Research article

Pharmacophore model, docking, QSAR, and molecular dynamics simulation studies of substituted cyclic imides and herbal medicines as COX-2 inhibitors

Nathalie Moussa a,*, Ahmad Hassan a,**, Sajjad Gharaghanib

a Department of Pharmaceutical Chemistry and Quality Control of Medicaments, Faculty of Pharmacy, Damascus University, Damascus, Syria
b Laboratory of Bioinformatics and Drug Design, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

ARTICLE INFO

Keywords:
Cyclooxygenase-2
Docking
QSAR
Pharmacophore
Molecular dynamics simulation
Virtual screening

ABSTRACT

Cyclooxygenase-2 (COX-2) enzyme inhibitors have not eliminated the necessity for developed drugs not only in the nonsteroidal anti-inflammatory drug (NSAIDs) area, but also in other therapeutic applications including prevention of cancer and Alzheimer's disease. A series of novel substituted cyclic imides have been reported as selective COX-2 inhibitors. To understand the structural features responsible for their activity, a 3D validated pharmacophore and quantitative structure–activity relationship (QSAR) model have been developed. The values of enrichment factor (EF), goodness of hit score (GH), area under the ROC curve (AUC), sensitivity, and specificity refer to the good ability of the pharmacophore model to identify active compounds. Multiple linear regression (MLR) produced statistically significant QSAR model with \( R^2_{\text{training}} = 0.763, R^2_{\text{test}} = 0.96 \) and predictability \( Q^2_{\text{training}} = 0.66, Q^2_{\text{test}} = 0.84 \). Then, using the pharmacophore and QSAR models, eight authenticated botanicals in two herbal medicines and the ZINC compounds database, were virtually screened for ligands to COX-2. The retrieved hits which also obey lipinski’s rule of five (RO5) were docked in the COX-2 3D structure to investigate their binding mode and affinity. Finally, based on the docking results, nine molecules were prioritized as promising hits that could be used as leads to discover novel COX-2 inhibitors. COX-2 inhibition of most of these hits has not been reported previously. Ten-nanosecond molecular dynamics simulation (10-ns MD) was performed on the initial structure COX-2 complex with ZINC000113253375 and ZINC000043170560 resulted from the docking. Our utilization of the 3D pharmacophore model, QSAR, molecular docking, and molecular dynamics simulation trials can be a potent strategy to successfully predict activity, efficiently design drugs, and screen large numbers of new compounds as active drug candidates.

1. Introduction

Non-steroidal anti-inflammatory drugs (NSAIDs) are a commonly prescribed for their established anti-inflammatory, antipyretic, and analgesic properties [1]. NSAIDs inhibit cyclooxygenase enzyme (COX), which mediates the bioconversion of arachidonic acid to inflammatory prostaglandins (PGs). There are two COX isoenzymes (COX-1 and COX-2). COX-1 is responsible for the maintenance of physiological homeostasis, while COX-2 is induced in response to pro-inflammatory conditions.

In spite of their benefits, conventional NSAIDs have significant gastrointestinal toxicity and irritation since they inhibit both COX-1 and COX-2 isoforms. These adverse side effects encouraged the improvement of selective COX-2 inhibitors as promising gastro protective agents [2]. Later on, the long term use of certain COX-2 inhibitor causes ulcer exacerbation in high-risk patients, thrombosis, and kidney toxicity. Thus COX-2 inhibitors are still of interest to researchers and need more work and improvement in the NSAIDs area [3, 4, 5]. In addition, more investigations in the role of COX-2 inhibitors in cancer chemotherapy [6, 7] and neurological diseases such as Parkinson [8] and Alzheimer’s diseases are still needed [9, 10]. Many valuable studies on COX-2 inhibitors have still published until now, as, these inhibitors are still subject to the interest of researchers worldwide [11, 12, 13, 14, 15, 16, 17].

* Corresponding author.
** Corresponding author.
E-mail addresses: nathalie86.moussa@damascusuniversity.edu.sy (N. Moussa), aha226@gmail.com (A. Hassan).

https://doi.org/10.1016/j.heliyon.2021.e06605
Received 8 January 2021; Received in revised form 22 February 2021; Accepted 24 March 2021
2405-8440/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Back to NSAIDs area, they are generally used for the treatment of pain and edema representing the choice of treatment for various inflammatory diseases such as Rheumatoid arthritis and osteoarthritis [18]. Since these drugs are related to negative side effects, new anti-inflammatory drugs are needed and complementary and alternative medicines are being sought [19]. These days about 70% of the world’s population relies on herbal drugs for its primary pharmaceutical care [20]. Natural sources are the most promising pool for drug candidates as revealed by statistics [21].

Examples of medicinal plants used locally to treat arthritis and related disorders are Voltarite® and Rheumax®. Voltarite® contains 5 herbs including Apium graveolens (Celery), Crataegus laevigata (Hawthorn berries), Curcuma longa (Turmeric), Harpagophytum procumbens (Devil’s claw), and Vaccinium myrtillus (Bilberry). Rheumax® contains 4 herbs including Curcuma longa (Turmeric), Boswellia serrata, Tinospora cordifolia, and Vitex negundo. These eight herbs contain 43 major active components, including steroids, terpenes, alkaloids, and flavonoids [22]. It will be a good idea to screen these components to search for unknown COX-2 inhibitors.

In the race to design and develop new therapeutic molecules with the maximum number of pharmacophoric features, computational techniques have emerged as new tools to minimize the time and resources required for their chemical syntheses, in vitro, and in vivo testing. Computer-aided or in silico drug design (CADD) mainly uses computing power to facilitate and accelerate the process of drug design by selecting the most likely lead candidate for biological testing. The most common approaches of CADD include ligand-based drug design (pharmacophore and quantitative-structure activity relationship (QSAR)) and target-based drug design (pharmacophore and molecular docking) [23, 24].

