Aspects of Molecular Medicine 5 (2025) 100085

Contents lists available at ScienceDirect

Aspects of Molecular Medicine

journal homepage: www.journals.elsevier.com/aspects-of-molecular-medicine

Identification and exploration of novel FGFR-1 inhibitors in the Lotus database for Cholangiocarcinoma (CCA) treatment

Samuel Aduramurewa Osunnaya [a,b,*], Wilberforce K. Ndarawit [a,c,d], Ifeoluwa Aderibigbe [e], Ibilola A. Omolopo [a,b], Precious O. Aribisala [f], Ayodele Oluwasegun Elekan [a,g], Adeola Sakirat Adeyemo [a,h], Sheriffdeen Abiola Amoo [a,g], Olatunde Simbiat Olamiposi [a,g], Njogu M. Kimani [c,d], Taiwo Hamidat Olaide [a,h], Adedoyin John-Joy Owolade [a,i], Damilola Samuel Bodun [a,j]

[a] ChemoInformatics Academy, Nigeria
[b] Nigerian Institute of Medical Research, Yaba, Nigeria
[c] Department of Physical Science, University of Embu, P.O BOX 6, 60100, Embu, Kenya
[d] Natural Products Chemistry and Computational Drug Discovery Laboratory, P.O Box 6, 60100, Embu, Kenya
[e] Department of Biochemistry, Federal University Oye-Ekiti, Nigeria
[f] Eureka Laboratory, Babcock University, Nigeria
[g] Department of Biochemistry, Adekunle Ajasin University Akungba, Akoko, Nigeria
[h] Department of Chemistry, Federal University of Technology, Akure, Nigeria
[i] Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Obafemi Awolowo University, Ile Ife, Osun State, Nigeria
[j] Covenant University Bioinformatics Research (CUBRe), Covenant University, Nigeria

ARTICLE INFO

Handling Editor: Prof A Angelo Azzi

Keywords:
Cholangiocarcinoma
LOTUS database
Machine learning
Classification
MD simulation

ABSTRACT

Cholangiocarcinoma (CCA) is a rare but aggressive cancer affecting the bile duct, with limited treatment options and a poor prognosis. This study employed a machine learning algorithm and molecular docking using Maestro to screen 215,925 compounds from the Lotus database, aiming to identify potential fibroblast growth factor receptor-1 (FGFR1) inhibitors as therapeutic agents. Five promising compounds were identified, with binding energies ranging from −10.018 to −8.439 kcal/mol, all outperforming the standard drug Dovitinib (−8.419 kcal/mol). Molecular mechanics calculations and MM/GBSA analysis confirmed the structural stability and favorable binding energies of the protein-ligand complexes. Additionally, 100-ns molecular dynamic simulations demonstrated that the top three compounds remained stable within FGFR1's active site, supported by root mean square deviation, root mean square fluctuation, and hydrogen bond interactions. Overall, these five compounds show promise as potential therapeutic agents for CCA and warrant further investigation for drug development.

1. Introduction

Over the last decade, intensive advancements in cancer genomic research have provided a foundation for the use of specific small molecules to target disrupted cellular processes. The dysregulation of fibroblast growth factor receptor (FGFR) signaling pathways as observed in a subset of various cancers, makes it a highly promising therapeutic target as evident in diverse pre-clinical and clinical studies (Chmiel et al., 2022).

The FGFR family constitute of four tyrosine kinase receptors, (FGFR 1–4). These enzymes are essential in carcinogenesis and several other physiological signaling pathways. During the early stages of embryonic development FGFRs are primarily involved in essential cellular interaction and functions (Ornitz and Legeai-Mallet, 2017). Moreover, these signaling pathways are involved in the regulation of fundamental metabolic functions such as bile acid, fatty acid, glucose, and mineral metabolism (Zhou et al., 2017). For example; activation of FGFR1 has been demonstrated to affect vascular endothelium proliferation positively (Cross and Claesson-Welsh, 2001). Therefore, activation of these receptors could likely initiate downstream signaling through key cellular pathways such as PI3K, AKT, mTOR, RAS/RAF/MEK/MAPK, JAK/STAT, and PLCγ, all of which play critical roles in tumor

* Corresponding author. Nigerian Institute of Medical Research, Yaba, Nigeria.
E-mail address: osunnayasammy@gmail.com (S.A. Osunnaya).

https://doi.org/10.1016/j.amolm.2025.100085
Received 25 February 2025; Received in revised form 6 April 2025; Accepted 23 April 2025
Available online 26 April 2025

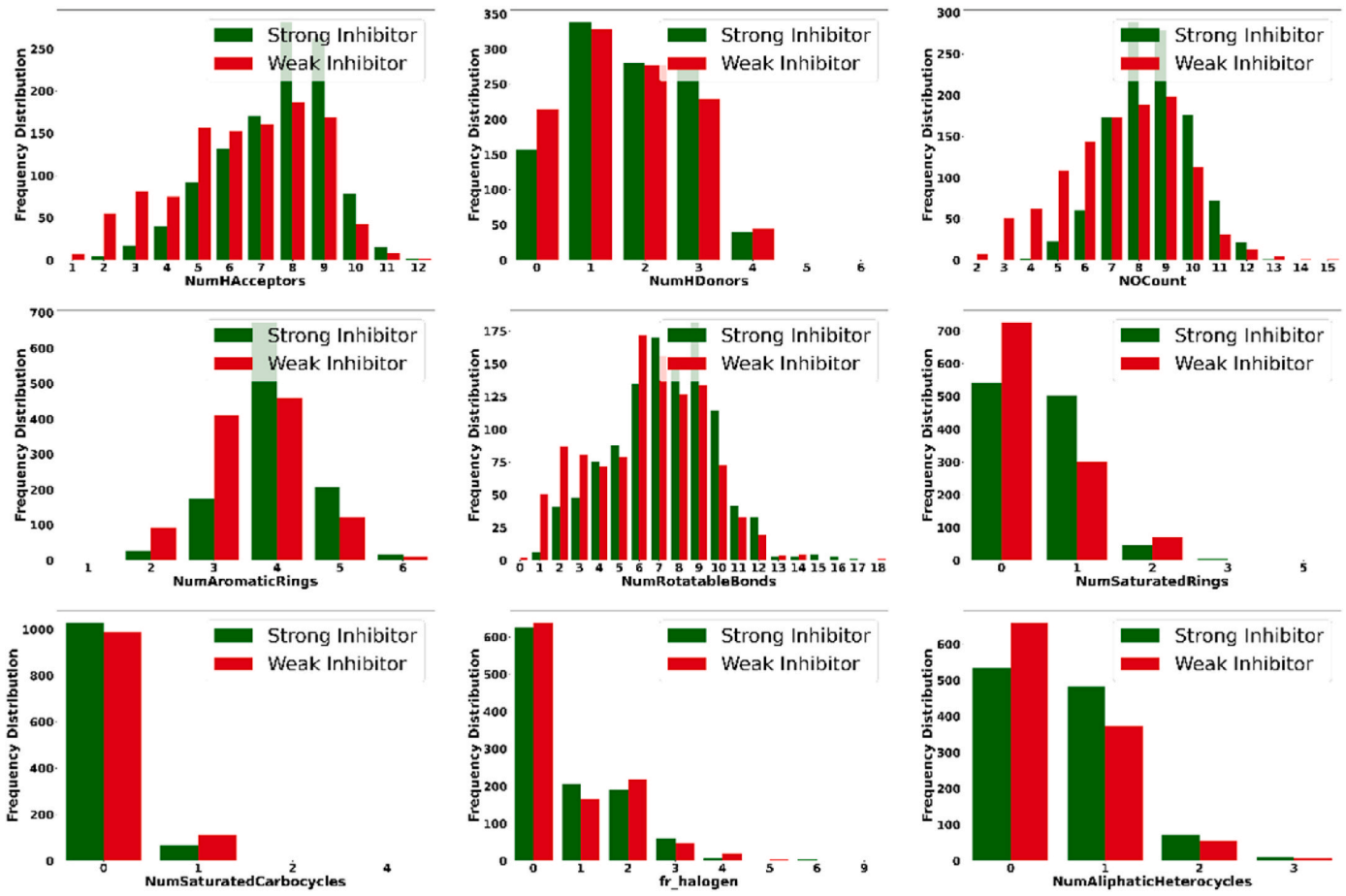

Fig. 1a. Bar charts of features counts of FGFR-1 inhibitors from CHEMBL database.

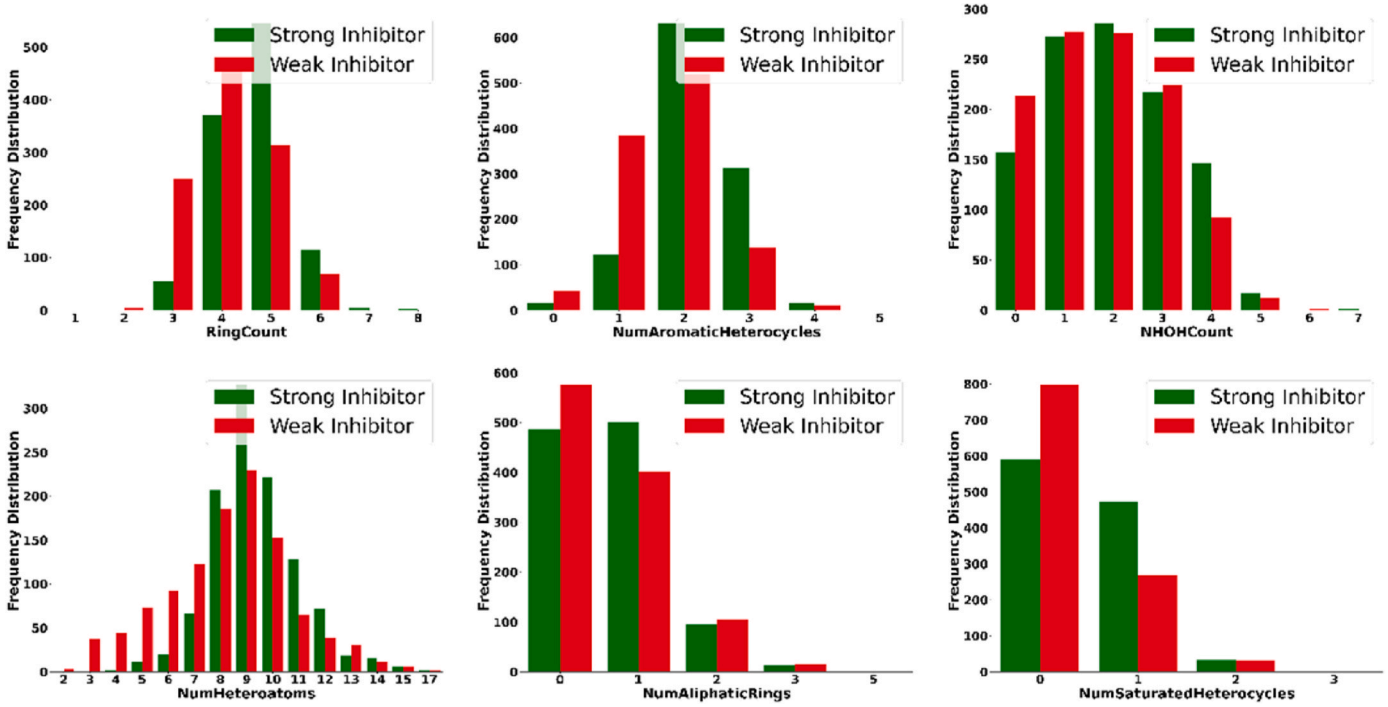

Fig. 1b. Bar charts of features counts of FGFR-1 inhibitors from CHEMBL database.

