Å. The majority of the RMSF of COX 2–126,967 ligand (Fig. 9D) oscillates within the range of 0.6 and 2.0 Å. Most COX 2–309,156 ligand (Fig. 9E) differs between 0.6 and 2.5 Å. The more significant part of the RMSF of the COX 2–339,222 ligand (Fig. 9F) was found to vary within the range of 0.8 and 2.5 Å. During the MD simulations, the RMSF of the protein–ligand interactions for all phytocompounds was consistent with the RMSF of the protein-co-crystal ligand complexes.

The properties of the ligands, such as the RMSD, radius of gyration (rGyr), molecular surface area (MolSA), solvent accessible surface area (SASA) and polar surface area (PSA), were studied during 150 ns MD simulations. The graphical results for the co-crystal ligand and the selected compounds (69,030, 71,423, 126,967, 309,156, 339,222) are illustrated in Fig. 10. RGyr reveals the compactness of a structure, where huge fluctuations indicate less stability. MolSA relates to the geometric surface property wherein the value of MolSA and van der Walls surface area are equivalent. SASA refers to the surface of a molecule in direct contact with the solvent system where it is present. PSA refers to the total surface area of polar atoms such as oxygen, nitrogen and corresponding hydrogens [60, 94–96].

Figure 10A depicts the various properties of co-crystal ligands. The RMSD of the co-crystal ligand can be in between the range of 1.5–4.5 Å, and it reaches the equilibrium around 3.5 Å. Most of the fluctuations of rGyrare were found to be between the range of 4.0–7.2 Å and reach the equilibrium at about 5.5 Å. The majority of MolSA fluctuations are between the range of 350–420 $Å^2$, and the equilibrium is at around 400 $Å^2$. During the 150 ns MD simulations, SASA oscillates between 00 and 15 $Å^2$ from 40 to 120 ns, and from that point on, the fluctuations are found to reach the equilibrium at 02 $Å^2$. The PSA fluctuates between 50 $Å^2$ and 105 $Å^2$, and the equilibrium is around 100 $Å^2$. The characteristics of compound 69,030 are depicted in Fig. 10B. The RMSD of 69,030 is between

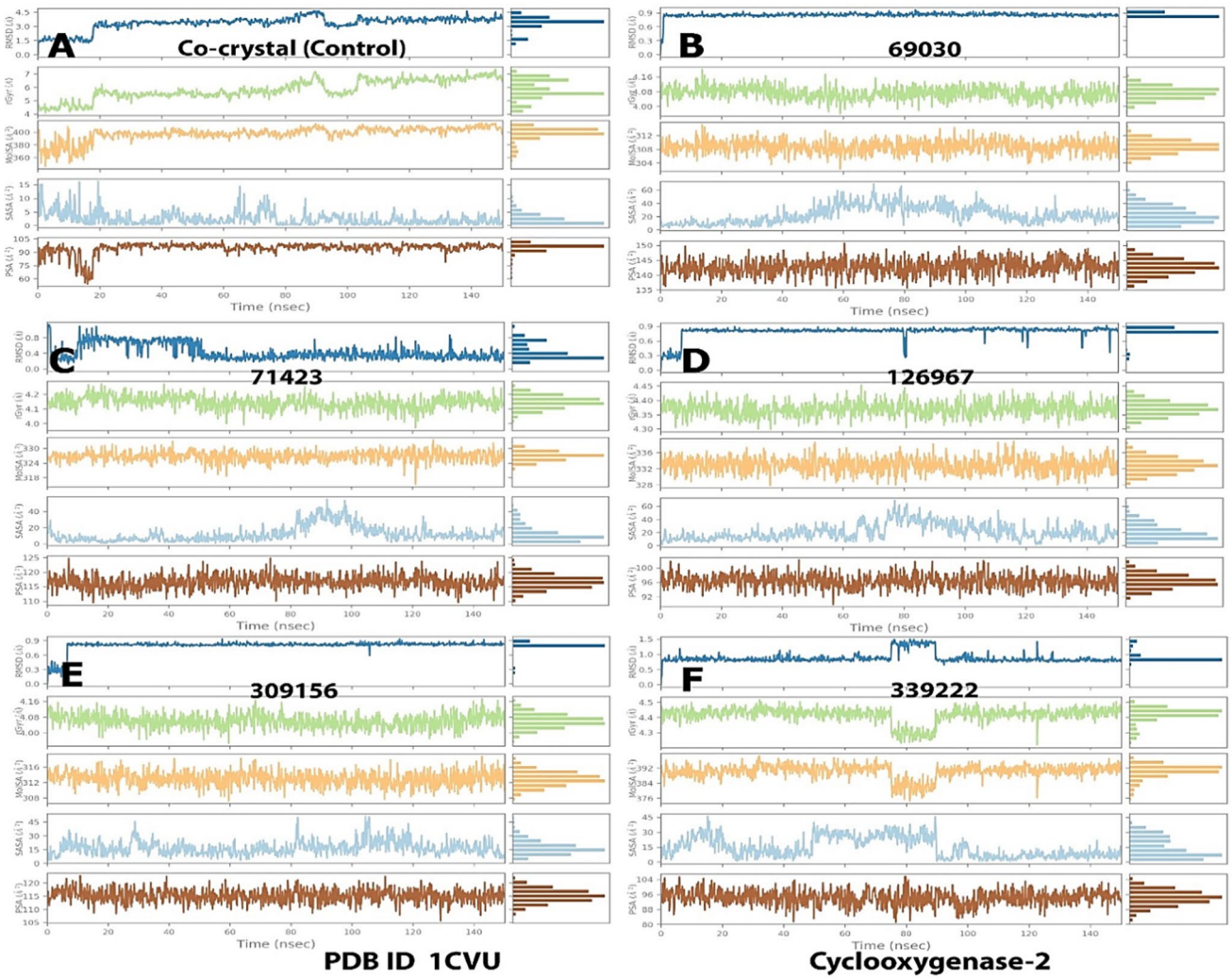


Fig. 10 The trajectory of different properties of **A** co-crystal ligand **B** 69030 ligand **C** 71423 ligand **D** 126967 ligand **E** 309156 ligand **F** 339222 ligand during the 150 ns of MD simulations