In this study, the Pharmacophore hypothesis and QSAR model were performed for a series of novel cyclic imide derivatives as COX-2 inhibitors. Then, the inhibitory activity towards COX-2 of 43 compounds (in-house library) and zinc “all now” (all purchasable in 2 weeks) database [25] was predicted relying on potent strategy which is the use of pharmacophore fit score (first filter), pIC50 predicted by QSAR model (second filter), docking score and also by compound-enzyme interactions (third filter). The compounds must pass all these filters to be considered as hits recommended for further bio studies. An additional 10-ns MD simulation was performed on the initial structure of the COX-2 complex with the two best compounds resulting from the docking. The root mean square deviations (RMSD) and the radius of gyration (Rg) of the enzyme were calculated to examine the stability of the system (enzyme, water, ions, etc.) [26].

The performance of our pharmacophore and QSAR models was validated by many strict parameters in order to use both of them effectively in virtual screening [27, 28]. We especially emphasize procedures used to define QSAR model applicability domain that should be used when the model is employed for the prediction of external compounds [29, 30]. In other words, we use machine learning to identify already-known chemical compounds as potential novel COX-2 inhibitors that have not yet been recognized as such.

2. Material and methods

2.1. Pharmacophore Model

2.1.1. Data set and pharmacophore modeling

A 3D ligand-based pharmacophore model was created using LigandScout v4.4.1 from five potent cyclic imide compounds collected from the literature. The values of the inhibitory activity of these COX-2 inhibitors ranged from 100 to 360 nM. The structures and experimental biological activities (IC50) of these compounds are listed in Table 1. The three dimensional (3D) structures of all compounds were constructed using Chemdraw. For pharmacophore modeling, LigandScout used espressso algorithm which has many stages including clustering and conformation generation [31].

LigandScout’s clustering method divides the ligands into test-set and training-set. The aim of this clustering is to choose compounds that are similar in terms of 3D pharmacophore characteristics and therefore bear a higher chance of delivering a large overlap of chemical features. The 3D clustering algorithm performs fast alignments and clusters based on a similarity value between 0 and 1. Since this algorithm basically performs combinatorial alignments of all conformations of all compounds, a low number of conformations is recommended. The cluster distance can be varied until the desired cluster size is reached. Conformations of the training-set molecules were generated. After ranking the molecules according to their number of conformations (flexibility), pharmacophore features were projected onto these molecules and all their conformations. All conformations of the two top-ranked (i.e. the least flexible) molecules are then aligned using IntelLigand’s molecular alignment algorithm [32].

2.1.2. Pharmacophore model validation

The validation of a pharmacophore model is considered an important step before its use in virtual screening. The Predictive ability, specificity, and sensitivity of a pharmacophore model are earnest metrics for the reliability of performance. We assessed the predictive ability of the model on a decoy set to determine the accuracy of recognition of active and inactive compounds.

Sensitivity states how good the model correctly classifies compounds and specificity shows how well the model is able to exclude inactive compounds [28]. The Sensitivity (TPR) and specificity (TNR) can be measured using Eqs. (1) and (2), respectively.

\[
\text{Sensitivity} = \frac{\text{true positives}}{\text{true positives in the database}} = \frac{TP}{A} = TPR \tag{1}
\]

\[
\text{Specificity} = \frac{\text{true negatives}}{\text{true negatives in the database}} = \frac{TN}{D} = TNR \tag{2}
\]

Other metrics to validate the pharmacophore model are the receiver operating characteristic (ROC) curve which indicates how well a model
can differentiate between active and inactive compounds [31] and the area under the ROC-curve (AUC) [33]. The ROC curve provides the true positive rate plotted against the false positive one of the hits. If the curve was sharp and then flattened, this means that the model ranked the active compounds higher than the inactive ones. The AUC value lies between 0 (bad classifier) when the model ranks all the inactive compounds first and 1 (excellent classifier) on the contrary situation. Hence, to evaluate the performance of the model, a set of 703 inactive compounds was obtained from DUD-E as a decay set for 5 active and selective COX-2 inhibitors. The decay and active compounds have similar physicochemical properties but different two dimensional topological ones [34, 35, 36]. Using the idbgen routine included in LigandScout, the compounds were converted into a LigandScout format [52]. In addition to these metrics, the statistical parameters of goodness of hit score (GH) [37], enrichment factor (EF), and accuracy (ACC) were determined to investigate the performance of the model. The formulas are written in Eqs. (3), (4), and (5) below:

\[ GH = \frac{H_a}{3H_a + H_t} \left[ 1 - \left( \frac{H_t - H_a}{D - A} \right) \right] \quad (3) \]

\[ EF = \frac{H(a+D)}{H_a} \quad (4) \]

\[ ACC = \frac{TP + TN}{A + D} \quad (5) \]

Ht is the total number of hits and Ha is the number of active hits. The range of the GH score is (0–1) with a threshold value equal to 0.6 [38]. Ideally, when the model picks all the active compounds with no inactive ones, it will have a steep slope for the ROC curve, a high value of AUC, a high EF value, and the highest value of sensitivity and specificity which is 1 [39].

2.1.3. Pharmacophore-based virtual screening (PBVS)

Virtual screening, based on the best pharmacophore model as a query, was performed using LigandScout v4.4.1 [40]. A part of the ZINC database “All Now” and a small dataset consisting of 43 active components in some herbal preparations were used for the screening study. The hit compounds were retrieved based on the Pharmacophore Fit Score values and then evaluated based on their drug-likeness properties using Lipinski’s ROS [41].