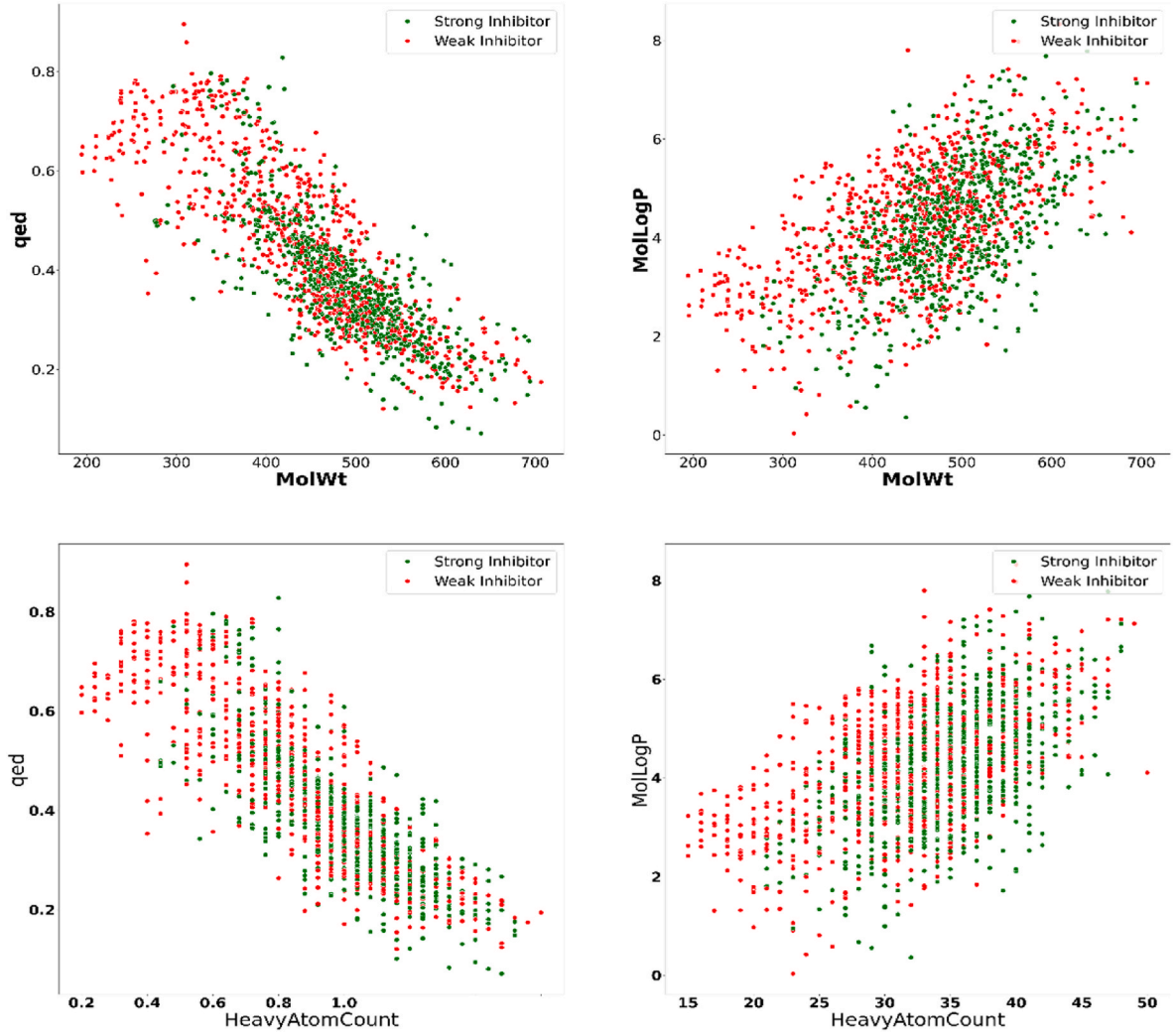

Fig. 2. Scatter plot of different features of FGFR-1 inhibitors from CHEMBL database.

Table 1
Best 10 descriptors importance.

Features	Scores
SMR_VSA10	0.185727718
TPSA	0.171837253
BCUT2D_MRHI	0.167809861
EState_VSA8	0.156903893
SlogP_VSA2	0.147997215
PEOE_VSA1	0.139312067
SMR_VSA7	0.133327107
PEOE_VSA8	0.132902051
BCUT2D_MWHI	0.126873558
SMR_VSA3	0.12616125

development, including mitogenesis via the MAPK pathway, cell survival via the PI3K pathway, and mobility via the PKC pathway, and these pathways are also present in Cholangiocarcinoma (CCA) (Goetz and Mohammadi, 2013; Teven et al., 2014; Turner and Grose, 2010).

Multiple studies regarding various forms of cancer demonstrated that there is a close association between FGFRs and tumor growth as well as tumor cell proliferation (Kim et al., 2013). For example, a study by Kim et al. (2013) demonstrated that FGFR2 promotes breast cancer tumorigenicity by maintaining breast tumor-initiating cells. Moreover, it is also possible that the FGFR1 signaling pathway plays a key role in tumor cell invasion (Coleman et al., 2014). Therefore, dysregulation of the FGFR pathways can also impact angiogenesis, a crucial stage in the development of cancer. Additionally, Wang et al. (2021) demonstrated that growth of tumors can be prevented by blocking angiogenesis through FGFR1 inhibition. In summary, gain-of-function coding mutations or gene amplifications have the potential to significantly impact each critical stage of cancer formation, ultimately resulting in dysregulation of the FGF/FGFR signaling pathway.

Consequently, new treatment avenues have emerged, such as the use of FGFR inhibitors in tumors like CCA, which have a poor prognosis and few available treatment options (Banales et al., 2016). Cholangiocarcinoma (CCA) is a rare type of cancer that originates in the bile ducts, which are responsible for carrying bile from the liver to the small intestine (Ghouri et al., 2015). CCA is frequently diagnosed at an advanced stage, mostly in older males, with a peak incidence age of 70 (Izquierdo-Sanchez et al., 2022). The incidence rate is higher in Eastern countries (e.g., Thailand) compared to Western countries, with risk factors such as cirrhosis and cholelithiasis. However, new potential risk factors are still being discovered (Khan et al., 2019). However, due to its rare occurrence, CCA continues to pose diagnostic and therapeutic challenges as the expected 5-year survival rate for CCA patients still oscillates at around 5 % (Sung et al., 2021). According to Chmiel et al. (2022), management of CCA currently relies on surgical procedures and chemotherapy which also have limited effectiveness, necessitating the need for finding an alternative therapeutic agent. However conventional methods of drug discovery through clinical trials have been overtaken

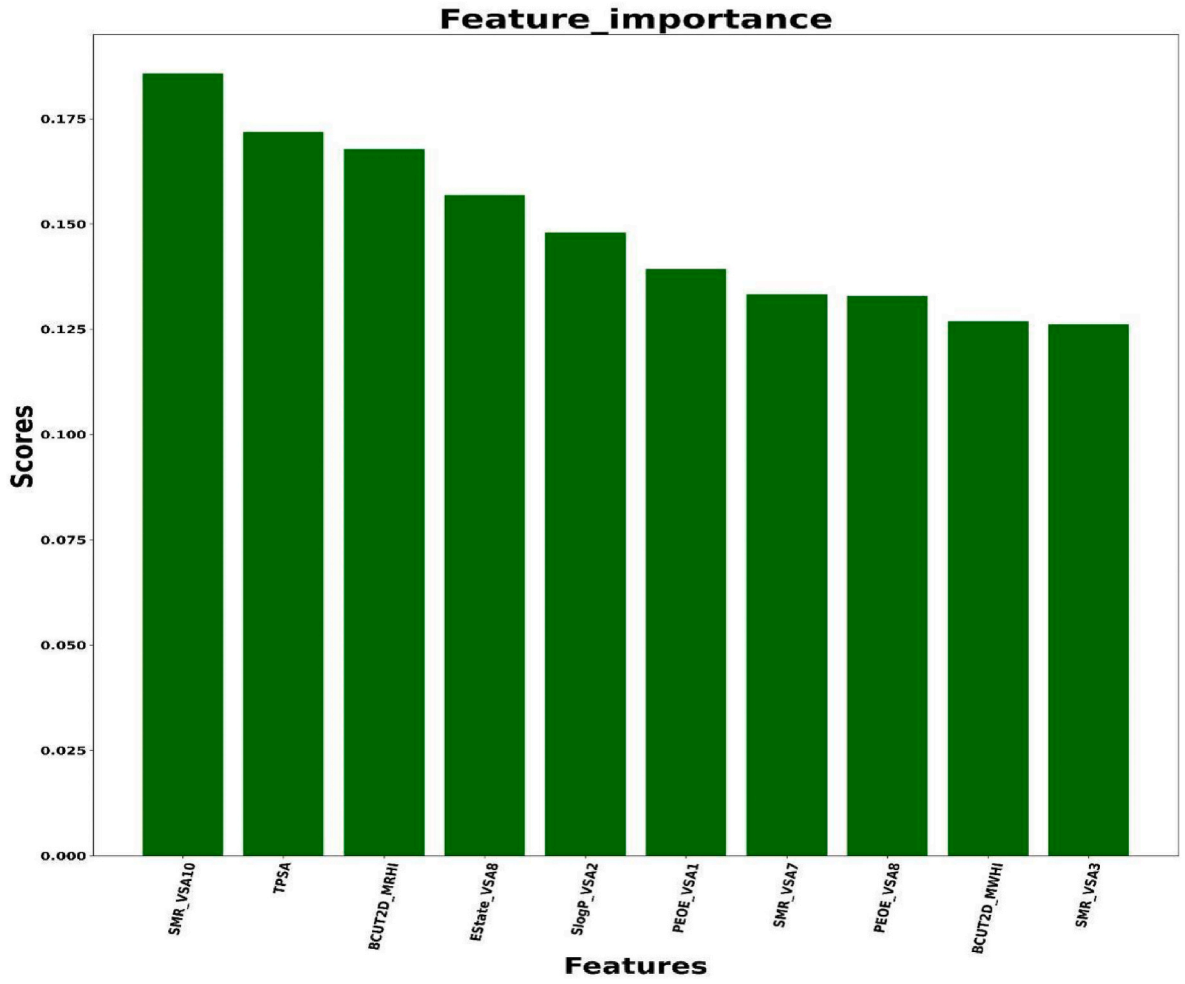

Fig. 3. Barplot of the best 10 features importance.

by events due to cost and the trial-and-error steps involved. It is estimated that a typical drug discovery cycle, from lead identification through to clinical trial can roughly take fourteen years with a cost of 800 million US dollars (Link, 2019). Such limitation brings into play the recent computational approach such as Machine Learning (ML) and Deep Learning (DL) which predicts new therapeutic agents using computer algorithms.

The development of machine learning (ML) models and the expansion of chemical and pharmacological data have given rise to AL paradigms (Vamathevan et al., 2019). This era of technology has created a space in the field of drug discovery for data-driven computer processes (Pasrija et al., 2022). As an offshoot of AI, ML-facilitated approaches place more emphasis on the transformation of massive biomedical big data into new enlightening, and sustainable expertise than conventional approaches do on the theoretical advancement of the complex and sustainable physicochemical tenets (Pasrija et al., 2022). Deep learning (DL) another tenet of AI allows its models to process information and make judgments or predictions without the need for explicit programming (Lavecchia, 2019). DL plays a pivotal role in drug development as it can analyze large datasets including both genetic and clinical data, hence helping in the identification of novel drug targets, with precise drug efficacy prediction, and drug optimization (Nag et al., 2022). Its ability to evaluate big and complicated datasets is one of its unique features. In contrast to labor-intensive and time-consuming traditional data analysis approaches such as statistical techniques and manual examination, deep learning models provide quick and efficient data analysis that may identify patterns and foresee outcomes, thereby expediting the drug development process (Nag et al., 2022).

Our study focuses on the identification of novel and natural FGFR-1 Inhibitors from the Lotus Database for CCA Treatment. Lotus database is an extensive natural product database designed to facilitate the exploration and discovery of bioactive chemical compounds (Rutz et al., 2022). The database comprises an array of more than 215,925 chemical entities derived from a variety of natural sources, such as fungi, plants, marine organisms, and microbes (Rutz et al., 2022). In this study, we employed a machine learning algorithm to train on active molecule datasets. Subsequently, we screened the Lotus database to identify potential FGFR-1 inhibitors. We further assessed the selected molecules using molecular docking techniques and molecular dynamic simulations.

2. Materials and methods

The workflow of this study is depicted in Supplementary Fig. 1.

2.1. Generation of FGFR1 inhibitors from CHEMBL

The experimental data on the activity of FGFR1 inhibitors ((n = 4896) were obtained from the CHEMBL database. The target protein's (5AM6) FASTA sequence was retrieved from the PDB and used to search for the antagonists in the CHEMBL database (https://www.ebi.ac.uk/chembl/). The list of antagonists was then preprocessed to contain just the Chembl ID of the compound, SMILES of the compound, and the pIC50 value in a CSV file.