0.3 and 0.9 Å, with an equilibrium value of roughly 0.8 Å. The rGyr ranges between 3.9 and 4.2 Å, with an equilibrium at approximately 4.10 Å. The MolSA ranges between 300 and 315 Å^2, with an equilibrium of about 309 Å^2. During the 150 ns MD simulations, SASA swings between 00 and 70 Å^2 from 40 to 120 ns, and the fluctuations gradually attain equilibrium at 20 Å^2. The PSA ranges between 135 and 150 Å^2, with an equilibrium value of approximately 143Å^2. The characteristics of compound 71,423 are depicted in Fig. 10C. The RMSD of 71,423 is between 0.2 and 1.0 Å, with an equilibrium value of roughly 0.3 Å. The rGyr swings between 4.30 and 4.43 Å, with an equilibrium at approximately 4.37 Å. The MolSA ranges between 315 and 335 Å^2, with an equilibrium of about 333 Å^2. During the 150 ns MD simulations, SASA fluctuates most between 0 and 60 Å^2 from 80 to 100 ns, and the fluctuations gradually achieve equilibrium at 10 Å^2. The PSA varies between 110 and 125 Å^2, with an equilibrium value of approximately 116 Å^2.

The characteristics of compound 126,967 are shown in Fig. 10D. The RMSD of compound 126,967 is between 0.3 and 0.9 Å, with an equilibrium value of roughly 0.8 Å. The rGyr ranges between 4.0 and 4.3 Å, with an equilibrium at approximately 4.15 Å. The MolSA ranges between 328 and 340 Å^2, with the equilibrium at around 327 Å^2. During the 150 ns MD simulations, SASA fluctuates most between 0 and 60 Å^2 from 80 to 100 ns, and the fluctuations gradually achieve equilibrium at 10 Å^2. The PSA varies from 90 to 103 Å^2, with an equilibrium value of approximately 95 Å^2. The characteristics of compound 309,156 are depicted in Fig. 10E. The RMSD of compound 309,156 is between 0.2 and 0.9 Å, with an equilibrium value of roughly 0.8 Å. The rGyr ranges between 3.95 and 4.17 Å, with an equilibrium at approximately 4.06 Å. The MolSA ranges between 308

and 319 Å^2, with an equilibrium of about 312 Å^2. During the 150 ns MD simulations, SASA swings most between 5 and 50 Å^2 from 80 to 100 ns, and the fluctuations then gradually attain equilibrium at 15 Å^2. The PSA varies between 108 and 123 Å^2, with an equilibrium value of approximately 115 Å^2. The characteristics of compound 339,222 are shown in Fig. 10F. The RMSD of compound 339,222 is between 0.5 and 1.5 Å, with an equilibrium value of roughly 0.8 Å. The rGyr swings between 4.2 and 4.5 Å, with an approximately 4.4 Å 5 equilibrium. The MolSA varies between 376 and 400 Å^2, with the equilibrium at 392 Å^2. During the 150 ns MD simulations, SASA fluctuates most between 0 and 45 Å^2 from 80 to 100 ns, and the fluctuations gradually achieve equilibrium at 10 Å^2. The PSA varies between 81 and 105 Å^2, with an equilibrium value of approximately 95Å^2.

The simulation interaction diagrams shown in Fig. 11 investigate the four protein–ligand interactions that occurred during the 150 ns MD simulations study: hydrogen bonds, hydrophobic interactions, ionic interactions, and water bridge interactions. The binding interactions of COX 2–co-crystal ligand complex, COX 2–69,030 ligand complex, COX 2–71,423 ligand complex, COX 2–126,967 ligand complex, COX 2–309,156 ligand complex, COX 2–339,222 ligand complex. After the MD simulation, protein interaction with the co-crystal ligand is validated by Fig. 11A, where the co-crystal ligand is found to interact with Met522 and Ser530 and shown to provide stable conformation of the native ligand at the active site throughout the simulated period. The MD simulation protein interaction with the 69,030 ligands is validated by Fig. 11B, where the ligand interacts with Phe518, Met522, Ser530 and Arg513 and is shown to provide stable conformation throughout the 150 ns simulation. The ligand 71,423 has been found to form hydrophobic interactions (Fig. 11C) with several amino acids of the hydrophobic aromatic pocket, and the interaction with the terminal amino group Arg120 remains intact throughout the 150 ns period. The simulation results of ligand 126,967 are shown in Fig. 11D, where it can be observed that apart from the initial interaction with His90, Phe518 and Tyr355, two significant hydrogen bonding interactions formed with Ile517 and Phe518. The ligand 309,156 interacts with His90, Phe518 and Ser530 via hydrogen bonding, hydrophobic