2.2. QSAR model

2.2.1. Data set

For QSAR studies a series of forty cyclic imide derivatives with their COX-2 reported IC50 values were compiled from recently published studies [42, 43, 44]. Various structures of dataset were selected. No large gaps were allowed between activity values. For QSAR modeling, the negative logarithm pIC50 of the biological activities IC50 was used. The pIC50 values ranged between 4.151 and 7.000 with an average of 5.401 [45]. The data set was divided randomly using MATLAB into a training-set (32 molecules) to build the model and the eight remaining molecules were used to test the performance of the model (test-set) [46]. The structures of the studied molecules and their corresponding experimental biological activities are listed in Table 2.

2.2.2. Optimization of compounds

The structures of the molecules were assembled using Hyperchem software (version 8.0; Hyperchem, Alberta, Canada) (13). The geometry of the compounds was optimized using Hyperchem. The molecular mechanics force field (MM+) then semi empirical method AM1 were applied [45].

2.2.3. Descriptors generation and features selection

For each compound, 1,875 descriptors were calculated using PaDEL software was used to calculate 1,875 descriptors [47]. The calculated descriptors which encoding different properties like physiochemical, electronic, and topological properties were analyzed for the existence of constant or near constant variables (standard deviation of 0.1 as a threshold) using MATLAB. The detected ones were then removed. Collinear descriptors (i.e. correlation coefficient between descriptors is greater than 0.9) were detected and the one displaying the most noteworthy correlation with the activity was held and others were expelled from the data. Finally, 191 descriptors remained. SPSS (version 13.0; SPSS Inc., Chicago, IL, USA) (15) statistical software was utilized to select the features that had the minimum number of descriptors (simplicity) and kept good performance [48]. The QSR equation consisted of the descriptors (X) and the measured pIC50 (Y). To correlate between X and Y, Multiple Linear Regression (MLR) method was used. Different models were generated with different R squares and numbers of descriptors for each model. These values are crucial for the selection of acceptable models. No less than five compounds should be involved in the equation for each descriptor. The performance of MLR model was evaluated by calculating root-mean-square error (RMSE) and from R2 values. RMSE was calculated by Eq. (6):

\[ \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{n}} \quad (6) \]

yi is the wanted output observation, y0 is the predicted value, and ns is the number of the total compounds in the data set. MATLAB was used to do all the calculation for the training and test sets like calculating R2, RMSE, and Q2.

2.2.4. Validation

To confirm the robustness of the model, the dataset was divided into a training-set consists of 32 compounds and test-set (8 compounds) and statistical parameters (Eqs. (7), (8), (9), (10), (11), and (12)), in addition to R2 in accordance with Tropsha et al, Roy and Roy were calculated.

\[ Q^2 > 0.5, \quad (7) \]

\[ R^2 > 0.6, \quad (8) \]

\[ \left| \frac{R^2_{\text{test}}}{R^2_{\text{tra}}} \right| < 0.1, \quad (9) \]

\[ \left| \frac{R^2_{\text{test}}}{R^2_{\text{tra}}} \right| < 0.1, \quad (10) \]

\[ 0.85 \leq k \leq 1.15 \text{ and } 0.85 \leq k' \leq 1.15 \quad (11) \]

\[ R^2 = r^2_{o} \times \left( 1 - \left( r^2_{p} - r^2_{o} \right)^{1/2} \right) \quad (12) \]

The R2 value should be larger than 0.5 [49] to express that the model has good external prediction.

2.2.5. Applicability of domain

The applicability of domain (AD) is widely comprehended in QSAR field to estimate the unreliability and vulnerability in the prediction of a specific molecule based on how similar it is to the compounds used to build the model [50]. In this study, we used the Williams plot to evaluate the AD of our QSAR model. The Williams plot provides leverage values (hi) plotted against the standardized residuals. The leverage hi for each compound, was calculated by Eq. (13) to predict its activity by QSAR model [51]:
Table 2. Structures of compounds used in the QSAR study, and pIC\textsubscript{50} values.

| No | R1       | R2       | pIC\textsubscript{50} | pIC\textsubscript{50} predicted |
|----|----------|----------|------------------------|---------------------------------|
| 3–4 |          |          |                        |                                 |
| 3t  | R1 = MeO |          | 5.283                  | 5.53                            |
| 4r  | R1 = Cl  |          | 4.978                  | 5.174                           |
| 5–10 |          |          |                        |                                 |
| 5r  | R1 = MeO, R2 = H |          | 5.193                  | 5.370                           |
| 6r  | R1 = Cl, R2 = H |          | 5.040                  | 5.051                           |
| 7r  | R1 = MeO, R2 = 5-NO2 |          | 7.000                  | 6.630                           |
| 9r  | R1 = MeO, R2 = 5-Me |          | 6.397                  | 5.433                           |
| 11t | R1 = MeO, R2 = 5-tert-But |          | 5.000                  | 4.567                           |
| 12r | R1 = Cl, R2 = 5-tert-But |          | 4.684                  | 4.571                           |
| 13r | R1 = MeO, R2 = 5,6- Dichloro |          | 4.906                  | 5.003                           |
| 14r | R1 = Cl, R2 = 5,6-Dichloro |          | 4.521                  | 4.704                           |
| 15r | R1 = MeO, R2 = 4,5,6,7-Tetrachloro |          | 4.347                  | 5.064                           |
| 18–19 |          |          |                        |                                 |
| 18r | R1 = MeO |          | 6.346                  | 6.747                           |
| 19r | R1 = Cl  |          | 6.000                  | 5.990                           |
| 20–21 |          |          |                        |                                 |
| 20r | R1 = MeO |          | 6.346                  | 6.251                           |
| 21t | R1 = Cl  |          | 6.000                  | 5.796                           |