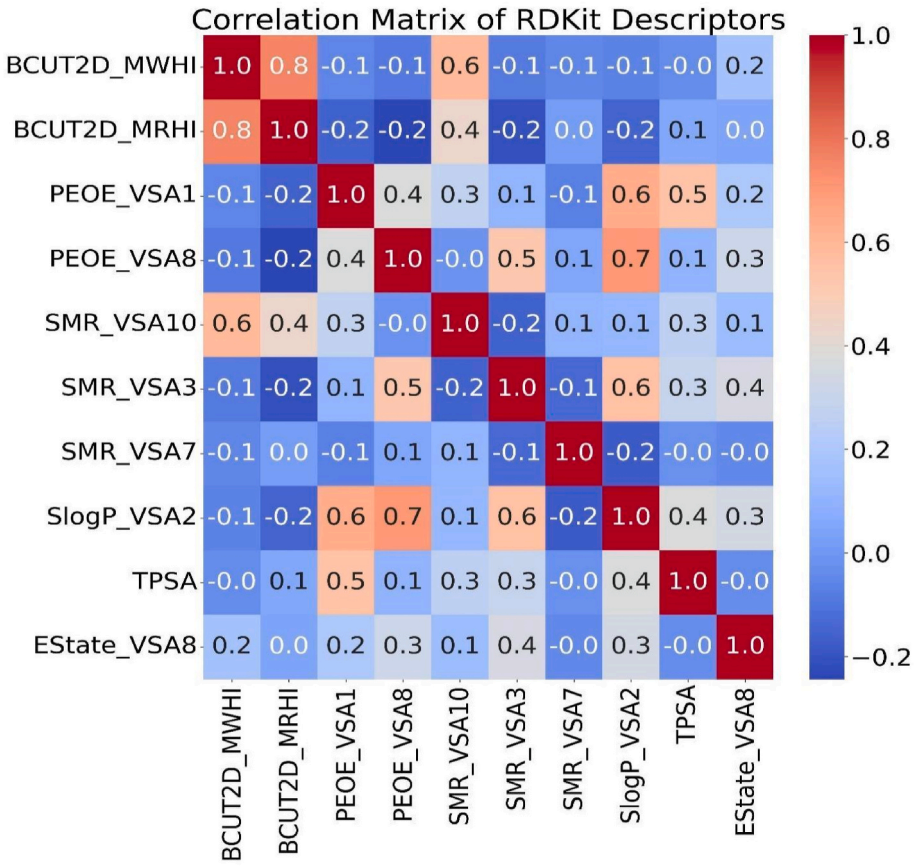

Fig. 4. Correlation matrix of the selected features (best 10).

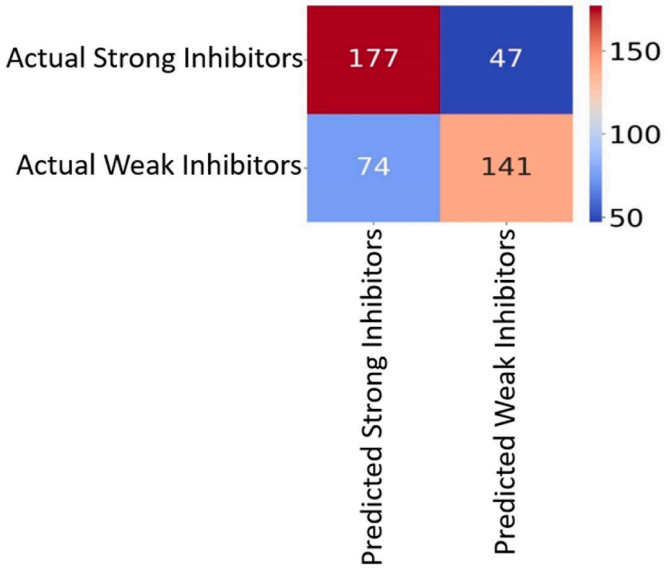

Fig. 5. Confusion matrix of the test data prediction.

2.2. Target library

The target library data set contains 215,925 compounds that were downloaded from the LOTUS database (https://lotus.naturalproducts.net/). The LOTUS database is one of the biggest resources for natural products. The database was downloaded in a SDF format.

Table 2
Classification report of the model classifier on the test data.

classification	precision	Recall	F1-score	support
Strong Inhibitor	0.71	0.79	0.75	224
Weak Inhibitor	0.75	0.66	0.70	215
accuracy			0.72	439
macro avg	0.73	0.72	0.72	439
weighted avg	0.73	0.72	0.72	439

2.3. Machine learning architecture

The FGFR1 inhibitors dataset was cleaned using pandas. Compounds with missing activity data were removed, and duplicates were also removed. The compounds were categorized into weak and strong inhibitors using a 6.5 pChEMBL value as the threshold (i.e., pChEMBL values greater than or equal to 6.5 were classified as strong inhibitors, while those less than 6.5 were classified as weak inhibitors). The data was balanced by randomly removing compounds to make the number of weak inhibitors equal to the number of strong inhibitors, resulting in a total of 2194 compounds. 210 descriptors were generated using the RDKit package. Exploratory data analysis was conducted using the Seaborn and Matplotlib packages. The final descriptors for the model were selected by: [1] Removing features with high correlation (>0.90) [2] Removing features with low variance.

The remaining features were transformed using the MinMaxScaler based on the fitness from the training set after dividing the entire dataset into training and test sets using a test size ratio of 0.2. Using the KBest algorithm (a traditional feature selection method) in the Scikit-learn library, the top 10 performing features were selected.

The machine learning model was built using the

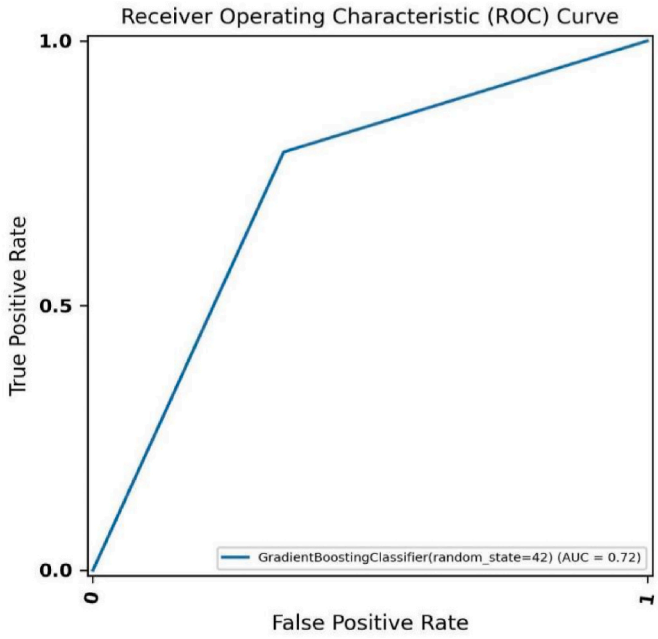

Fig. 6. Roc curve of the model classifier on the test data.

GradientBoostingClassifier, an ensemble technique that can operate with a small amount of data. The model was used to screen the LOTUS database of 215,925 compounds, classifying the database into 147,310 weak inhibitors and 68,508 strong inhibitors.

2.4. Virtual screening by Lipinski rule of 5, QED, and SAScore

The resulting 68,508 strong inhibitors were further screened using Lipinski's rule of five (ROV = 0), which resulted in 34,738 compounds. The compounds were further screened using the QED score (threshold 0.6) and SA score (threshold 5.0), resulting in 19,417 compounds and 14,500 compounds, respectively. Both the QED and SA scores were calculated using the RDKit library. The SAScorer was used for the SA score calculation as this is not part of the CalculateDescriptors class.

2.4.1. Ligand preparation

The filtered compounds (14500) were prepared using the LigPrep module, employing the OPLS3 force field (Bodun et al., 2024). Tautomer generation was excluded, while stereoisomer calculation was limited to a single isomer per ligand.

2.4.2. Protein preparation

The crystal structure of the FGFR-1 was obtained from the Protein Data Bank (PDB) with PDB ID 5AM6. The process started by importing the protein into the workspace, followed by Prime-facilitated filling of missing loops and chains. Subsequently, water molecules were eliminated during chain refinement. PROPKA pH 7.5 was applied to optimize the protein, followed by removal of water beyond a 5 Å radius and protein minimization. Finally, an energy minimization process was executed to achieve a lower-energy state for the protein.

2.5. Grid generation

The active site of the FGFR-1 was determined using the receptor grid generation module in Schrödinger Maestro software. The coordinates of the active site were calculated based on the position of co-crystallized

Table 3
Top compounds and their fitness scores.

Compound	Lotus ID	Fitness score
(S)-2-(2-oxo-1,2-dihydroquinoline-4-carboxamido) succinic acid	73350744	1.615
(2R)-4-methoxy-4-oxo-2-[[2-[(3S)-2-oxo-1,3-dihydroindol-3-yl]acetyl]amino]butanoic acid	162848022	1.952
2-demethylcyclopiamide E	137206669	1.826
(4bR,5R,10aS)-11,11-dimethyl-4b,5,10,10a-tetrahydrobenzo[*b*]fluorene-5,7,9-triol	102410411	1.345
Chandrananimycin D	46939591	1.359

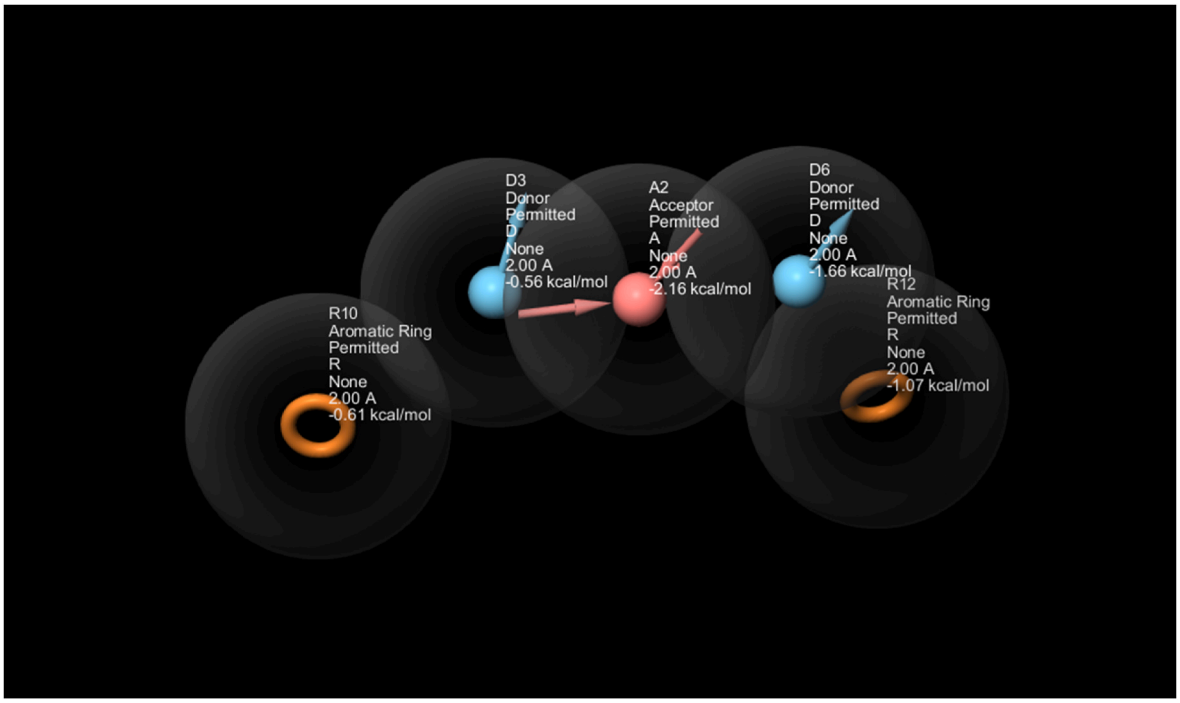

Fig. 7. The E-pharmacophore hypothesis developed with five features (2 Donors, 1 Acceptor and 2 Aromatic Rings).