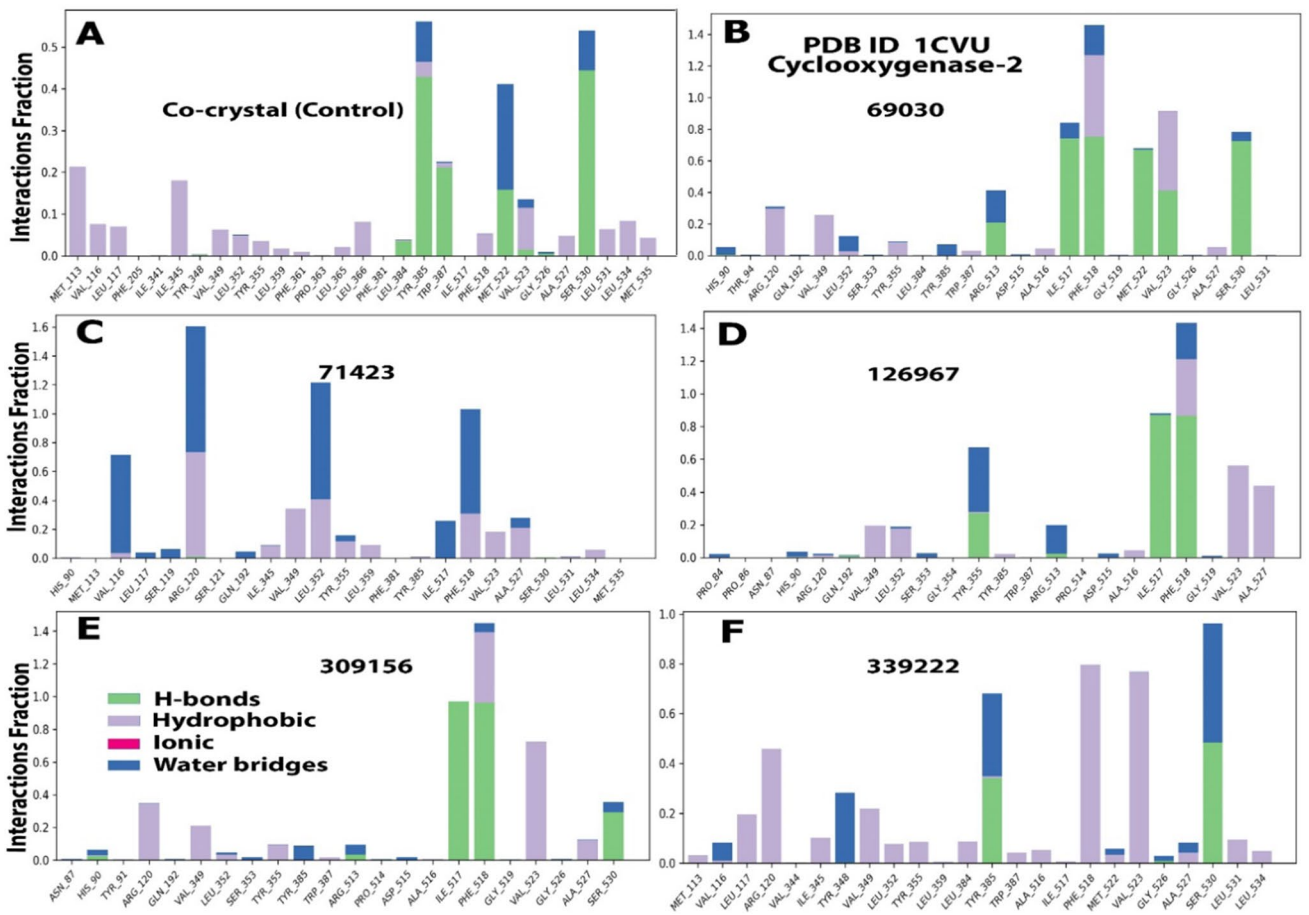


Fig. 11 The plot represents the hydrogen bonding interactions of **A** co-crystal ligand **B** 69030 ligand **C** 71423 ligand **D** 126967 ligand **E** 309156 ligand **F** 339222 ligand with the amino acid residues of COX-2 protein during 150 ns MD simulation

and water bridge interaction, validated by Fig. 11E. The MD simulation results of ligand 339,222 are validated by Fig. 11F, wherein it can be observed that apart from the initial interaction with Tyr348, it has also formed hydrogen bonds with Tyr385 and Ser530 utilising a water bridge. So, it can be concluded that the residues mentioned above played a significant role at the active binding site and have been vital for stabilising the co-crystal molecule and the selected molecules inside the cyclooxygenase active site of COX-2 [32, 76, 97].

Figure 12 depicts a schematic representation of the protein–ligand interactions (hydrogen bonds, hydrophobic interactions, ionic interactions and water bridge interactions) during MD simulations. The data provided in Fig. 12 also show the total number of interactions formed by the amino acids of COX-2–s cyclooxygenase active site with the co-crystal ligand during the MD simulations. Figure 12 shows how frequently COX-2 amino acids interact with co-crystal ligands, compound 69,030, compound 71,423, compound 126,967, compound 309,156 and compound 339,222. The scale on the right side of each diagram shows that a deeper shade of orange corresponds to numerous interactions of amino acids with the compounds [14]. The protein–ligand interaction plot and schematic illustration (Figs. 11 and 12) support the MD simulation study results because the amino acid residues of COX-2 were found to interact with the co-crystal ligand and test compounds during docking were also found to interact with the amino acid residues of cyclooxygenase active site of COX-2 during the 150 ns MD simulations.

Furthermore, numerous amino acids that were not predicted to interact with the chosen compounds during docking were discovered to interact with them at a certain period during the 150 ns simulations. This suggests a vitalprotein–ligand interaction profile and entirely occupied the cyclooxygenase active site of COX-2. From Figs. 11 and 12, it can be concluded that at a particular time, almost all the selected molecules interact with the amino acid residue range of Tyr385-Ser530 in the same way as the co-crystal molecule does in the cyclooxygenase active site of COX-2