(continued on next page)
| No   | pIC_{50} Predicted | pIC_{50} predicted | pIC_{50} | pIC_{50} predicted |
|------|--------------------|--------------------|----------|--------------------|
| 22–31 |                    |                    |          |                    |
| 23r  | R = H, Ar = b      | 4.602              | 4.525    |                    |
| 24t  | R = 5-Me, Ar = a   | 4.301              | 4.025    |                    |
| 27r  | R = 5-tert-But, Ar = b | 4.301          | 5.074    |                    |
| 28t  | R = 5,6-Dichloro, Ar = a | 6.397           | 6.299    |                    |
| 30r  | R = 5-nitro, Ar = a | 7.000              | 7.221    |                    |
| 31r  | R = 5-nitro, Ar = b | 6.522              | 6.239    |                    |
| 32–33 |                    |                    |          |                    |
| 32r  | Ar = a             | 6.096              | 6.310    |                    |
| 33r  | Ar = b             | 6.000              | 5.434    |                    |
| 34–35 |                    |                    |          |                    |
| 34r  | Ar = a             | 5.522              | 5.617    |                    |
| 35r  | Ar = b             | 5.397              | 4.810    |                    |
| 37r  |                    | 4.151              |          |                    |
| 38–52 |                    |                    |          |                    |
| 38r  | R1 = 3,4,5-trimethoxy | 4.471            |          | 4.336              |
|      | R2 = H             |                    |          | 5.597              |
| 41t  | R1 = 3,4,5-trimethoxy | 4.619            |          |                    |

(continued on next page)
where $x_i$ is the descriptor row vector of the query molecule and $x$ is the $k \times n$ matrix where $k$ is the descriptor value for $n$ molecule from the training-set. The diagonal elements in this matrix represent the leverage values ($h$) for the molecules in the dataset. The warning leverage, $h^*$, was fixed at $3p/n$, where $p$ is the number of descriptors plus one and $n$ is the number of training samples. A molecule with leverage higher than $h^*$ may refer to unreliable predictions. The developed QSAR model could be used to predict the activity of certain compound only if it was inside the applicability domain as Williams plot showed.

MLR was used to determine the standardized residuals of inhibitory activity. The calculated leverage and standardized residuals values were used for definition the AD. The AD was identified as a square area between $\pm 3$ standard deviation and $h^*$. Leverage values for all of training-set and test-set compounds were calculated. The compound was considered inside of the applicability domain only if had standardized residual $< 3$ standard deviation units and leverage value not exceeding $h^*$.

Table 2 (continued)

| No  | $pIC_{50}$ | $pIC_{50}$ predicted |
|-----|------------|----------------------|
| 42r | R1 = 4-methoxy | 4.442 |
|     | R2 = CH3 | 4.289 |
| 44t | R1 = 3,4,5-trimethoxy | 4.492 |
|     | R2 = tert-butyl | 4.295 |
| 45r | R1 = 4-methoxy | 4.536 |
|     | R2 = tert-butyl | 4.388 |
| 46r | R1 = 3,4,5-trimethoxy | 6.744 |
|     | R2 = NO2 | 6.500 |
| 47t | R1 = 4-methoxy | 6.619 |
|     | R2 = NO2 | 6.448 |
| 48t | R1 = 4-fluoro | 4.507 |
|     | R2 = NO2 | 5.752 |
| 50r | R1 = 4-methoxy | 6.443 |
|     | R2 = 5,6-Dichloro | 5.408 |
| 52r | R1 = 4-methoxy | 5.387 |
|     | R2 = 3,4,5,6-tetrachloro | 4.968 |
| 53r | R1 = 3,4,5-trimethoxy | 5.070 |
|     | | 4.860 |
| 55r | R1 = 3,4,5-trimethoxy | 5.136 |
|     | R2 = NO2 | 4.974 |
| 56r | R1 = 4-methoxy | 5.317 |
|     | R2 = NO2 | 5.317 |
| 57r | R1 = 3,4,5-trimethoxy | 6.522 |
|     | R2 = NO2 | 6.148 |

$r$: training, *t: test.
2.2.6. Prediction

QSAR is an important part in computational drug design [52]. The QSAR equation was applied to the compounds that we got by pharmacophore virtual screening on ZINC “all now” database and the 43 herbal components in order to predict their inhibitory activity on the COX-2 enzyme. The expected inhibitory activity must be considered reliable just for inhibitors that fall within the AD pool on which the model was developed. Leverage values and the predicted pIC50 of the retrieved compounds are listed in Table 5.

Later on, only the compounds that fall inside the AD and pass the QSAR filter will succeed to reach the docking study.

2.3. Molecular docking

All the compounds that achieved both good pharmacophore fit score in pharmacophore-virtual screening and good pIC50 predicted by QSAR (PDB code: 5KIR, 2.697 Å) besides celecoxib, rofecoxib, and diclofenac were docked to COX-2 (PDB code: 5KIR, 2.697 Å) [53]. The Schrödinger suite of software (Maestro, version 12.1) was used to do all the docking procedure with its calculations and scores.

Molecular docking was performed using the extra-precision (XP) mode of Glide (Grid-based Ligand Docking with Energetics). Glide algorithm searches for ligand orientations, positions, and conformations in the enzyme-binding pocket. The final evaluation of binding energy was done with Glide score (G Score). Glide performance was validated by a re-docking approach. Before docking the nine molecules, the bound ligand found in the X-ray crystal structure was docked back into the binding pocket of COX-2 enzyme. This was done to confirm that Glide could imitate the position and orientation of the inhibitors as observed in the crystal structure.

2.3.1. Ligand preparation

The two-dimensional (2D) structures of the studied inhibitors were sketched using the freeware ACD/ChemSketch version 12.01 (ACD/Labs Release, Canada) [54]. Then, they were converted into the standard structure-data file (SDF) format using the freely available open source toolbox, Open Babel [55].

Energy minimization was conducted using optimized potential for liquid simulations (OPLS3) force field [56, 57] using Lig Prep Module provided by Maestro version 12.1 (Schrödinger, LCC, New York, 2019) [58]. One three-dimensional (3D) conformer that has the lowest energy and correct chirality was generated for each ligand to be used during the docking procedure.