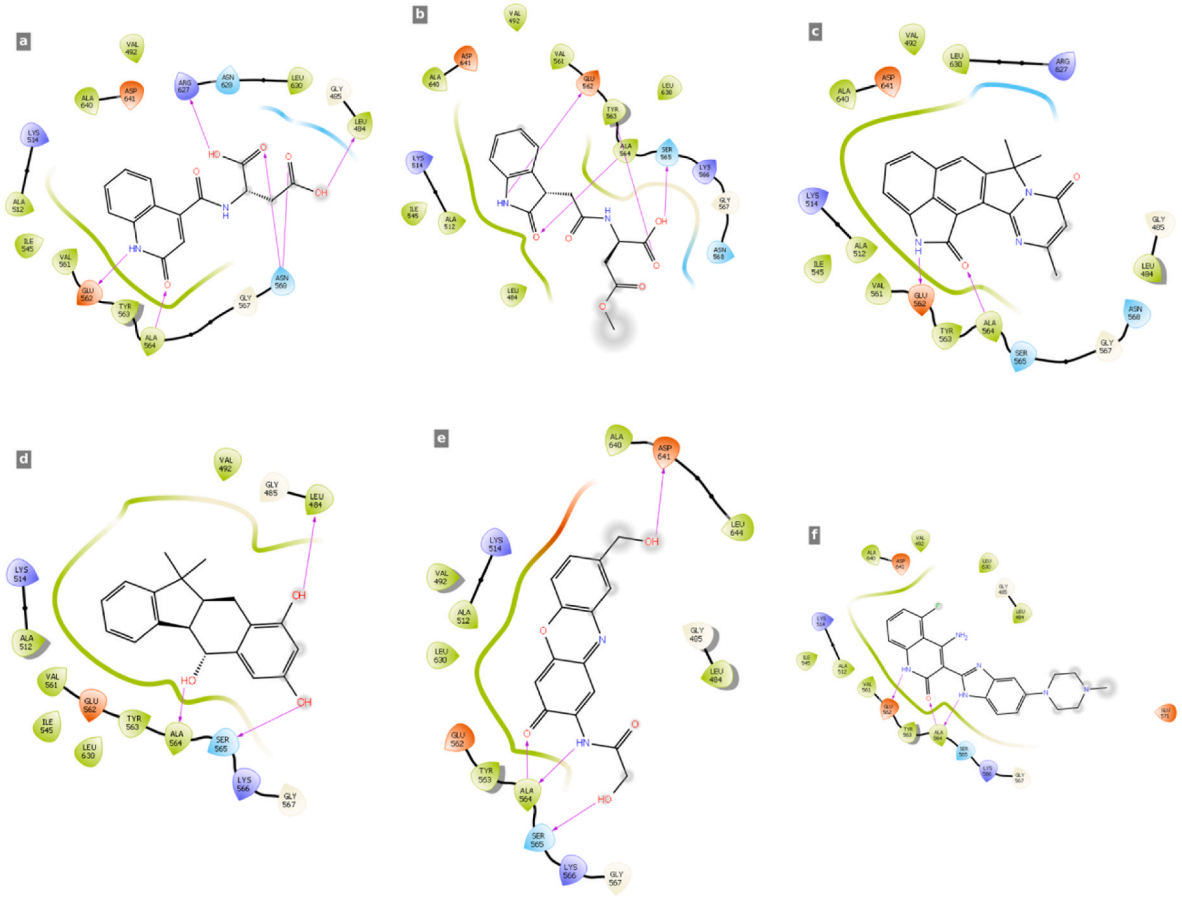

Fig. 8a. 2D interaction of the top compounds with the FGFR-1 protein.
a – Compound 73350744, b – Compound 162848022, c – Compound 137206669, d – Compound 102410411, e – Compound 46939591, f – Standard drug (Dovitinib).

ligand that was present within the FGFR-1 binding pocket. The x, y, and z grid coordinates were measured to be 217.92, −8.14, and 24.83 respectively.

2.6. E-pharmacophore generation and screening

Pharmacophore screening was conducted using Schrödinger Maestro software version 12.8. A pharmacophore model was created based on the interactions between a co-crystallized ligand and the target protein. This model featured five distinct characteristics. Strict screening criteria (requiring at least 4 out of 5 features) were applied to filter 14,500 compounds, resulting in 1937 compounds passing the initial screening.

2.7. Structure-based virtual screening

Out of the initial 1937 compounds that passed the pharmacophore screening, further screening was performed using molecular docking methods with different precision levels. Initially, HTVS (High Throughput Virtual Screening) was applied, followed by filtering based on a docking threshold of −6.0 kcal/mol, resulting in 1770 compounds.

Subsequently, Standard Precision (SP) docking was employed with a stricter threshold score of −6.5 kcal/mol, reducing the list to 227 compounds. Finally, Extra Precision (XP) docking was used on the top 50 compounds from SP, leading to the selection of the top five compounds.

Throughout this process, the binding affinity of each compound was evaluated against the co-crystallized ligand, 4-amino-5-fluoro-3-[5-(4-methylpiperazin-1-yl)-1H-benzimidazol-2-yl]quinolin-2(1H)-one also known as Dovitinib.

2.8. Estimation of binding energy

Following XP docking, which demonstrated a positive correlation between favorable poses and high scores among the docked complexes, the next step involved evaluating the potential biological response of these complexes. This assessment was conducted by calculating the free binding energy using the molecular mechanics generalized Born surface area (MM-GBSA) method within the Prime module of Maestro. The study focused on the top five ranked compounds, including the standard drug, to ascertain the stability of the protein-ligand complexes. This analysis aimed to determine whether the interactions between the ligands and the target protein were sufficiently strong to potentially induce a biological response (Bodun et al., 2023).

$$\Delta Gbind = \Delta Gcomplex - (\Delta Gligand + \Delta Gprotein) \quad (1)$$

2.9. Evaluation of DrugLikeness and ADMET properties

The ADMET properties of the top five compounds were evaluated using the SwissADME (http://www.swissadme.ch/) and ADMETLab (https://admetlab3.scbdd.com/). These processes—absorption, distribution, metabolism, and excretion—are critical in drug development for understanding how drugs are absorbed into the body, distributed within tissues, metabolized by enzymes, and ultimately eliminated. These web servers provide predictions for various ADMET parameters, which are essential for assessing the potential effectiveness and safety of drug candidates during the early stages of development.

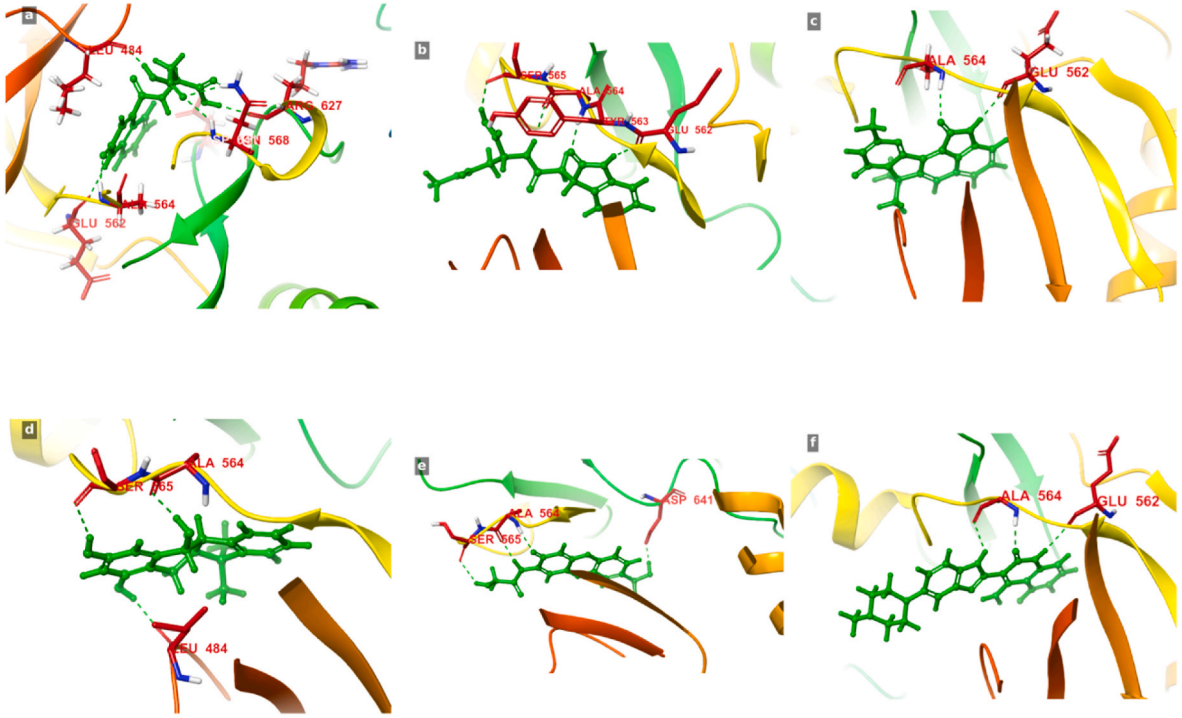

Fig. 8b. 3D interaction of the top compounds with the FGFR-1 protein
a – Compound 73350744, b – Compound 162848022, c – Compound 137206669, d – Compound 102410411, e – Compound 46939591, f – Standard drug (Dovitinib).

Table 4
Docking scores and the free energy of the top compounds.

Compound	Docking score	MMGBSA score
(S)-2-(2-oxo-1,2-dihydroquinoline-4-carboxamido) succinic acid (Compound 73350744)	−10.018	−59.10
(2R)-4-methoxy-4-oxo-2-[[2-[(3S)-2-oxo-1,3-dihydroindol-3-yl]acetyl]amino]butanoic acid (Compound 162848022)	−0.9.981	−51.80
2-demethylcyclopiamide E (Compound 137206669)	−8.851	−47.61
(4bR,5R,10aS)-11,11-dimethyl-4b,5,10,10a-tetrahydrobenzo[*b*]fluorene-5,7,9-triol (Compound 102410411)	−8.442	−45.14
Chandrananimycin D (Compound 46939591)	−8.439	−45.67
Standard drug (Dovitinib)	−8.419	−46.28

2.10. Molecular dynamic simulation

The top three compounds based on docking scores with the co-crystalized ligand were subjected to a 100 ns molecular dynamics (MD) simulation using the Desmond tool of the Schrodinger (version 2017-1). The simulation was carried out with an explicit solvent system, and to maintain neutrality, Na+, and Cl-were added at 300^{O}K and 1.01325 bar. An orthorhombic box with a 10 × 10 × 10 Å buffer and 0.15 M salt was used to prepare the model for the physiological environment, with the transferable intermolecular potential-4 point (TIP4P) force field. After the MD simulation, the trajectories were analyzed for RMSF, RMSD, and protein-ligand contacts using the interaction diagram module of the Desmond-Schrodinger suite.

3. Results and discussion

3.1. Exploratory data analysis

Fig. 1 shows the histogram plots of some selected descriptors. Fig. 2a shows a scatterplot of the Quantitative Estimate of Drug-likeness (QED) and Molecular Weight (Mol Wt). There was a negative correlation between these two descriptors: the higher the molecular weight, the lower the QED. We also noticed that strong inhibitors formed a noticeable cluster with low QED values. This was also the case in Fig. 2c, where the

Table 5
Interacting amino acids at the active site with the lead compounds.

Compound name	Docking score	H-bond	Interacting residue	Hydrophobic interaction	Other interaction
Compound 73350744	−10.018	6	ARG 627, ASN 568, GLU 562, LEU 484, ALA 564	LEU 484, ALA 640, ALA 512, ILE 545, VAL 561, TYR 563, ALA 564, VAL 492, LEU 630	None
Compound 162848022	−9.981	4	GLU 562, TYR 563, ALA 564, SER 565	ALA 640, ILE 545, ALA 512, VAL 561, TYR 563, ALA 564, VAL 492, LEU 630, LEU 484	None
Compound 137206669	−8.851	2	GLU 562, ALA 564	VAL 492, LEU 630, ALA 640, ILE 545, ALA 512, VAL 561, TYR 563, ALA 564, LEU 484	None
Compound 102410411	−8.442	3	LEU 484, ALA 564, SER 565	ILE 565, ALA 512, VAL 561, TYR 563, LEU 630, ALA 564, LEU 484, VAL 492	None
Compound 46939591	−8.439	4	ASP 641, ALA 564, SER 565	LEU 644, ALA 640, VAL 492, ALA 512, LEU 630, TYR 563, ALA 564, LEU 484	None
STANDARD DRUG (Dovitinib)	−8.419	3	GLU 562, ALA 564	ALA 640, ILE 545, VAL 492, ALA 512, VAL 561, TYR 563, ALA 564, LEU 630, LEU 484	None

Table 6
In-silico drug likeness prediction of the top compounds.