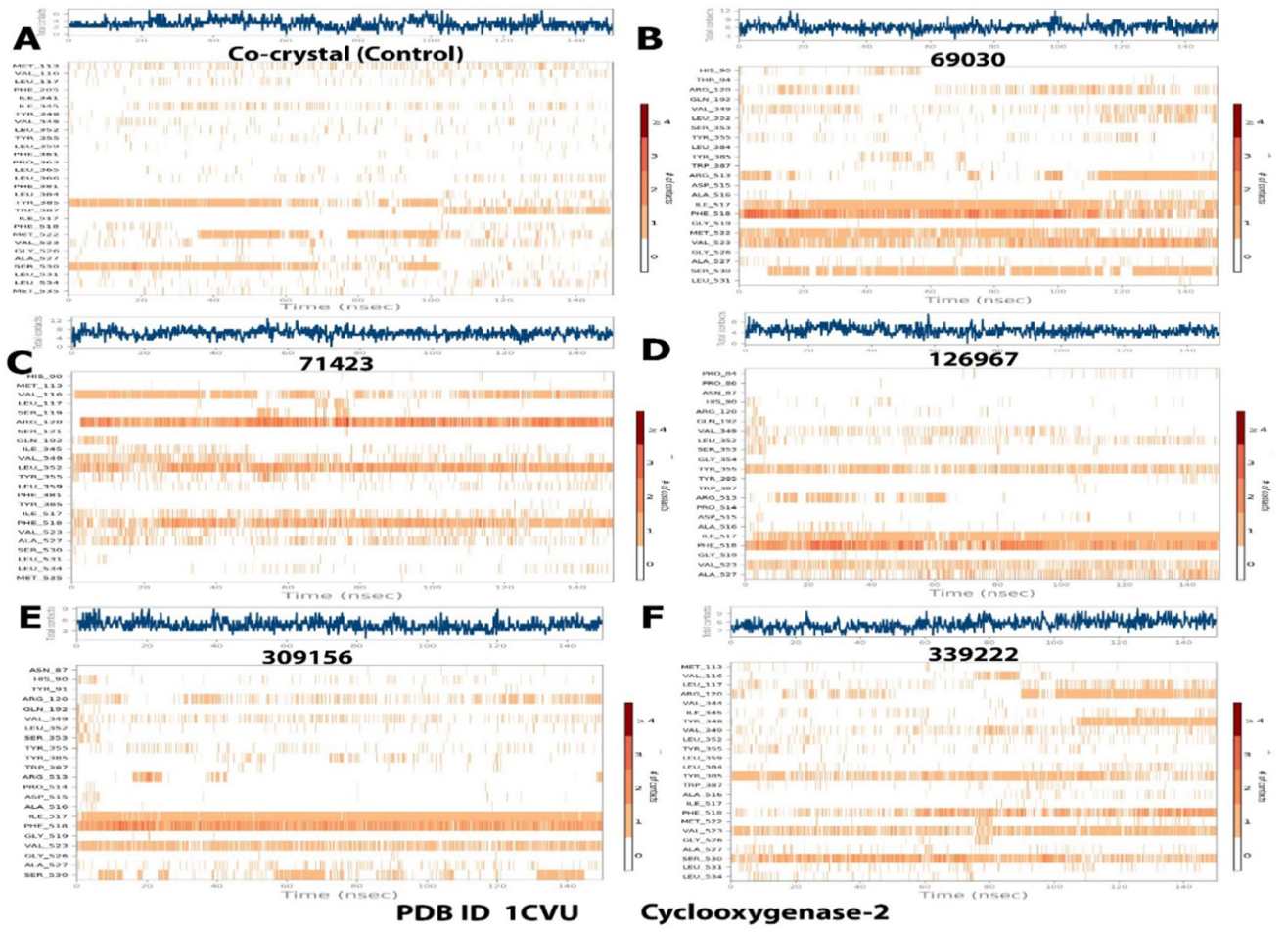


Fig. 12 Timeline interaction data of the amino acids of the protein with **A** co-crystal ligand **B** 69030 ligand **C** 71423 ligand **D** 126967 ligand **E** 309156 ligand **F** 339222 ligand throughout the entire 150 ns MD simulations

Table 4 Binding free energy values were found in the MM-GBSA study

Serial number	Ligand	MM-GBSA dG Bind (–kcal/mol)
1	Co-crystal	–72.0758401106395
2	69,030	–74.0993193920694
3	71,423	–64.6818324606174
4	126,967	–68.2292497261369
5	309,156	–70.640260057653
6	339,222	–86.0914082292365

and completely occupied the hydrophobic pocket and catalytical binding site [32, 75–77].

Binding free energy calculation (MM-GBSA)

To investigate the ligand-receptor binding energies, the five compounds were subjected to a molecular mechanics generalised born solvent accessibility (MM-GBSA) analysis. The MM-GBSA technique determines the binding free energy of the ligand–receptor complex to ensure the compounds' stability after binding to the target protein's active binding site [14, 92].

MM-GBSA dG bind values are the energy difference between the optimised ligand–receptor complex (prime) energy and a combination of optimised free ligand and receptor energy [92]. The MM-GBSA dG bind values are given in Table 4. From the MM-GBSA analysis, it can be observed that the ligand 339,222 has been found to have best binding energy (– 86.0914082292365 kcal/mol) followed by the ligand 69,030 (– 74.0993193920694 kcal/mol), ligand 309,156 (– 70.640260057653 kcal/mol), ligand 126,967 (– 68.2292497261369 kcal/mol) and ligand 71,423 (– 64.6818324606174 kcal/mol). The MM-GBSA dG bind the value of the Co-crystal molecule is found to be – 72.0758401106395 Kcl/mol. So, based on MMGBSA analysis, ligand 339,222 is found to be more stable than the co-crystal ligand and can be considered to be the toughest among the other five ligands after it binds to the cyclooxygenase active site of COX-2 and the MM-GBSA result correlates well with the results given in Fig. 9F and 10F.

Conclusion

This study used explainable artificial intelligence (XAI) models and bioinformatics tools to design and discover phenolic COX-2 inhibitors. XAI models were developed to predict the bioactivity of COX-2 inhibitors and used to identify potential drug candidates by scanning CHEMBL, an extensive database of natural products. Moreover, molecular fingerprint analysis of the scanned compounds was performed and the common structural properties of phenolic compounds that are effective as COX-2 inhibitors were analysed using bioinformatics tools. This study demonstrates the potential of XAI models and bioinformatics tools in natural product research.

Author contributions MR and KKK conceptualized the idea and contributed to the design of methodology. MR, KKK, MA and SM wrote and edited the paper. MR and KKK performed the explainable AI, methodology and algorithm parts. MA and SM performed the data collection, analysis and interpretation parts. All authors reviewed and approved the paper.