2.3.2. Protein preparation

The 3D structure of COX-2 (PDB code: 5KIR, 2.697 Å) was retrieved from the Protein Data Bank (PDB) and processed using Protein Preparation Wizard from Maestro version 12.1 (Schrödinger, LCC, New York, 2019). Protein preparation was executed for chain A only, while the other chain and extra molecules like water molecules, and arachidonic acid were removed. Steps of protein preparation involve adding explicit hydrogen atoms, assigning bond orders and formal charges, creating disulfide bonds, capping termini, and finding overlaps. The corrected structure was energetically optimized and minimized to relieve any strain and to fine-tune the placement of various groups [59]. Protein-energy minimization was conducted using OPLS3 force field.

2.3.3. Receptor grid generation

Glide Receptor Grid Generation platform as a part of Maestro was employed to create a rigid grid around the interacting residues of the active site from the prepared protein. The default grid box size of 20 Å was used and the size of the inner grid box was changed to (12 × 12 × 12 Å3). The centroid of grid box was located on the ligand. Grids for molecular docking with Glide16 were calculated with a hydrogen bond constraint to Arg 513 of COX-2 [60].

2.3.4. Extra precision (XP) molecular docking

To predict the binding pose of ligands COX-2, extra precision (XP) Glide docking procedure was utilized to dock the ligands into the generated receptor grid. XP refines the predicted docking modes using an anchor-and-grow algorithm to more thoroughly sample ligand degrees of freedom [61, 62]. Glide workflow from Maestro version 12.1 (Schrödinger, LCC, New York, 2019) was used during this step. Ligands were docked flexibly into the rigid docking box using Glide's internal conformation generator and the “Alternate Protocol 2” which is the process of applying constraints in a flexible ligand docking experiment. For each ligand, one docking hit with the lowest Glide docking score was generated and used to analyze docking results [63].

2.4. Molecular dynamics simulation of COX-2

The MD simulations were performed using the GROMACS 2019.1 [64]. The topology parameters of COX-2 were created. The interaction parameters were computed using the charmm36 force field. The system was immersed in a cubic water box (9.906 + 9.906 + 9.90618 nm3) of extended simple point charge (SPC) water molecules. The solvated system was neutralized by adding one chloride ions in the simulation, and

Figure 1. Mapping the most active training compound on the model in 2D (a) and 3D (b).
the entire system was composed of 8790 atoms of COX-2, 1 Cl⁻
counters, and 28997 solvent atoms. The energy was minimized using
the steepest descent method of 1000 steps. Then, simulations were
performed in a constant number of molecules, constant pressure,
and constant temperature (NPT) ensemble with coupled temperature
and pressure at 310 K and 1 bar, respectively. In the next step, the
solute (protein and counterion) was fixed and the position-restrained
dynamics simulation of the system, in which the atom positions of COX-2
were restrained at 300 K. The water and the counterion permitted to
relax about the protein. The relaxation time of water was 20 ps. Finally, the full
system was subjected to 10-ns MD at 300 K temperature and 1 bar
pressure. The MD simulation and results analysis were performed on the
Ubuntu 18.04.3 LTS Linux on an Intel® Xeon(R) CPU E5-2650 v4 @ 2.20GHz
/C2 24 and 64 GB of RAM.

2.5. Molecular dynamics simulation on COX-2 complex with 375 and 560

The complex of ZINC000113253375 (375) and ZINC000043170560 (560) with COX-2 was chosen for MD simulation. The topology parameters
of 375 and 560 were built by the CHARMM General Force Field
(CGenFF) program [65], and were embedded to the topology parameters
of COX-2. An additional 10-ns MD simulation was performed on the
initial structure COX-2 complex with 375 and 560 resulted from the
docking. All the simulation processes were performed by the GROMACS
2019.1 and the same as previous section and at the end of simulation
results of the complex of 375 and 560 with COX-2 were analyzed.

3. Results and discussions

3.1. Pharmacophore

3.1.1. Pharmacophore modeling

The aim of this study is finding new COX-2 inhibitors by ligand-based
pharmacophore modeling. Based on five potent and diverse cyclic imides
collected from the literature (mentioned in Table 1 with their biological
activities IC₅₀ [42,43], a pharmacophore model was created. The
pharmacophore model of the most active compound is displayed in
Figure 1. The eleven features of the generated pharmacophore were two
aromatic (R) showed as blue sphere, one hydrophobic (H) colored by
yellow sphere, one H-bond donors (D) showed as green sphere, six
H-bond acceptors (A) showed as red sphere, one positive charge (P)
colored by blue lines. The Pharmacophore Fit Score values are used to
measure the overlapping between the features of pharmacophore and
chemical functionalities of the compound and listed in Table 3. The
values of Pharmacophore Fit Score of the compounds in training-set are
high (103.87–106.94) and indicate good mapping with the model.

Accordingly, the model could be a hopeful query in virtual screening in
order to find promising active hits so screening decoy and active sets was
done to evaluate the pharmacophore model.

3.1.2. Pharmacophore model validation

Two datasets including five COX-2 inhibitors and 703 inactive com-
ounds were used to validate the model. The values of enrichment factor
(EF) and goodness of hit score (GH) refer to the good ability of the model
to identify active compounds. The receiver operating characteristic
(ROC) curve of the pharmacophore model is shown in Figure 2. The
retrieved hits by the model are 3 and the area under the ROC curve (AUC)
value is 0.8%. The pAUC (partial area under the curve) values are 1.00,
1.00, and 1.00 at 1%, 5% and 10% of the screened database respectively.
Also, the model is sensitive because it retrieved 3 active compounds from
5 ones (60% of the total active compounds) and very specific because it
selected no decoys. This high selectivity and sensitivity indicates that our

Table 3. The Pharmacophore-Fit Score values of the training and test compounds.