Pubchem ID	Compound name	MW	HBA	HBD	TPSA	ILOGP	LOGKP	ROV
Compound 73350744	(S)-2-(2-oxo-1,2-dihydroquinoline-4-carboxamido)succinic acid	304.25	6	4	136.56	0.5	−8.54	0
Compound 162848022	(2R)-4-methoxy-4-oxo-2-[[2-[(3S)-2-oxo-1,3-dihydroindol-3-yl]acetyl]amino] butanoic acid	320.3	6	4	128.72	1.19	−7.76	0
Compound 137206669	2-demethylcyclopiamide E	317.34	4	0	62.63	2.44	−8.27	0
Compound 102410411	(4bR,5R,10aS)-11,11-dimethyl-4b,5,10,10a-tetrahydrobenzo[*b*]fluorene-5,7,9-triol	296.36	3	3	60.69	2.2	−5.76	0
Compound 46939591	Chandrananimycin D	300.27	5	4	107.89	1.04	−7.49	0

Table 7
The bio-availability, pharmacokinetic properties and Cytochrome P450 metabolizing enzymes inhibitory potentials of the top compounds.

Models	(S)-2-(2-oxo-1,2-dihydroquinoline-4-carboxamido)succinic acid	(2R)-4-methoxy-4-oxo-2-[[2-[(3S)-2-oxo-1,3-dihydroindol-3-yl]acetyl] amino]butanoic acid	2-demethylcyclopiamide E	(4bR,5R,10aS)-11,11-dimethyl-4b,5,10,10a-tetrahydrobenzo[*b*]fluorene-5,7,9-triol	Chandrananimycin D	Standard drug
Bioavaliability Score	0.56	0.56	0.55	0.55	0.55	0.55
CYP1A2 inhibitor	No	No	Yes	Yes	No	Yes
CYP2C19 inhibitor	No	No	No	No	No	No
CYP2C9 inhibitor	No	No	No	No	No	No
CYP2D6 inhibitor	No	No	No	Yes	No	Yes
CYP3A4 inhibitor	No	No	No	No	No	No
GI Absorption	High	High	High	High	High	High
BBB permeant	No	No	Yes	Yes	No	
Pgp substrate	No	Yes	No	Yes	Yes	Yes
Mutagenicity	No	No	No	No	Yes	Yes
Carcinogenicity	No	No	No	No	No	No
Hepatotoxicity	No	No	No	No	No	No
Cytotoxicity	No	No	No	No	No	No

heavy atom count of the molecules was compared to their QED values. In Fig. 2b, there was a positive correlation between the two descriptors, MolLogP (which measures the lipophilicity of a compound) and Mol Wt. The higher the weight of a compound, the more probable it is that it can cross lipid layers.

3.2. ML model

Prior to model building, the top 10 features were selected using the KBest algorithm from the Scikit-learn library. The features selected include SMR_VSA10, TPSA, BCUT2D_MRHI, EState_VSA8, SlogP_VSA2, PEOE_VSA1, SMR_VSA7, PEOE_VSA8, BCUT2D_MWHI, and SMR_VSA3. SMR_VSA10 represents the sum of van der Waals surface areas (VSA) for atoms with a specific contribution range to molecular refractivity (SMR). It highlights the portion of the molecule contributing to refractivity within a defined range. TPSA (Topological Polar Surface Area) calculates the sum of the surface areas of polar atoms (usually oxygen and nitrogen, including their attached hydrogen atoms). It is useful for predicting a molecule's ability to permeate cells and cross biological membranes, making it important in drug discovery. BCUT2D_MRHI refers to the high eigenvalue related to molecular refractivity in the BCUT descriptor family, which is based on the eigenvalues of a modified connectivity matrix. This provides insights into the electronic properties and diversity of a molecule. EState_VSA8 sums the van der Waals surface areas for atoms with a specific EState (electronic state) index range, relating the VSA to the electronic environment of the atoms. SlogP_VSA2 relates the VSA of atoms to their logP contributions (partition coefficient between octanol and water. SlogP_VSA2 corresponds to atoms within a specific logP contribution range.

PEOE_VSA1 is derived from Partial Equalization of Orbital Electronegativities (PEOE) and sums the VSA for atoms with specific partial charge ranges. PEOE_VSA1 corresponds to atoms with a specific range of partial charges. SMR_VSA7 is similar to SMR_VSA10, summing the VSA for atoms with a different specific contribution range to molecular refractivity (SMR), emphasizing another aspect of refractivity contribution. PEOE_VSA8, another PEOE_VSA descriptor, sums the VSA for atoms with partial charges in a specific range, focusing on a different partial charge range compared to PEOE_VSA1. BCUT2D_MWHI is a BCUT descriptor related to the high eigenvalue associated with molecular weight (MW) (Zhang et al., 2023). SMR_VSA3 sums the VSA for atoms with a specific SMR contribution range, highlighting a different aspect of the molecular refractivity contribution compared to SMR_VSA10 and SMR_VSA7.

Table 1 presents the scores of the 10 most important features. The feature with the highest score was reported to be SMR_VSA10 (0.185727718), while the feature with the least score was SMR_VSA3 (0.12616125). The higher the value, the better it contributes to building an accurate model. Fig. 3 Shows a bar plot representing the best 10 features of importance with the scores on the y-axis and features on the x-axis, ranging from the highest features importance (left) to the lowest (right) on the plot.

Fig. 4 indicates the relationships between the top 10 selected features for the model. Each cell in the matrix shows the correlation coefficient (similarity) between the features. A lower value indicates better feature independence, which is crucial to build a good model for classification. On the other hand, a higher value indicates high feature dependence. The order of feature independence was represented with color range from blue to red (i.e the closer to blue, the better degree of independence). As observed from the correlation matrix, almost all the features are highly independent which indicated that the features contributed

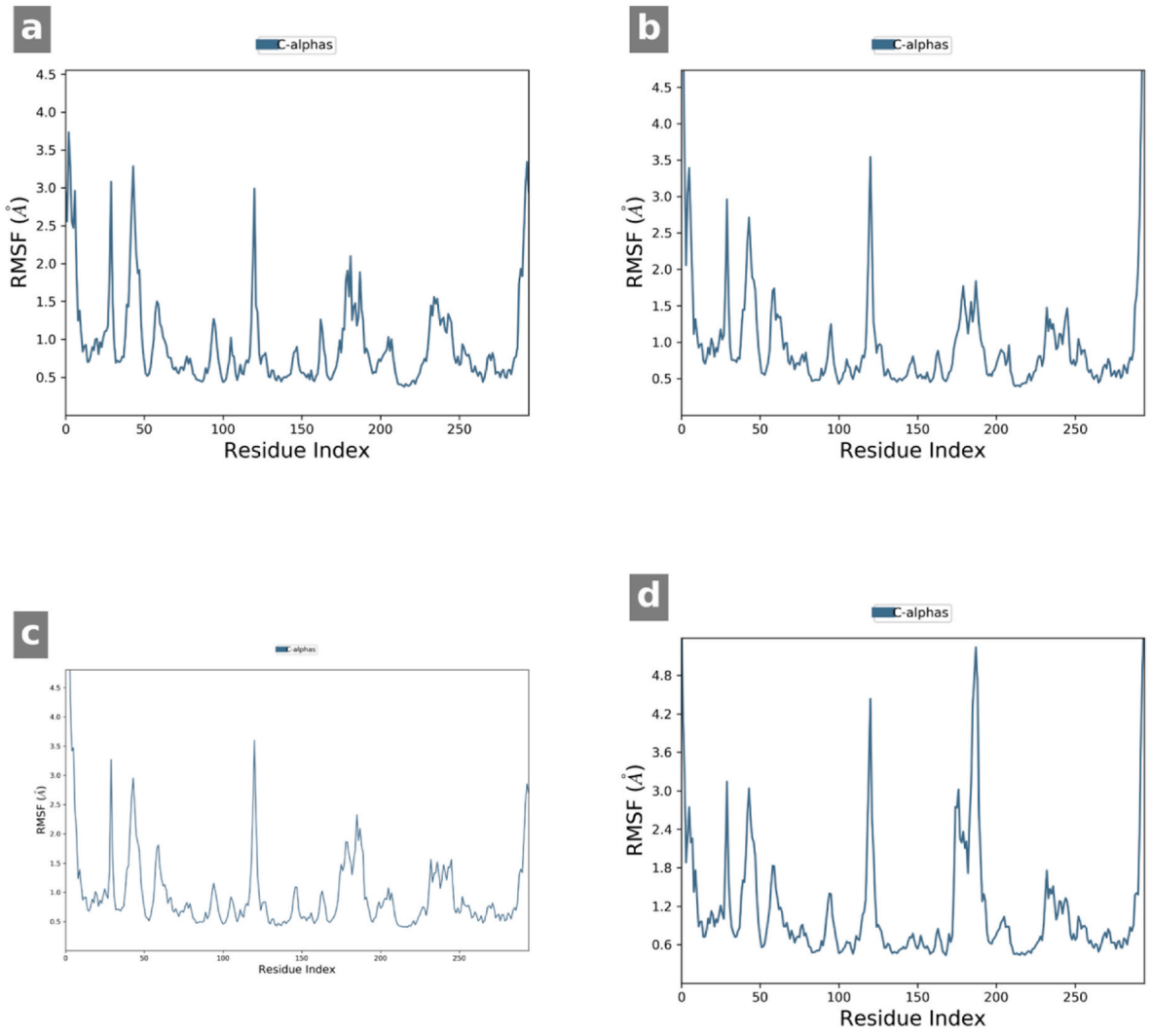

Fig. 9a. Fig. 10 The Root Mean Square Fluctuation (RMSF) line plot used for characterizing local changes along the protein chain. A shows the RMSF values for the 73350744 complex with FGFR1, while B displays the RMSF values for the 162848022 complex with FGFR1. C presents the RMSF values for the 137206669 complex with FGFR1. Lastly, D shows the RMSF values of the reference ligand Dovitinib in complex with FGFR1.

independently to the model's performance.

The confusion matrix of the test data prediction is shown in Fig. 5. The matrix shows the summary of the actual results and the predicted results. The model correctly predicted 177 actual negative results as negative and 47 were misclassified as positives while 141 actual positives were correctly predicted as positives and 74 were misclassified as negative. This aligns with the precision and recall values reported in Table 2, where the classification report of the model classifier on the test data was presented including precision, recall, and F1-scores. As seen in Table 2, the model performed reasonably well in predicting the weak and strong inhibitors with a recall and precision rate ranging from 0.66 to 0.79.

Receiver Operating Characteristic Curve (ROC curve) is an important tool for evaluating the performance of machine learning models. ROC curves demonstrate the correlation between true positive rates (sensitivity) for a given model and false positive rates (1-specificity). Fig. 6 shows the ROC curve for the model for this study. The area under the curve (AUC) quantifies the overall ability of the model to distinguish between the two classes. As shown in Fig. 6, the model performed reasonably well in distinguishing between the strong and weak inhibitors in the test data.

All the evaluating parameters employed in assessing the model's performance (classification report, confusion matrix and ROC curve) confirmed that our model performed well on the test dataset.

3.3. *Pharmacophore-based screening*

The pharmacophore model is a computational representation of the essential features and constraints that a molecule needs to possess in order to interact with a specific biological target, such as a receptor or enzyme. These features include things like hydrogen bond donors/acceptors, hydrophobic regions, or specific functional groups; the pharmacophore model is constructed based on the known characteristics of the target and the expected binding interactions. In this study, the e-pharmacophore model for the human FGFR-1 complex with the reference drug was developed using four partitioning features for the identification of the hypothesis model. The generated hypothesis from the target with the reference drug is shown in Fig. 7. The features used for predicting the model with the best fitness score include two hydrogen bond donors, one hydrogen bond acceptor and two aromatic ring. The screening of the five top ranked compounds was carried out according to these features. The fitness scores for the top ranked compounds are shown in Table 3. Among the compounds, compound 2 has the best fitness score of 1.952. The fitness score serves as a measure for evaluating the geometric alignment of the features of a compound to the pharmacophore model (Table 3).