Declarations

Competing interest The authors declare no competing interests.

References

1. Ju Z, Li M, Xu J, Howell DC, Li Z, Chen FE (2022) Recent development on COX-2 inhibitors as promising anti-inflammatory agents: the past 10 years. Acta Pharm Sin B 12(6):2790–2807. https://doi.org/10.1016/j.apsb.2022.01.002
2. Desai SJ, Prickril B, Rasooly A (2018) Mechanisms of phytonutrient modulation of cyclooxygenase-2 (COX-2) and inflammation related to cancer. Nutr Cancer 70(3):350–375. https://doi.org/10.1080/01635581.2018.1446091
3. Zarghi A, Arfaei S (2011) Selective COX-2 inhibitors: a review of their structure—activity relationships. Iran J Pharm Res 10(4):655–683
4. Ricciotti E, FitzGerald GA (2011) Prostaglandins and inflammation. Arterioscler Thromb Vasc Biol 31(5):986–1000. https://doi.org/10.1161/atvbaha.110.207449
5. Das UN (2005) Can COX-2 inhibitor-induced increase in cardiovascular disease risk be modified by essential fatty acids? J Assoc Physicians India 53:623–627
6. Prusakiewicz JJ, Duggan KC, Rouzer CA, Marnett LJ (2009) Differential sensitivity and mechanism of inhibition of COX-2 oxygenation of arachidonic acid and 2-arachidonoylglycerol by ibuprofen and mefenamic acid. Biochemistry 48(31):7353–7355. https://doi.org/10.1021/bi900999z
7. Brock TG, McNish RW, Peters-Golden M (1999) Arachidonic acid is preferentially metabolized by cyclooxygenase-2 to prostacyclin and prostaglandin E2. J Biol Chem 274(17):11660–11666. https://doi.org/10.1074/jbc.274.17.11660
8. Ambriz-Pérez DL, Leyva-López N, Gutierrez-Grijalva EP, Heredia JB (2016) ?Phenolic compounds: natural alternative in inflammation treatment: a review. Cogent Food Agric 2(1):1131412. https://doi.org/10.1080/23311932.2015.1131412
9. Rahman MM et al (2021) Role of phenolic compounds in human disease: current knowledge and future prospects. Molecules 27(1):233. https://doi.org/10.3390/molecules27010233
10. Hu F et al (2023) Phenolic compounds from Chaenomeles speciosa alleviate inflammation in lipopolysaccharide-treated RAW264.7 macrophages via the NF-κB and MAPK pathways. Food Sci Human Wellness 12(4):1071–1080. https://doi.org/10.1016/j.fshw.2022.10.025
11. Abdelgawad MA et al (2021) Novel phenolic compounds as potential dual EGFR and COX-2 inhibitors: design, semisynthesis,