| Compounds | COX-2 IC₅₀ μM | Pharmacophore-Fit score | Matched Feature Pairs |
|-----------|--------------|------------------------|-----------------------|
| 30        | 0.1          | 106.943                | 11                    |
| 46        | 0.18         | 100.663                | 10                    |
| 47        | 0.24         | 103.873                | 10                    |
| 49t       | 0.28         | 71.428                 | 7                     |
| 50t       | 0.36         | 73.634                 | 7                     |
| Celecoxib | 0.3          | 52.520                 | 6                     |

| Features | AAAAAAAAAHDRRP |
|----------|----------------|
| Ha       | 3              |
| Ht       | 3              |
| A        | 5              |
| D        | 703            |
| TPR      | 60%            |
| FPR      | 0%             |
| AUC      | 0.8            |
| ACC      | 0.997          |
| Sensitivity | 0.6          |
| specificity | 1            |
| GH       | 0.6            |
| EF       | 141.6          |

Features: A: H-bond acceptors, H: hydrophobic groups, D: H-bond donors, R:
aromatic ring, P: positive charge.
Ha: active hits, Ht: total hits, A: total number of actives in the dataset, D:
total number of decoys in the dataset, TP: true positive, FP: false positive, EF:
enrichment factor, GH: goodness of hit, AUC: area under ROC curve, ACC:
accuracy.
The pharmacophore is a very good filter for recognizing COX-2 inhibitors \cite{27}. All the parameters of validation process are shown in Table 4 and state good quality of the pharmacophore model so it can be used successfully in virtual screening. 

### 3.1.3. Pharmacophore-based virtual screening (PBVS)

The most active training compound (30), which has high Pharmacophore Fit Score (106.94) mapped on the model in Figure 1. Virtual screening was done to identify new inhibitors of COX-2 by the validated pharmacophore model. ZINC database containing of 5002 molecules were converted into a LigandScout format using (idbgen) function. Six molecules of ZINC database were selected by screening using the model. The values of Pharmacophore Fit Score of the hits ranged from 79.65 to 96.92. In addition, small herbal components data contained 43 compounds was exposed to all mentioned steps. Table 5 shows the pharmacophore fit scores of all the retrieved hits. 17 active components have had pharmacophore fit scores as good as the pharmacophore fit score of celecoxib and/or rofecoxib. These 23 compounds...
3.2. QSAR study

3.2.1. QSAR model

The best QSAR model was built using 7 descriptors such as extended topochemical atom, information content, auto correlation, RDF, BCUT, barysz matrix, atom type electrotopological state and molecular linear topochemical atom, information content, auto correlation, RDF, BCUT, barysz matrix, atom type electrotopological state and molecular linear free energy relation. For the selection of the most important descriptors, stepwise MLR method was used.

The MLR analysis with a stepwise selection was carried out to relate the pIC50 to a 7 set of descriptors. The SPSS software (version 13.0; SPSS Inc., Chicago, IL, USA) (15) was used for the MLR analysis). It is defined by Eq. (14):

\[
p\text{IC}50 = -6.2128 (±2.496) +0.1658 (±0.038) \text{ETA-beta} +2.7284 (±0.491) \text{IC1} +0.9179 (±0.239) \text{GATS8m} -0.0017 (±0.000) \text{VR1_Dz} +0.1979 (±0.056) \text{HdCH} -0.0402 (±0.011) \text{RDF70} -0.3436 (±0.159) \text{BCUTp-II}
\]

(14)

The built model produced good results for the training-set and the test-set. The 7 descriptors with their contribution to the model and their statistical parameters are shown in Table 6.

3.2.2. Validation results

Results show that this model has a cross-validated correlation coefficient (Q^2) value higher than 0.5 (0.8407). Q^2 is important parameter but not enough to judge the power of the model. Indeed, the true predictive power of a QSAR model can be established only through model validation procedure which consists of prediction of activities of compounds in test set (i.e. not included in model building). A strong validation should be performed to insure good reliable, prediction, and conclusion of the QSAR equation. This implies quantitative evaluation and a set of statistical criteria. Such as R^2 which is a measure of the degree of the fit between the predicted outcomes and the experimental ones. Other statistical parameters were performed to insure good quality of the model [29, 30].

R^2 (calculated for the test set) value larger than 0.6. R^2 is the correlation coefficient for regressions between predicted against experimental values through the origin. R^2 is the correlation coefficient for regressions between experimental against predicted values through the origin. R^2 is an additional parameter was specified as a worthy parameter of the external prediction. The value of R^2 is here 0.8778 (larger than 0.5) and that means the model has good external prediction [49].

The values of slopes (k and k') of regression lines through the origin are very close to 1 and fall in the acceptable range between 0.85 and 1.15. It could be seen from the results that all criteria were satisfied thus giving power and trust for the developed model. Statistical parameters of the external test set for the MLR model are given in Table 7. The predicted pIC50 values versus their experimental values were plotted in Figure 3 for the training-set and test-set.

3.2.3. Applicability of domain

The warning leverage, h*, was fixed at 3p/n (h* = 0.75) in this study, where p is the number of descriptors plus one and n is the number of training samples.

The applicability domain (AD) was defined as a square region between leverage threshold h* of 0.75 and ±3 standard deviation. Leverage values for all compounds in training-set and test-set were calculated (Figure 4). As mentioned before, the compound was considered inside of the AD only if it had leverage value ≤0.75 and standardized residual ≤3 standard deviation units. It can be concluded from the plot that all compounds of training-set and the test-set are situated inside the assigned domain. One of the inhibitors with response over the threshold of ±3 standard deviation unit of standardized residual.

Concerning the previously conditions, the suggested model can be used with a high level of confidence right now.