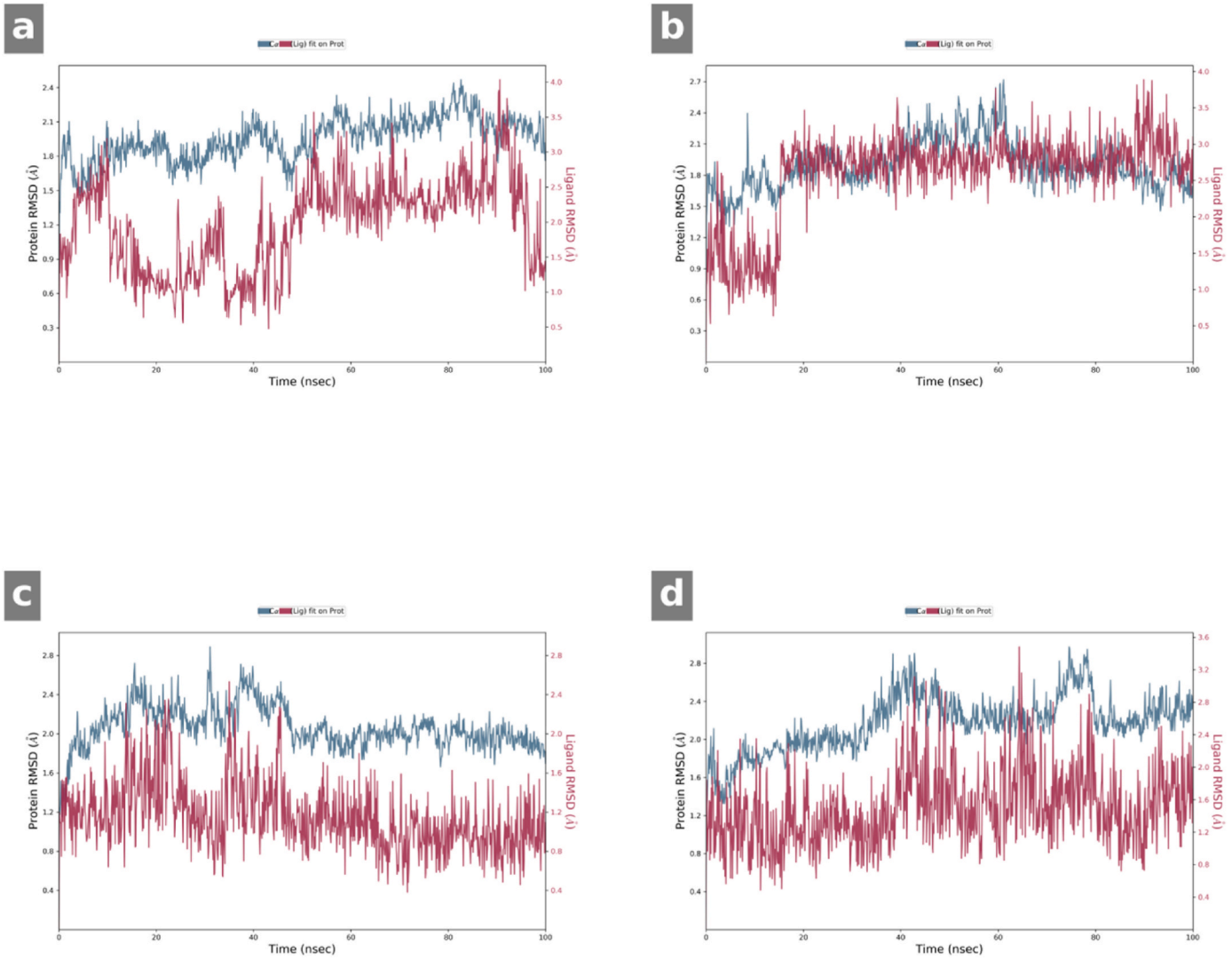

Fig. 9b. illustrates the Root Mean Square Deviation (RMSD) line plots, which measure the average change in displacement of selected atoms for each frame relative to a reference frame. Panel A shows the RMSD values for the 73350744 complex with FGFR1, while Panel B displays the RMSD values for the 162848022 complex with FGFR1. Panel C presents the RMSD values for the 137206669 complex with FGFR1. Lastly, Panel D shows the RMSD values of the reference ligand Dovitinib in complex with FGFR1.

The involvement of aromatic rings, in addition to hydrogen bond formation, in the interaction of the test compounds with the enzyme might have contributed to the higher binding affinity of the compounds compared with the standard ligand. Aromatic rings are important residues for molecular interactions and frequently exist in several protein-ligand and protein-protein interactions. Owing to their natural existence in amino acid residues like histidine, tryptophan, phenylalanine, and tyrosine, they are considered essential for protein stability and molecular recognition processes. Furthermore, aromatic rings are frequently used in drug design due to their role in the improvement of binding affinity and specificity of drug-like molecules (Ojo et al., 2021).

3.4. Structure-based screening

Molecular docking remains an important and established computational structural based virtual screening method employed in drug discovery and design. It predicts potential drug targets and molecular ligand-target interactions at the atomic level (Ferreira et al., 2015).

The 2D and 3D interaction diagrams for the five top ranked compounds are presented in Figure Fig. 8a and b. Through comprehensive analysis of these figures, we examined the binding poses and interactions of the hit compounds with the active site amino acid residues of the target, FGFR1. The specific amino acid residues involved in these interactions are listed in Table 4.

Table 5 shows a comparative evaluation of the docking scores among the top five compounds and reference drug, Dovitinib. The docking scores ranged from −10.018 to −8.419 kcal/mol, where lower scores indicated stronger binding affinity. Compound 73350744, the lead compound, exhibited a binding energy of – 10.018 kcal/mol and demonstrated six hydrogen bond interactions with ARG 627, ASN 568, GLU 562, LEU 484 and ALA 564. It also interacted with several hydrophobic amino acids, including LEU 484, ALA 640, ALA 512, ILE 545, VAL 561, TYR 563, ALA 564, VAL 492 and LEU 630 at the active site.

Compound 162848022, ranking second with a docking score of −9.981 kcal/mol, formed four hydrogen bond interactions with GLU 562, TYR 563, ALA 564, SER 565. Additionally, hydrophobic amino acids such as ALA 640, ILE 545, ALA 512, VAL 561, TYR 563, ALA 564, VAL 492, LEU 630 and LEU 484 interacted with compound 2 at the active site. Compound 137206669, with a docking score of −8.851 kcal/mol, formed two hydrogen bond interactions with GLU 562 and ALA 564. Similar hydrophobic interactions occurred with various amino acids such as VAL 492, LEU 630, ALA 640, ILE 545, ALA 512, VAL 561,

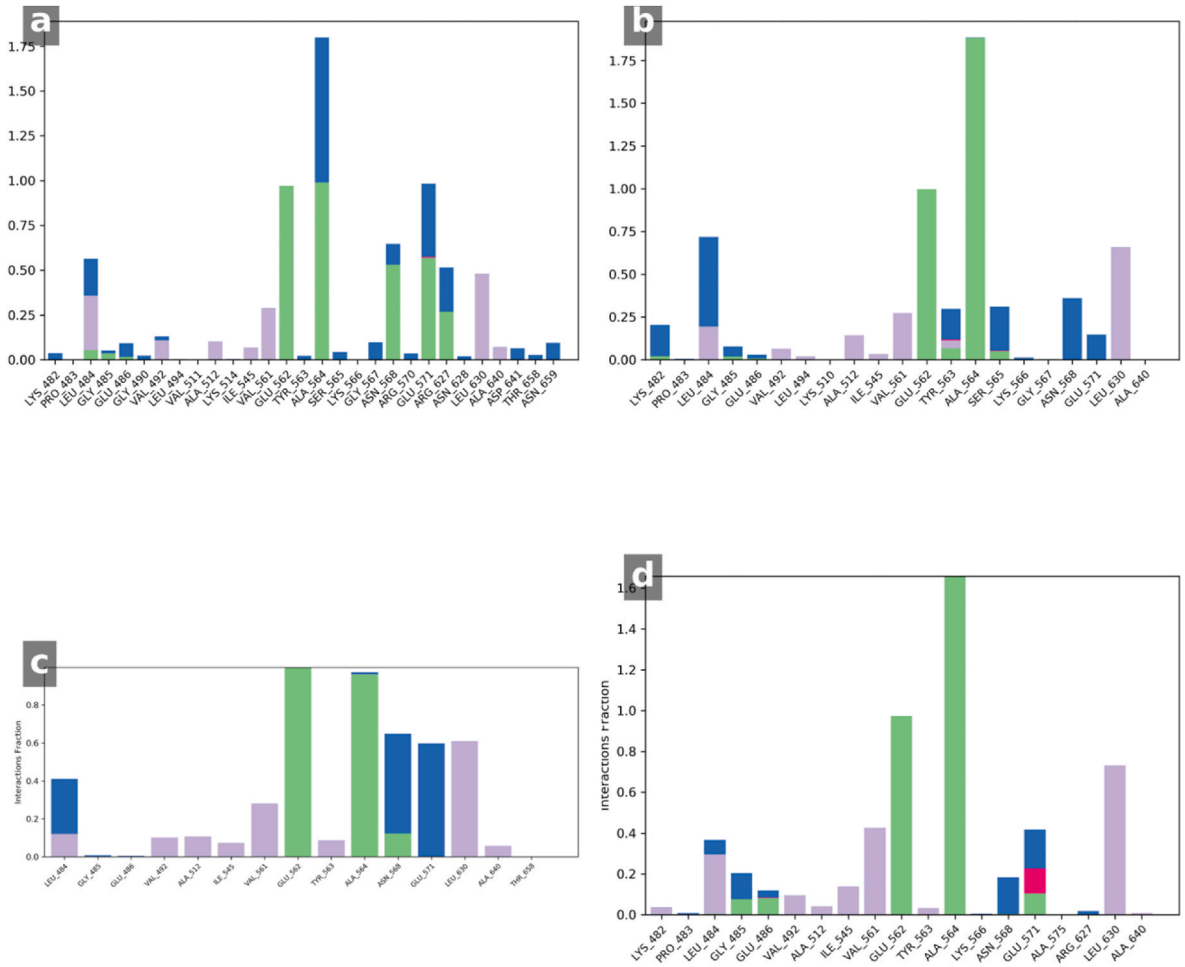

Fig. 9c. The protein-ligand contact histogram used for comprehensive interaction fraction within the simulation trajectory. A presents the protein-ligand contact histogram of 73350744 complex with FGFR1, B presents the protein-ligand contact histogram of 162848022 and FGFR1, C presents the protein-ligand contact histogram of 137206669 and FGFR1 and D presents the protein-ligand contact histogram of the reference ligand Dovitinib and FGFR1.

TYR 563, ALA 564 and LEU 484. Compound 102410411 had a good docking score of −8.442 kcal/mol, forming hydrogen bonds with LEU 484, ALA 564 and SER 565. Hydrophobic contacts were formed with ILE 565, ALA 512, VAL 561, TYR 563, LEU 630, ALA 564, LEU 484 and VAL 492.

Similarly, Compound 46939591 with a docking score of −8.439 kcal/mol, engaged in four hydrogen bond interactions with ALA 564, SER 565 and ASP 641. Hydrophobic contacts involving LEU 644, ALA 640, VAL 492, ALA 512, LEU 630, TYR 563, ALA 564 and LEU 484 were also established. The standard drug, Dovitinib, had a high docking score of −8.419 kcal/mol, forming three hydrogen interaction with GLU 562 and ALA 564 and formed hydrophobic contacts with ALA 640, ILE 545, VAL 492, ALA 512, VAL 561, TYR 563, ALA 564, LEU 630 and LEU 484.

3.5. Binding free energy calculation

Although molecular docking is commonly utilized in computer-aided drug design (CADD), it has been criticized for its lack of some energy parameters, such as solvation energy systems. To address this issue, the validation procedure was further supported by prime MM/GBSA calculation. MM/GBSA is known for its dependability and accuracy in determining the structural stability and binding energy of protein-ligand complexes (Bodun et al., 2023). A positive binding free energy (MM/GBSA) score depicts false binding energy (docking score). Remarkably, the results of the binding free energy obtained from the investigated natural compounds and inhibitor formed favorable stability with FGFR1 with negative binding free energy values (Table 4). Compound 1 (−59.10) and Compound 2 (−51.80) complexes were computed with the highest binding free energy among the ligands and the FGFR1-inhibitor complex (−46.28) was observed to give the lower binding free energy.

3.6. Drug likeness and ADMET

Majority of potential molecules fail during the final stages of clinical trials due to undesirable side effects and toxicity. The primary reasons for these failures are typically related to toxicity issues. So predicting the pharmacokinetic and toxicological characteristics of a potential lead compound during the initial stages of drug discovery is a critical strategy to mitigate future risks. In this study, we examined the absorption, distribution, metabolism, excretion and toxicity (ADMET) profiles of the top five lead compounds using the SwissADME web server.