in vitro biological evaluation and in silico insights. Drug Des Devel Ther 15:2325–2337. https://doi.org/10.2147/dddt.S310820
12. Ali DE, Gedaily RAE, Ezzat SM, Sawy MAE, Meselhy MR, Abdel-Sattar E (2023) In silico and in vitro anti-inflammatory study of phenolic compounds isolated from Eucalyptus maculata resin. Sci Rep 13(1):2093. https://doi.org/10.1038/s41598-023-28221-y
13 Gaulton A et al (2017) The ChEMBL database in 2017. Nucl. Acids Res. 45(D1):D945–D954. https://doi.org/10.1093/nar/gkw1074
14 Zothantluanga J, Abdalla M, Rudrapal M, Tian Q, Chetia DD, Li J (2022) Computational investigations for identification of bioactive molecules from Baccaurea ramiflora and Bergenia ciliata as inhibitors of SARS-CoV-2 M pro. Polycyclic Aromatic Compd 43:2459–2487. https://doi.org/10.1080/10406638.2022.2046613
15. Pérez-Sacau E et al (2007) Synthesis and pharmacophore modeling of naphthoquinone derivatives with cytotoxic activity in human promyelocytic leukemia HL-60 cell line. J Med Chem 50(4):696–706. https://doi.org/10.1021/jm060849b
16. Prachayasittikul V et al (2014) Synthesis, anticancer activity and QSAR study of 1,4-naphthoquinone derivatives. Eur J Med Chem 84:247–263. https://doi.org/10.1016/j.ejmech.2014.07.024
17. Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput. Chem. 32(7):1466–1474. https://doi.org/10.1002/jcc.21707
18. Carracedo P et al (2021) A review on machine learning approaches and trends in drug discovery. Comput Struct Biotechnol J 19:4538–4558. https://doi.org/10.1016/j.csbj.2021.08.011
19. Gramatica P, Sangion A (2016) A historical excursus on the statistical validation parameters for QSAR models: a clarification concerning metrics and terminology. J Chem Inf Model 56(6):1127–1131. https://doi.org/10.1021/acs.jcim.6b00088
20. Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/a:1010933404324
21. Mi X, Zou B, Zou F, Hu J (2021) Permutation-based identification of important biomarkers for complex diseases via machine learning models. Nat Commun 12(1):3008. https://doi.org/10.1038/s41467-021-22756-2
22. Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140. https://doi.org/10.1007/bf00058655
23. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D (2019) A survey of methods for explaining black box models. ACM Comput Surv 51(5):1–42. https://doi.org/10.1145/3236009
24. Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. Presented at the proceedings of the 31st international conference on neural information processing systems. Long Beach
25. Hu L, Liu B, Ji J, Li Y (2020) Tree-based machine learning to identify and understand major determinants for stroke at the neighborhood level. J Am Heart Assoc. https://doi.org/10.1161/jaha.120.016745
26. Lundberg SM et al (2018) Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2(10):749–760
27 Shapley LS (1953) 17. A value for n-person games. In: Harold William K, Albert William T (eds) Contributions to the theory of games (AM-28), vol II. Princeton University Press, Princeton, pp 307–318
28. Molnar C (2020) Interpretable machine learning. Self published. https://christophm.github.io/interpretable-ml-book/
29. Mitchell R, Frank E, Holmes G (2022) ?GPUTreeShap: massively parallel exact calculation of SHAP scores for tree ensembles. PeerJ Comput Sci 8:e880. https://doi.org/10.7717/peerj-cs.880
30. Moncada-Torres A, Van Maaren MC, Hendriks MP, Siesling S, Geleijnse G (2021) Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. Sci Rep. https://doi.org/10.1038/s41598-021-86327-7
31 Kirboga KK, Abbasi S, Küçüksille EU (2023) Explainability and white box in drug discovery. Chem Biol Drug Design. https://doi.org/10.1111/cbdd.14262
32. Kiefer JR et al (2000) ?Structural insights into the stereochemistry of the cyclooxygenase reaction. Nature 405(6782):97–101. https://doi.org/10.1038/35011103
33 Allen AEA, Robertson MJ, Payne MC, Cole DJ (2019) Development and validation of the quantum mechanical bespoke protein force field. ACS Omega 4(11):14537–14550. https://doi.org/10.1021/acsomega.9b01769
34. Ortuso F, Langer T, Alcaro S (2006) GBPM: GRID-based pharmacophore model: concept and application studies to protein?protein recognition. Bioinformatics 22(12):1449–1455. https://doi.org/10.1093/bioinformatics/btl115
35 Elokely KM, Doerksen RJ (2013) Docking challenge: protein sampling and molecular docking performance. J Chem Inf Model 53(8):1934–1945. https://doi.org/10.1021/ci400040d
36. Csizmadia P (1999) MarvinSketch and MarvinView: molecule applets for the World Wide Web
37 MadhaviSastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aid Mol Design 27(3):221–234. https://doi.org/10.1007/s10822-013-9644-8
38 Rudrapal M et al (2023) ?Dual synergistic inhibition of COX and LOX by potential chemicals from Indian daily spices investigated through detailed computational studies. Sci Rep 13(1):8656. https://doi.org/10.1038/s41598-023-35161-0
39. Abdolmaleki A, Ghasemi JB, Ghasemi F (2017) Computer aided drug design for multi-target drug design: SAR /QSAR, molecular docking and pharmacophore methods. Curr Drug Targets 18(5):556–575. https://doi.org/10.2174/1389450117666160101120822
40. Badavath VN, Sinha B, Jayaprakash V (2015) Design, in-silico docking and predictive ADME properties of novel Pyrazoline derivatives with selective hMAO inhibitory activity. Int J Pharm Pharm Sci 7:56–61
41. Floquet N, Richez C, Durand P, Maigret B, Badet B, Badet-Denisot MA (2007) ?Discovering new inhibitors of bacterial glucosamine-6P synthase (GlmS) by docking simulations. Bioorg Med Chem Lett 17(7):1966–1970. https://doi.org/10.1016/j.bmcl.2007.01.052
42 Gentile F et al (2020) Deep docking: a deep learning platform for augmentation of structure based drug discovery. ACS Central Sci 6(6):939–949. https://doi.org/10.1021/acscentsci.0c00229
43. Ghosh S, Chetia D, Gogoi N, Rudrapal M (2021) ?Design, molecular docking, drug-likeness, and molecular dynamics studies of 1,2,4-trioxane derivatives as novel Plasmodium falciparum falcipain-2 (FP-2) inhibitors. BioTechnologia (Pozn) 102(3):257–275. https://doi.org/10.5114/bta.2021.108722
44. Jayanna DND, Vagdevi H, Dharshan J, Ramappa R, Telkar S (2013) Synthesis, antimicrobial, analgesic activity, and molecular docking studies of novel 1-(5, 7-dichloro-1, 3-benzoxazol-2-yl)-3-phenyl-1H-pyrazole-4-carbaldehyde derivatives. Med Chem Res 22:1–9. https://doi.org/10.1007/s00044-013-0565-9
45. Joshi SD et al (2018) ?Pharmacophore mapping, molecular docking, chemical synthesis of some novel pyrrolyl benzamide derivatives and evaluation of their inhibitory activity against enoyl-ACP reductase (InhA) and mycobacterium tuberculosis. Bioorg Chem 81:440–453. https://doi.org/10.1016/j.bioorg.2018.08.035
46. Krishnaswamy G, Desai NR, Naika R, Naika H, Mahadevan KM, Satyendra RV, Kumar DBA (2016) Design, synthesis, antibacterial, antioxidant activity and molecular docking studies of 6-hydroxybenzofuran derivatives. J Chem Pharm Res 8(4):19–28