3.2.4. Prediction

Twenty-three molecules from ZINC database and our in-house library were selected by virtual screening using the pharmacophore model. Then only nine compounds from those twenty-three passed our QSAR model and its applicability domain filter. These nine compounds were ZINC000002939626, ZINC000000009029, ZINC000114185151, ZINC000113253375, ZINC000043170560, Astragal, Curcumin, cyanidin and isoquercitrin. Curcumin is the main bioactive compound isolated from rhizomes of curcuma longa. Isoquercitrin and astragalin, naturally occurring flavonoids, have been identified in a variety of medicinal plants such as the eight herbs in Voltarit® (Billberry and Hawthorn berries) (Voltarit®) [22]. The pIC50 of all tested compounds listed before in Table 5.

3.3. Docking results and discussion

Finally, to confirm the discovery of new lead compounds, we finished with the docking study of the compounds retrieved from ZINC and herbal data to choose the best hits that have the best glide docking score. For validation the reliability of docking, the heavy-atom root mean squared deviation (RMSD) value was determined between the crystal ligand and re-docked ligand using Schrodinger. The value of RMSD equaled to 0.5 Å (no more than 2 Å) and that reveal good agreement between the experimental and predicted binding pose [66]. The hits that showed good pharmacophore score with good predicted pIC50 from QSAR model (5 compounds from ZINC and 4 compounds from herbal data that are listed in Table 8) were exposed to docking with the 3D structure of COX-2 (PDB code: 5KIR, 2.697 Å) by GLIDE.
Types of interactions of the hits, celecoxib and rofecoxib with the binding site of COX-2.

| Compound | Glide docking score (kcal/mol) | Interaction type with Arg513 | Distance A (Å) |
|----------|-------------------------------|-----------------------------|---------------|
| ZINC000029396226 | -7.956 | H-bond and positive charge | 2.347 |
| ZINC000113253375 | -9.293 | positive charge | 2.761 |
| Curcumin | -9.096 | H-bond and positive charge | 2.417 |
| Cyanidin | -6.360 | positive charge | 2.063 |
| Celecoxib | -10.452 | H-bond and positive charge | 2.177 |
| Rofecoxib | -9.734 | positive charge | 2.063 |

The structure of the catalytic domain, which is the largest region in COX enzyme, is similar between COX-1 and COX-2 (RMS deviation of 0.4 Å). The Arg120, Tyr355, and Glu524 residues form the entrance of the active site while we find Tyr385 and Tyr348 in the apex of the active site. The main differences between both active sites are at the positions 513, 434 and 523. In other words, Ile434 and Ile523 in COX-1 are replaced by valine residues in COX-2. The hydrophobic His513 in COX-1 was replaced by the positively charged arginine in COX-2. These differences are the key points of selective inhibitors [67, 68]. The residue 513 plays a vital role for ligands to distinguish the structural difference between the active sites of COX isoforms [69]. The reason mentioned above elucidates one way in which many hit compounds of the present study selectively inhibits COX-2.

The extra precision glide docking scores for the hits ranged from -7.279 to -9.764. The glide docking score for sodium diclofenac, rofecoxib and celecoxib and as a reference potent COX-2 inhibitors were -7.839, -9.735 and -10.452 respectively. Analysis of the binding poses disclosed that the compounds oriented in the COX-2 binding cavity in a similar way; i.e. binding with residues from both the membrane-binding domain (MBD) and the catalytic domain as summarized in Table 8. Moreover, binding modes of all studied compounds were close to that of the selective COX-2 inhibitors rofecoxib and celecoxib, especially the interaction with Arg513 which is essential for selective COX-2 inhibition [70]. However, celecoxib binds to Arg513 through positive charge interaction between the sulfonamide and guanidinium. The hits interact with Arg513 via positive charge and/or hydrogen bond. Access of celecoxib and hits to the adjacent pocket was through hydrophobic interaction with Tyr91. Other types of interactions like polar, hydrogen bonding, negative/positive charges, etc. with other residues in the binding site were also involved in the binding pose and will be explained for only the significant COX-2 inhibitors [71]. In the side pocket of the COX-2 channel, the methyl sulfone moiety of rofecoxib binds to it and the phenyl ring reaches the side chain of Tyr385. Rofecoxib contacts COX-2 channel's residues 42 times. All the interactions are hydrophobic ones and only one hydrophilic placed between O atoms of the methyl sulfone moiety of the inhibitor and the side-chain N atoms of His90 and Arg513 found in the base of the side pocket [72].

ZINC00043170560, ZINC00113253375 and Astragalin (as a plant component) had the lowest docking score amongst the hits as shown in Table 8 suggesting good binding poses and stable ligand-enzyme complexes. Figure 5 shows that ZINC00013253375 binds to the residues His90, Phe518 and Arg513 sited at the base of the side pocket (polar interaction with His90, hydrogen bond and hydrophobic interaction with Ph518, positive charge with Arg513). Figure 6 shows that Astragalin also binds to residues His90, Phe518 and Arg513 (polar interaction with His90, hydrogen bond and hydrophobic interaction with Ph518, positive charge and hydrogen bond with Arg513). They also bind to the hydrophobic side pocket contained of residues Leu352, Ser353, Ile517, and Phe518 (polar interaction with Ser353 and hydrophobic ones with the rest residues). We also see the hydrophobic interaction with residues Tyr355 lie at the mouth of the COX active site, and the catalytic Tyr385 located at the apex of the hydrophobic channel.

Astragalin binds to the residue Glu192 that contributes to the outer shell of the side pocket with polar interaction and hydrogen bond and also binds to Tyr385 with hydrogen bond [71, 72].

### 3.4. Molecular dynamics simulation

The stability of the system (enzyme, water, ions, etc.) was examined by the calculation of root mean square deviations (RMSD) and the radius of gyration (Rg) of the enzyme with respect to its initial structure.