The results of the predictions are presented in Tables 6 and 7. All the top 5 lead compounds obeyed the Lipinski rule of five, a rule in which a compound is considered a drug candidate if it does not violate one of the following rules: hydrogen bond donor not greater than 5 (HBD $\leq$5), hydrogen bond acceptor not greater than 10 (HBA $\leq$10), molecular weight not greater than 500 Da (MW $\leq$ 500 Da), and octanol-water partition coefficient not greater than 5 (LogP$\leq$5) (Lipinski et al., 2012). The results shown on Table 6 revealed that all the top ranked compounds including the reference drug, Dovitinib obeyed all Lipinski's rule of five. This suggests that all the top ranked compounds can be

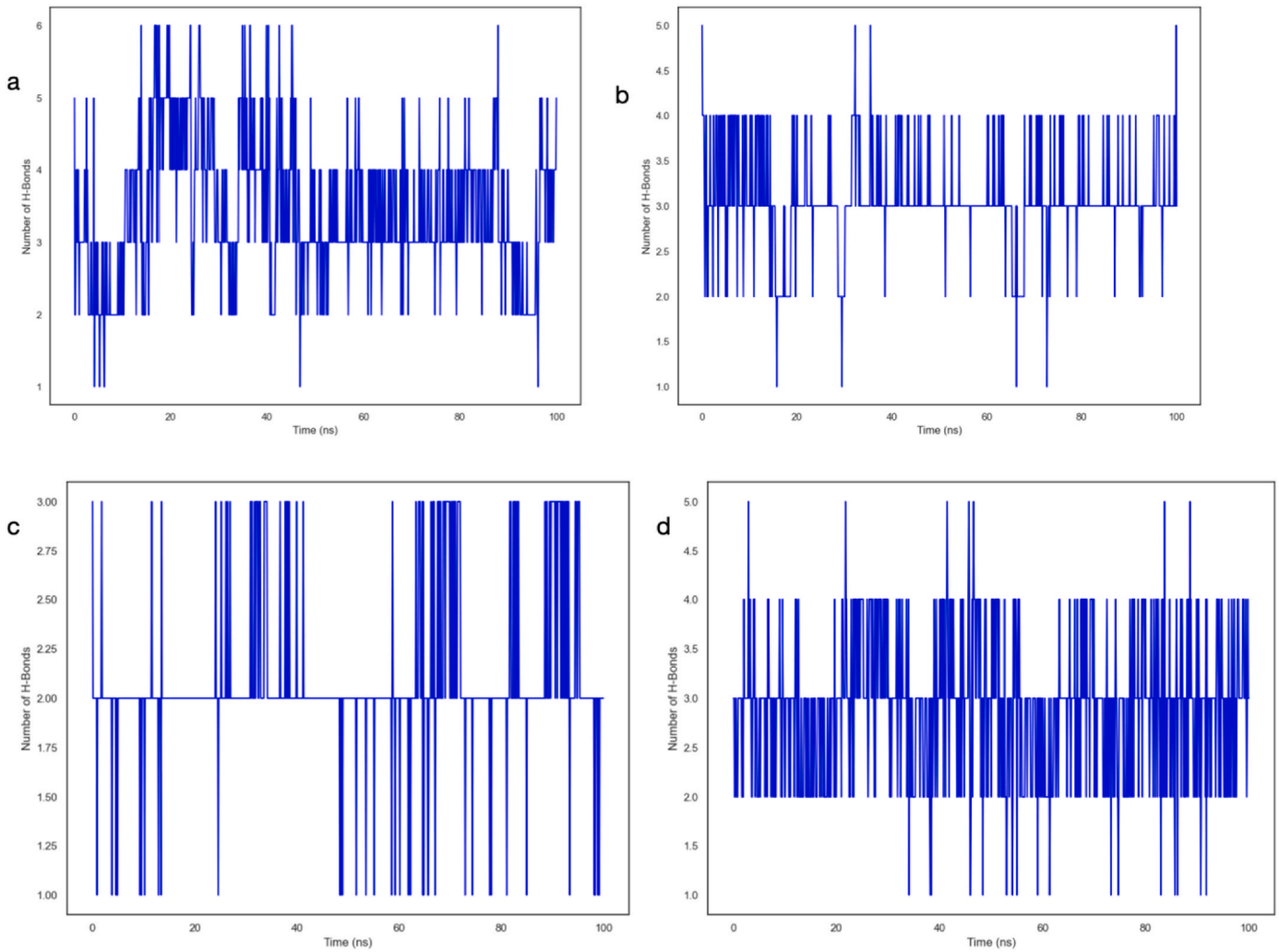

Fig. 10. Number of H-bond contacts. a) 73350744 b) 162848022 c) 137206669 d) Dovitinib (Reference Ligand).

predicted as good therapeutic candidates. Furthermore, the compounds showed good metabolism profile with the CYP450 isoenzymes (the key enzymes in drug metabolism) which is characterized by the non-inhibition of CYP450 2C9, CYP450 2C19 and CYP450 3A4. CYP450 1A2 also showed non-inhibitory properties for compounds 73350744, 162848022, and 46939591 whereas the enzyme was inhibited by 137206669, 102410411 and the standard drug. Similarly, the CYP450 2D6 was also inhibited by compound 4 and the standard drug as seen in Table 7.

The oral bioavailability is the critical factor for any potential drug candidate. To be orally bioavailable, a compound must possess specific functions that enable it to penetrate cellular membranes. Without these functions, the compound may become trapped within the barrier and pose a serious health risk. These functions are desolvation, diffusion, resolvation and physicochemical properties including lipophilicity (Kullappan et al., 2023).

As shown in Table 6, all the five top ranked compounds, including the standard drug are permeable to the human intestinal membrane and Caco-2 cell permeability, signifying favorable absorption through the human intestine. The p-glycoprotein is a significant drug transporter (Fortuna et al., 2012), and thus, examining the potential for the top-ranked compounds to act as substrates or inhibitors of p-glycoprotein is clinically relevant. Compounds 73350744, 162848022, and 46939591 inhibited p-glycoprotein while compounds 3, 4 and the standard were predicted to be non-substrate. Additionally, all the top-ranked compounds including the standard have an oral bioavailability rate of 0.55.

Lastly, the toxicity assessment revealed that all compounds, including the standard, were non-carcinogenic, non-eye corrosive, and non-irritating. Furthermore, none of the compounds were projected to cause skin irritation. Additionally, all compounds, except for Compound 5 and the standard, were predicted to have no mutagenic potential according to the Ames mutagenesis test, which evaluates mutagenic or carcinogenic effects. From our observations, we can infer that the top-ranked compounds could be safe to prescribe as medicinal candidates.

Based on the key features important for classifying actives and inactives against FGFR1, TPSA, EState_VSA8, PEOE_VSA1, and PEOE_VSA8 all provide insights into the polar or hydrophilic nature of molecules. Higher TPSA values indicate increased polarity, influencing solubility and membrane permeability. The positive correlation between TPSA and molecular weight is notable, as most compounds classified as active (Fig. 2) have a molecular weight above 300, falling within the optimal TPSA range of 60–140 $Å^2$ (Prasanna and Doerksen, 2009). This suggests that active compounds possess a balance between polarity and molecular size, which is crucial for their biological activity hence the significance of these descriptors.

4. Molecular dynamic simulation

4.1. RMSF

The local changes and flexibility of the FGFR1 and 73350744,

162848022, 137206669, and Dovitinib in respective complexes were evaluated for root mean square fluctuations (RMSF) analysis throughout the 100 ns simulation trajectories. RMSF was done to study the flexibility of protein-ligand complexes and fluctuation in interactive amino acid residues in the secondary structure of the target protein (Bodun et al., 2023). The RMSF plot characterizes the local fluctuations of the residues of the protein. The peaks indicate the residues of the protein that fluctuate the most. Based on the resulting output, the amino acid residues of FGFR1 of the docked complexes did not have similar fluctuation, especially between the 160–280 amino acid residues during the simulation process of 100 ns.

RMSF plots are instrumental in identifying regions within the protein that exhibit considerable variability during the simulation, with pronounced peaks indicating areas of constant and frequent fluctuation, thus suggesting less stability (Olanrewaju et al., 2024). Typically, the peripheral or tail regions of proteins are less stable compared to their core, leading to higher fluctuations observed in these areas, a phenomenon that was noted in the reference ligand complex. In Frame 1 of Fig. 9a, the FGFR1-73350744 complex initially shows a peak at 3.4 Å, followed by continuous and consistent residue fluctuations ranging between 1.5 Å and a peak of 3.0 Å. Significant fluctuations are observed at amino acid residues 0–50 and 110–130. Frame 2 illustrates the FGFR1-162848022 complex's fluctuation over a 100 ns simulation period, with the highest fluctuation occurring at residues 110–130, reaching a distance of 3.5 Å, which is within the permissible range. Frame 3 shows the FGFR1-137206669 complex interaction and fluctuation over the same 100 ns period, with residue 0–50 exhibiting fluctuations of 3.2 Å and 3.0 Å, and the highest fluctuation at residues 110–130 with a range of 3.5 Å. Finally, a comparative simulation study with the standard Dovitinib reveals similar fluctuations around residues 0–50 and 110–120, comparable to Frames 1, 2, and 3. However, Dovitinib exhibits a maximum fluctuation range exceeding 4.8 Å at residues 170–200, indicating a less stable interaction over the 100 ns simulation period compared to 73350744, 162848022, and 137206669.

5. RMSD

Fig. 9b represents the ligand RMSD value, which measures the stability of the docked ligand pose within the binding pocket on the Y-axis. The term "Lig Fit Prot" refers to the RMSD values of the ligand to the protein backbone, providing additional insight into the conformational changes and structural compatibility between the ligand and the protein. Typically, the "Lig Fit Prot" values are slightly higher than the protein RMSD, which suggests some flexibility or movement of the ligand relative to the protein backbone. However, if this value is significantly higher, it may indicate substantial changes in the ligand's pose compared to its initial docked position. Such deviations could imply a less stable interaction between the ligand and the protein or notable conformational adjustments within the binding pocket, potentially affecting the ligand's effectiveness in binding to its target (Olanrewaju et al., 2024).

Fig. 9b, Frame 1, illustrates the RMSD deviation of the FGFR1-73350744 complex over a 100 ns simulation period. The results indicate an unstable interaction with a deviation range of 0.5–4.0 Å which is above the acceptable value. The complex appears stable at the beginning of the interaction but experiences a significant deviation, peaking at approximately 4 Å towards the end of the simulation period. Notably, the FGFR1-73350744 complex exhibits a significantly reduced "Lig Fit Prot" value, indicating a normalization of the complex interaction by the end of the 100 ns simulation. Frame 2 shows the FGFR1-162848022 complex, which demonstrates a significant deviation in the first 20ns of the simulation period. However, after the 20 ns mark, there is a somewhat stable interaction from 1.7 Å to a "Lig Fit Prot" value of 3.8 Å. Despite the initial fluctuations, the deviation remains within the permissible range after the 20ns mark, suggesting a stable interaction. Frame 3 displays the FGFR1-137206669 complex, which maintains a more stable interaction, as indicated by a relatively consistent "Lig Fit Prot" value, with minor fluctuations from 0.4 Å to a "Lig Fit Prot" value of 2.6 Å, suggesting a stable interaction across the entire simulation period. The reference ligand-complex in frame 4 also shows somewhat stable interaction, with a "Lig Fit Prot" value of 0.5 Å to 3.5 Å, reflecting a less fluctuating interaction with the FGFR1 protein throughout the simulations.

5.1. Interaction fraction analyses

Fig. 9c, Frame 1, shows that 73350744 formed hydrogen bonds with eight specific amino acids, including LEU484, GLU485, GLU486, GLU562, ALA564, ASN568, GLU571, and ARG627, each displaying varying durations of interaction. Notably, hydrogen bonds with GLU562, ALA564, ASN568, and GLU571 were maintained for over 50 % of the simulation time, with GLU571 engaged for approximately 25 % of the simulation duration. In contrast, LEU484, GLU485, and GLU486 sustained hydrogen bonds for less than 10 % of the simulation time, with ALA564 demonstrating the most prolonged interaction, being maintained 100 % of the time. Furthermore, ALA564, ASN568, GLU571, and ARG627 also participated in water bridge formations—hydrogen bonds mediated by water molecules occurring approximately 40 %–80 % of the simulation, respectively, underscoring the dynamic nature of ligand-protein interactions. Hydrophobic interactions were observed with LEU484, VAL492, ALA512, ILE545, VAL561, LEU630, and ALA640, all of which were maintained for less than 50 % of the simulation duration.