47. Li M et al (2019) ?Discovery of Keap1-Nrf2 small-molecule inhibitors from phytochemicals based on molecular docking. Food Chem Toxicol 133:110758. https://doi.org/10.1016/j.fct.2019.110758
48. Lohidakshan K, Rajan M, Ganesh A, Paul M, Jerin J (2018) Pass and Swiss ADME collaborated in silico docking approach to the synthesis of certain pyrazoline spacer compounds for dihydrofolate reductase inhibition and antimalarial activity. Bangladesh J Pharmacol 13(1):23. https://doi.org/10.3329/bjp.v13i1.33625
49 McGaughey GB et al (2007) ?Comparison of topological, shape, and docking methods in virtual screening. J Chem Inf Model 47(4):1504–1519. https://doi.org/10.1021/ci700052x
50. Morris GM et al (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19(14):1639–1662
51. Reddy AS, Mao J, Krishna LS, Badavath VN, Maji S (2019) Synthesis, spectral investigation, molecular docking and biological evaluation of Cu(II), Ni(II) and Mn(II) complexes of (E)-2-((2-butyl-4-chloro-1H-imidazol-5-yl)methylene)-N-methylhydrazinecarbothioamide (C10H16N5ClS) and its DFT studies. J Mol Struct 1196:338–347. https://doi.org/10.1016/j.molstruc.2019.06.085
52. Rudrapal M et al (2022) Phytocompounds as potential inhibitors of SARS-CoV-2 Mpro and PLpro through computational studies. Saudi J Biol Sci 29(5):3456–3465. https://doi.org/10.1016/j.sjbs.2022.02.028
53 Rudrapal M et al (2022) Repurposing of phytomedicine-derived bioactive compounds with promising anti-SARS-CoV-2 potential: molecular docking, MD simulation and drug-likeness/ADMET studies. Saudi J Biol Sci 29(4):2432–2446. https://doi.org/10.1016/j.sjbs.2021.12.018
54. Tomi IHR, Al-Daraji AHR, Abdula AM, Al-Marjani MF (2016) Synthesis, antimicrobial and docking study of three novel 2,4,5-triarylimidazole derivatives. J Saudi Chem Soc 20:S509–S516. https://doi.org/10.1016/j.jscs.2013.03.004
55. Vijesh AM, Isloor AM, Telkar S, Arulmoli T, Fun H-K (2013) Molecular docking studies of some new imidazole derivatives for antimicrobial properties. Arab J Chem 6(2):197–204. https://doi.org/10.1016/j.arabjc.2011.10.007
56. Wan Y, Tian Y, Wang W, Gu S, Ju X, Liu G (2018) In silico studies of diarylpyridine derivatives as novel HIV-1 NNRTIs using docking-based 3D-QSAR, molecular dynamics, and pharmacophore modeling approaches. RSC Adv 8(71):40529–40543. https://doi.org/10.1039/C8RA06475J
57 Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 7(1):42717. https://doi.org/10.1038/srep42717
58. Badavath VN et al (2022) Determination of potential inhibitors based on isatin derivatives against SARS-CoV-2 main protease (m(pro)): a molecular docking, molecular dynamics and structure-activity relationship studies. J Biomol Struct Dyn 40(7):3110–3128. https://doi.org/10.1080/07391102.2020.1845800
59. Lipinski C, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46(3):00129–0
60. Abdalla M, Eltayb WA, El-Arabey AA, Singh K, Jiang X (2022) Molecular dynamic study of SARS-CoV-2 with various S protein mutations and their effect on thermodynamic properties. Comput Biol Med 141:105025. https://doi.org/10.1016/j.compbiomed.2021.105025
61. Bowers K et al. (2006) Molecular dynamics—scalable algorithms for molecular dynamics simulations on commodity clusters, p. 84.
62 Ezugwu JA et al (2022) Design, synthesis, molecular docking, molecular dynamics and in vivo antimalarial activity of new dipeptide-sulfonamides. ChemistrySelect 7(5):e202103908. https://doi.org/10.1002/slct.202103908
63. Ghosh P et al (2021) Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques. IEEE Access 9:19304–19326. https://doi.org/10.1109/ACCESS.2021.3053759
64. Nimgampalle M, Devanathan V, Saxena A (2020) Screening of chloroquine, hydroxychloroquine and its derivatives for their binding affinity to multiple SARS-CoV-2 protein drug targets. J Biomol Struct Dyn 39:1–13. https://doi.org/10.1080/07391102.2020.1782265
65. Rudrapal M et al (2022) In silico screening of phytopolyphenolics for the identification of bioactive compounds as novel protease inhibitors effective against SARS-CoV-2. J Biomol Struct Dyn 40(20):10437–10453. https://doi.org/10.1080/07391102.2021.1944909
66 Jin Z et al (2020) Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. Nature 582(7811):289–293. https://doi.org/10.1038/s41586-020-2223-y
67 Li J, Abel R, Zhu K, Cao Y, Zhao S, Friesner RA (2011) The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling. Proteins 79(10):2794–2812. https://doi.org/10.1002/prot.23106
68. Jamal S, Goyal S, Shanker A, Grover A (2017) Predicting neurological adverse drug reactions based on biological, chemical and phenotypic properties of drugs using machine learning models. Sci Rep 7(1):872. https://doi.org/10.1038/s41598-017-00908-z
69. Kýrboða KK, Küçüksille EU, Köse U (2022) Ignition of small molecule inhibitors in Friedreich?s Ataxia with explainable artificial intelligence research square platform LLC, 2022-03-07
70. Shaikh MM, Patel AP, Patel SP, Chikhalia KH (2019) Synthesis, in vitro COX-1/COX-2 inhibition testing and molecular docking study of novel 1,4-benzoxazine derivatives. New J Chem 43(26):10305–10317. https://doi.org/10.1039/C9NJ00684B
71. Romero-Estrada A et al (2022) Synthesis, biological evaluation, and molecular docking study of 3-amino and 3-hydroxy-seco A derivatives of ?-amyrin and 3-epilupeol as inhibitors of COX-2 activity and NF-kB activation. J Nat Products 85(4):787–803. https://doi.org/10.1021/acs.jnatprod.1c00827
72. Hawash M et al (2023) Design, synthesis, molecular docking studies and biological evaluation of thiazole carboxamide derivatives as COX inhibitors. BMC Chem 17(1):11. https://doi.org/10.1186/s13065-023-00924-3
73. Ayuk E, Afoke P, Samuel A, Olowolafe T (2020) Synthesis and preliminary molecular docking studies of novel ethyl-glycinate amide derivatives. Int J Res Granthaalayah 8:368–382. https://doi.org/10.29121/granthaalayah.v8.i9.2020.338
74 Kalid O, Toledo Warshaviak D, Shechter S, Sherman W, Shacham S (2012) Consensus Induced Fit Docking (cIFD): methodology, validation, and application to the discovery of novel Crm1 inhibitors. J Comput Aided Mol Des 26(11):1217–1228. https://doi.org/10.1007/s10822-012-9611-9
75. Liu HX, Zhang RS, Yao XJ, Liu MC, Hu ZD, Fan BT (2004) QSAR and classification models of a novel series of COX-2 selective inhibitors: 1,5-diarylimidazoles based on support vector machines. J Comput Aided Mol Des 18(6):389–399. https://doi.org/10.1007/s10822-004-2722-1
76. Neophytou N et al (2011) Docking and molecular dynamics calculations of pyrrolidinone analog MMK16 bound to COX and LOX enzymes. Mol Inf 30:473–486. https://doi.org/10.1002/minf.201000131
77. PuttaVenkat Reddy CN, Sreenivas E, Satyanarayana S (2016) Protein modeling of COX 2 and evaluating docking for prediction of binding affinities of Ru(II)/Co(III) polypyridyl complexes with COX 2 and CDK2 proteins. J Chem Pharm Res 8(4):980–987