Figure 7 shows the time history of RMSD for COX-2, COX-2-375 and COX-2-560. Analysis of this figure indicates that the RMSD of these systems reaches equilibration and oscillates around in average value after 3.5 ns. The RMSD values indicated that conformation of COX-2 has been
Figure 5. Schematic representation of the interactions between compound ZINC000113253375 and COX-2 active site.

Figure 6. Schematic representation of the interactions between Astragalin and COX-2 active site.
Astragalin (the direct inhibition of COX-2 not through mediators) have not been reported before. We find our contribution significant for future in vitro studies of these compounds or their derivatives.

**Declarations**

**Author contribution statement**

Nathalie Moussa: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Ahmad Hassan; Sajjad Gharaghani: Conceived and designed the experiments; Analyzed and interpreted the data.

**Funding statement**

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

**Data availability statement**

Data will be made available on request.

**Declaration of interests statement**

The authors declare no conflict of interest.

**Additional information**

No additional information is available for this paper.

**Acknowledgements**

We thank University of Vienna for providing an opportunity to learn many approaches and techniques used in computational drug design during EUROPIN Summer School on Drug Design 2019.

**References**

[1] K. Abouzid, P. Frohberg, J. Lehmann, M. Decker, 6-Aryl-4-oxohexanoic acids: synthesis, effects on eicosanoid biosynthesis, and anti-inflammatory in vivo-activities, Med. Chem. 3 (2007) 433–438.

[2] H. Ulbrich, B. Fiebich, G. Dannhardt, Cyclooxygenase-1/2 (COX-1/COX-2) and 5-lipoxygenase (5-LOX) inhibitors of the 6,7-diaryl-2,3-H-dihydroprroloazine type, Cheminform 34 (20) (2003) 953–959.

[3] G. Dannhardt, S. Laufer, Structural approaches to explain the selectivity of COX-2 inhibitors: is there a common pharmacophore? Curr. Med. Chem. 7 (11) (2000) 1101–1112.

[4] H.-J. Kim, G.H. Chae, K.Y. Yi, K.-L. Park, S. Yoo, Computational studies of COX-2 inhibitors: 3D-QSAR and docking, Bioorg. Med. Chem. 12 (7) (2004) 1629–1641.

[5] C. Almansa, A.F. de Arriba, F.L. Cavalcanti, I.A. Gómez, A. Miralles, M. Merboa, J. Fortm, Synthesis and SAR of a new series of COX-2-selective inhibitors: pyrazolo[1,5-a]pyrimidines, J. Med. Chem. 44 (3) (2001) 350–361.

[6] M. Tsuji, S. Kawanow, S. Tsuji, H. Sawasaka, M. Hori, R.N. Doldois, Cyclooxygenase regulates angiogenesis induced by colon cancer cells, Cell 93 (5) (1998) 705–716.

[7] M. Tsuji, S. Kawanow, R.N. Doldois, Cyclooxygenase-2 expression in human colon cancer cells increases metastatic potential, PNAS USA 94 (7) (1997) 3336–3340.

[8] M. Asanuma, I. Miyazaki, Nonsteroidal anti-inflammatory drugs in experimental Parkinsonian models and Parkinson’s disease, Curr. Pharmaceut. Des. 14 (2008) 1428–1434.

[9] J.C. Breitner, P.P. Zandi, Do nonsteroidal anti-inflammatory drugs reduce the risk of Alzheimer’s disease? N. Engl. J. Med. 345 (2001) 1567–1568.

[10] P.L. McGee, J. Rogers, E.G. McGee, Inflammation, anti-inflammatory agents and Alzheimer disease: the last 12 years, J. Alzheimers Dis. 9 (2006) 271–276.

[11] Al-Hussain Alu-Hashem, Zaki, Synthesis of novel benzodifuranyl; 1,3,5-triazines; 1,3,5-oxadiazepines; and thiazolopyrimidines derived from visnaginone and khellinoline as anti-inflammatory and anaglyces agents, Molecules 25 (11) (2020) 226.

[12] Y. Zhang, Y. Zheng, W. Shi, Y. Guo, T. Xu, Z. Li, J. Li, Design, synthesis and investigation of the potential anti-inflammatory activity of 7-O-amide hesperetin derivatives, Molecules 24 (20) (2019) 3663.

[13] M.T.-E. Maghraby, O.M.F. Abou-Ghadir, S.G. Abdel-Moty, A.Y. Ali, O.I.A. Salem, Novel class of benzimidazole-thiazole hybrids: the privileged scaffolds of potent anti-inflammatory activity with dual inhibition of cyclooxygenase and 15-lipoxygenase enzymes, Bioorg. Med. Chem. 26 (7) (2020) 1154–1159.
N. Mousa et al.  
Heliyon 7 (2021) e06605
N. Moussa et al. Heliyon 7 (2021) e06605

[75] R.A. Sharma, A.J. Gescher, W.P. Steward, Curcumin: the story so far, Eur. J. Canc. 41 (2005) 1955–1968.

[76] R. González, I. Ballester, R. López-Posadas, M.D. Suárez, A. Zarzuelo, O. Martínez-Augustin, F.S.D. Medina, Effects of flavonoids and other polyphenols on inflammation, Crit. Rev. Food Sci. Nutr. 51 (4) (2011) 331–362.

[77] J.R. Soberón, M.A. Sgariglia, D.A. Sampietro, E.N. Quiroga, M.A. Vattuone, Free radical scavenging activities and inhibition of inflammatory enzymes of phenolics isolated from Tripodanthus acutifolius, J. Ethnopharmacol. 130 (2) (2010) 329–333.

[78] Y. Noreen, G. Serrano, P. Perera, L. Bohlin, Flavan-3-ols isolated from some medicinal plants inhibiting COX-1 and COX-2 catalysed prostaglandin biosynthesis, Planta Med. 64 (6) (1998) 520–524.

[79] S. Gibbons, Bioactive compounds from natural sources isolation, characterization and biological properties, Phytother Res. 16 (6) (2002), 600–600.