78. Ruslin R et al (2022) The search for cyclooxygenase-2 (COX-2) inhibitors for the treatment of inflammation disease: an in-silico study. J Multidiscip Healthc 15:783–791. https://doi.org/10.2147/jmdh.s359429
79. Singh G, Singh A, Singh P, Bhatti R (2019) Bergapten ameliorates vincristine-induced peripheral neuropathy by inhibition of inflammatory cytokines and NF?B signaling. ACS Chem Neurosci 10(6):3008–3017. https://doi.org/10.1021/acschemneuro.9b00206
80. Stiller CO, Hjemdahl P (2022) Lessons from 20 years with COX-2 inhibitors: importance of dose-response considerations and fair play in comparative trials. J Intern Med 292(4):557–574. https://doi.org/10.1111/joim.13505
81. Hamberg M (1998) Stereochemistry of oxygenation of linoleic acid catalyzed by prostaglandin-endoperoxide H synthase-2. Arch Biochem Biophys 349(2):376–380. https://doi.org/10.1006/abbi.1997.0443
82. Hamberg M, Samuelsson B (1967) On the mechanism of the biosynthesis of prostaglandins E1 and F1?. J Biol Chem 242(22):5336–5343. https://doi.org/10.1016/S0021-9258(18)99433-0
83. Karthein R, Dietz R, Nastainczyk W, Ruf HH (1988) Higher oxidation states of prostaglandin H synthase: EPR study of a transient tyrosyl radical in the enzyme during the peroxidase reaction. Eur J Biochem 171(1–2):313–320. https://doi.org/10.1111/j.1432-1033.1988.tb13792.x
84. Shimokawa T, Kulmacz RJ, DeWitt DL, Smith WL (1990) Tyrosine 385 of prostaglandin endoperoxide synthase is required for cyclooxygenase catalysis. J Biol Chem 265(33):20073–20076. https://doi.org/10.1016/S0021-9258(17)30468-4
85. Tsai A, Palmer G, Xiao G, Swinney DC, Kulmacz RJ (1998) Structural characterization of arachidonyl radicals formed by prostaglandin H synthase-2 and prostaglandin H synthase-1 reconstituted with mangano protoporphyrin IX. J Biol Chem 273(7):3888–3894. https://doi.org/10.1074/jbc.273.7.3888
86. Bhattacharyya DK, Lecomte M, Rieke CJ, Garavito M, Smith WL (1996) Involvement of arginine 120, glutamate 524, and tyrosine 355 in the binding of arachidonate and 2-phenylpropionic acid inhibitors to the cyclooxygenase active site of ovine prostaglandin endoperoxide H synthase-1. J Biol Chem 271(4):2179–2184. https://doi.org/10.1074/jbc.271.4.2179
87 Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997) 3–25.1. Adv Drug Deliv Rev 46(1):3–26. https://doi.org/10.1016/S0169-409X(00)00129-0
88 Saravanabhavan M, Badavath VN, Maji S, Muhammad S, Sekar M (2019) Novel halogenated pyrido [2,3-a]carbazoles with enhanced aromaticity as potent anticancer and antioxidant agents: rational design and microwave assisted synthesis. New J Chem 43(44):17231–17240. https://doi.org/10.1039/C8NJ06504G
89. Aita S et al (2021) Novel α-aminophosphonates of imatinib intermediate: synthesis, anticancer activity, human Abl tyrosine kinase inhibition, ADME and toxicity prediction. Bioorg Chem 109:104718. https://doi.org/10.1016/j.bioorg.2021.104718
90. Guner OF, Bowen JP (2013) Pharmacophore modeling for ADME. Curr Top Med Chem 13(11):1327–1342. https://doi.org/10.2174/15680266113139990037
91. Yamashita F, Hashida M (2004) In silico approaches for predicting ADME properties of drugs. Drug Metab Pharmacokinet 19(5):327–338. https://doi.org/10.2133/dmpk.19.327
92 Hou T, Wang J, Li Y, Wang W (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1: the accuracy of binding free energy calculations based on molecular dynamics simulations. J Chem Inf Model 51(1):69–82. https://doi.org/10.1021/ci100275a
93 Othman IMM et al (2021) Toward a treatment of antibacterial and antifungal infections: design, synthesis and in vitro activity of novel arylhydrazothiazolylsulfonamides analogues and their insight of DFT, docking and molecular dynamic simulations. J Mol Struct 1243:130862. https://doi.org/10.1016/j.molstruc.2021.130862
94. Ali Eltayb W et al (2023) Exploring particulate methane monooxygenase (pMMO) proteins using experimentation and computational molecular docking. J King Saud Univ Sci 35(4):102634. https://doi.org/10.1016/j.jksus.2023.102634
95. Abdalla M et al (2021) In silico studies on phytochemicals to combat the emerging COVID-19 infection. J Saudi Chem Soc 25:101367. https://doi.org/10.1016/j.jscs.2021.101367
96 Abdalla M, Rabie AM (2023) Dual computational and biological assessment of some promising nucleoside analogs against the COVID-19-Omicron variant. Comput Biol Chem 104:107768. https://doi.org/10.1016/j.compbiolchem.2022.107768
97. Karunakar P et al (2014) In silico docking analysis of piperine with cyclooxygenases. J Biochem Technol 3:122–127