Frame 2 depicts the interaction fraction between FGFR1 and 162848022, showing hydrogen bonds established with four amino acid residues: GLY485, GLU486, and SER565, all maintained for less than 20 % of the simulation time. Prolonged hydrogen bond interactions were observed with GLU562 and ALA564, lasting about 80 % and above 175 % of the simulation time, respectively. Notably, there was a pronounced hydrophobic interaction with LEU630, which persisted for more than 50 % of the simulation period. Frame C analyzes the interaction fraction of FGFR1 complexed with 137206669, involving three major amino acids in hydrogen bond interactions: GLU562 and ALA564, both sustained for 100 % of the simulation time, and ASN568 for less than 20 %. Additionally, LEU485, ASN568, and GLU571 participated in water bridge formations for approximately 40 %–60 % of the simulation duration. This complex exhibited increased hydrophobic bond interactions with seven amino acid residues, including LEU485, VAL492, ALA512, ILE545, VAL561, TYR563, LEU630, and ALA640, with interaction durations ranging from 10 % to 60 % of the simulation time.

In the MDS analysis, the reference ligand Dovitinib engaged in five hydrogen bond interactions with amino acids GLY485, GLU486, GLU562, ALA564, and GLU571, demonstrating significant stability (Fig. 10). Notably, interactions with GLU562 and ALA564 were maintained for 80 % and above 160 % of the simulation time, respectively. This finding is consistent with the initial docking results, emphasizing the crucial roles of GLU562 and ALA564 in the binding of the reference ligand. Additionally, of the six observed water bridge interactions, only those involving LEU484 and GLU571 were sustained for about 40 % of the simulation duration. A significant hydrophobic interaction with LEU630 persisted for 60 % of the simulation, highlighting its role in the stability of the protein-ligand complex. Furthermore, an ionic bond with GLU571 was maintained for approximately 20 % of the simulation time. Remarkably, GLU562 and ALA564 consistently demonstrated the most prolonged hydrogen bond interactions across all frames, underscoring their crucial roles in the active site of FGFR1. Their consistent involvement highlights their significance in ligand binding interactions and stability within the active site.

6. Conclusion

FGFR (1–4) enzymes play a significant role in carcinogenesis, including cholangiocarcinoma (CCA), through their involvement in key

signaling pathways. Current treatment options for CCA remain limited. In this study, we screened a database of natural compounds to identify potential FGFR-1 inhibitors for CCA therapy. We identified five compounds with stronger binding affinities and more favorable binding free energies compared to the reference ligand. All five compounds also demonstrated drug-like properties based on their ADMET profiles. Furthermore, molecular dynamics simulations over 100 ns revealed that compounds 162848022 and 137206669 exhibited lower fluctuations at the active site, indicating greater stability than the reference ligand. Based on these findings, we recommend further *in vitro* and *in vivo* studies to validate the potential of these compounds as FGFR-1 inhibitors for the treatment of CCA.

CRediT authorship contribution statement

Samuel Aduramurewa Osunnaya: Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Formal analysis, Conceptualization. **Wilberforce K. Ndarawit:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Ifeoluwa Aderibigbe:** Writing – review & editing, Methodology, Investigation. **Ibilola A. Omolopo:** Writing – review & editing, Writing – original draft, Methodology. **Precious O. Aribisala:** Writing – original draft, Visualization. **Ayodele Oluwasegun Elekan:** Visualization, Data curation. **Adeola Sakirat Adeyemo:** Writing – original draft, Data curation. **Sheriffdeen Abiola Amoo:** Visualization, Data curation. **Olatunde Simbiat Olamiposi:** Writing – original draft, Formal analysis. **Njogu M. Kimani:** Formal analysis. **Taiwo Hamidat Olaide:** Visualization, Formal analysis. **Adedoyin John-Joy Owolade:** Writing – review & editing. **Damilola Samuel Bodun:** Writing – review & editing, Writing – original draft, Supervision, Investigation, Conceptualization.

Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request. Additional supporting data are provided in the supplementary materials associated with this publication.

Disclosure statement

The authors declare no conflicts of interest concerning the authorship, research work and/or publication of this work.

Funding information

No fund was received for this study.

Declaration of competing interest

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

Abbreviations

PI3K	Phosphoinositide 3-kinase
AKT	Protein kinase B
mTOR	Mammalian target of rapamycin
RAF	Rapidly accelerated fibrosarcoma
MAPK	Mitogen-activated protein kinase
JAK	Janus kinase
STAT	Signal transducers and activators of transcription
PLCγ	Phospholipase C Gamma

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.amolm.2025.100085.

References

Banales, J.M., Cardinale, V., Carpino, G., Marzioni, M., Andersen, J.B., Invernizzi, P., Lind, G.E., Folseraas, T., Forbes, S.J., Fouassier, L., Geier, A., Calvisi, D.F., Mertens, J.C., Trauner, M., Benedetti, A., Maroni, L., Vaquero, J., Macias, R.I.R., Raggi, C., et al., 2016. Expert consensus document: cholangiocarcinoma: current knowledge and future perspectives consensus statement from the European Network for the Study of Cholangiocarcinoma (ENS-CCA). Nat. Rev. Gastroenterol. Hepatol. 13 (5), 261–280. https://doi.org/10.1038/nrgastro.2016.51.

Bodun, D.S., Omoboyowa, D.A., Omotuyi, O.I., Olugbogi, E.A., Balogun, T.A., Ezeh, C.J., Omirin, E.S., 2023. QSAR-based virtual screening of traditional Chinese medicine for the identification of mitotic kinesin Eg5 inhibitors. Comput. Biol. Chem. 104, 107865.

Bodun, D.S., Abdulwahab, A.A., Temitope, A.I., Chidinma, E.O., Aduramurewa, O.S., Alonge, M.O., et al., 2024. Machine learning-based drug repositioning of novel human aromatase inhibitors utilizing molecular docking and molecular dynamic simulation. J. Comput. Biophysics and Chem.

Chmiel, P., Gęca, K., Rawicz-Pruszyński, K., Polkowski, W.P., Skórzewska, M., 2022. FGFR inhibitors in cholangiocarcinoma—a novel yet primary approach: where do we stand now and where to head next in targeting this Axis? Cells 11 (23). https://doi.org/10.3390/cells11233929.

Coleman, S.J., Chioni, A., Ghallab, M., Anderson, R.K., Lemoine, N.R., Kocher, H.M., Grose, R.P., 2014. Nuclear translocation of FGFR1 and FGF2 in pancreatic stellate cells facilitates pancreatic cancer cell invasion. EMBO Mol. Med. 6 (4), 467–481. https://doi.org/10.1002/emmm.201302698.

Cross, M.J., Claesson-Welsh, L., 2001. FGF and VEGF function in angiogenesis: signalling pathways, biological responses and therapeutic inhibition. Trends Pharmacol. Sci. 22 (4), 201–207. https://doi.org/10.1016/S0165-6147(00)01676-X.

Ferreira, L.G., Dos Santos, R.N., Oliva, G., Andricopulo, A.D., 2015. Molecular docking and structure-based drug design strategies. Molecules 20 (7), 13384–13421.

Fortuna, A., Alves, G., Falcão, A., Soares-da-Silva, P., 2012. Evaluation of the permeability and P-glycoprotein efflux of carbamazepine and several derivatives across mouse small intestine by the Ussing chamber technique. Epilepsia 53 (3), 529–538.

Ghouri, Y.A., Mian, I., Blechacz, B., 2015. Cancer review: cholangiocarcinoma. J. Carcinog. 14, 1.

Goetz, R., Mohammadi, M., 2013. Exploring mechanisms of FGF signalling through the lens of structural biology. Nat. Rev. Mol. Cell Biol. 14 (3), 166–180. https://doi.org/10.1038/nrm3528.

Izquierdo-Sanchez, L., Lamarca, A., La Casta, A., Buettner, S., Utpatel, K., Klümpen, H.J., Banales, J.M., 2022. Cholangiocarcinoma landscape in Europe: Diagnostic, prognostic and therapeutic insights from the ENSCCA Registry. J. Hepatol. 76 (5), 1109–1121.

Khan, S.A., Tavolari, S., Brandi, G., 2019. Cholangiocarcinoma: Epidemiology and risk factors. Liver Int. 39, 19–31.

Kim, S., Dubrovska, A., Salamone, R.J., Walker, J.R., Grandinetti, K.B., Bonamy, G.M.C., Orth, A.P., Elliott, J., Porta, D.G., Garcia-Echeverria, C., Reddy, V.A., 2013. FGFR2 promotes breast tumorigenicity through maintenance of breast tumor-initiating cells. PLoS One 8 (1), e51671. https://doi.org/10.1371/journal.pone.0051671.

Kullappan, M., Mary, U., Ambrose, J.M., Veeraraghavan, V.P., Surapaneni, K.M., 2023. Elucidating the role of N440K mutation in SARS-CoV-2 spike–ACE-2 binding affinity and COVID-19 severity by virtual screening, molecular docking and dynamics approach. J. Biomol. Struct. Dyn. 41 (3), 912–929.

Lavecchia, A., 2019. Deep learning in drug discovery: opportunities, challenges and future prospects. Drug Discov. Today 24 (10), 2017–2032. https://doi.org/10.1016/j.drudis.2019.07.006.

Link, W., 2019. Drug discovery and development. In: Link, W. (Ed.), Principles of Cancer Treatment and Anticancer Drug Development. Springer International Publishing, pp. 87–136. https://doi.org/10.1007/978-3-030-18722-4_4.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J., 2012. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 64, 4–17.

Nag, S., Baidya, A.T.K., Mandal, A., Mathew, A.T., Das, B., Devi, B., Kumar, R., 2022. Deep learning tools for advancing drug discovery and development. 3 Biotech 12 (5), 110. https://doi.org/10.1007/s13205-022-03165-8.

Ojo, O.A., Adegboyega, A.E., Johnson, G.I., Umedum, N.L., Onuh, K., Adeduro, M.N., Nwobodo, V.O., Elekan, A.O., Alemika, T.E., Johnson, T.O., 2021. Deciphering the interactions of compounds from *Allium sativum* targeted towards identification of novel PTP 1B inhibitors in diabetes treatment: a computational approach. Inform. Med. Unlocked 26 (15), 100719.

Olanrewaju, J.A., Arietarhire, L.O., Soremekun, O.E., Olugbogi, E.A., Aribisala, P.O., Alege, P.E., et al., 2024. Reporting the anti-neuroinflammatory potential of selected spondias mombin flavonoids through network pharmacology and molecular dynamics simulations. In Silico Pharmacology 12 (2), 74.

Ornitz, D.M., Legeai-Mallet, L., 2017. Achondroplasia: development, pathogenesis, and therapy. Dev. Dyn. 246 (4), 291–309. https://doi.org/10.1002/dvdy.24479.

Pasrija, P., Jha, P., Upadhyaya, P., Khan, Mohd S., Chopra, M., 2022. Machine learning and artificial intelligence: a paradigm shift in big data-driven drug design and

discovery. Curr. Top. Med. Chem. 22 (20), 1692–1727. https://doi.org/10.2174/1568026622666220701091339.

Prasanna, S., Doerksen, R.J., 2009. Topological polar surface area: a useful descriptor in 2D-QSAR. Curr. Med. Chem. 16 (1), 21–41.

Rutz, A., Sorokina, M., Galgonek, J., Mietchen, D., Willighagen, E., Gaudry, A., Graham, J.G., Stephan, R., Page, R., Vondrášek, J., Steinbeck, C., Pauli, G.F., Wolfender, J.-L., Bisson, J., Allard, P.-M., 2022. The LOTUS initiative for open knowledge management in natural products research. Elife 11, e70780. https://doi.org/10.7554/eLife.70780.

Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F., 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 71 (3), 209–249. https://doi.org/10.3322/caac.21660.

Teven, C.M., Farina, E.M., Rivas, J., Reid, R.R., 2014. Fibroblast growth factor (FGF) signaling in development and skeletal diseases. Genes & Diseases 1 (2), 199–213. https://doi.org/10.1016/j.gendis.2014.09.005.

Turner, N., Grose, R., 2010. Fibroblast growth factor signalling: from development to cancer. Nat. Rev. Cancer 10 (2), 116–129. https://doi.org/10.1038/nrc2780.

Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., Zhao, S., 2019. Applications of machine learning in drug discovery and development. Nat. Rev. Drug Discov. 18 (6), 463–477. https://doi.org/10.1038/s41573-019-0024-5.

Wang, H., Yang, J., Zhang, K., Liu, J., Li, Y., Su, W., Song, N., 2021. Advances of fibroblast growth factor/receptor signaling pathway in hepatocellular carcinoma and its pharmacotherapeutic targets. Front. Pharmacol. 12. https://doi.org/10.3389/fphar.2021.650388.

Zhang, S., He, X., Xia, X., Xiao, P., Wu, Q., Zheng, F., Lu, Q., 2023. Machine-learning-enabled framework in engineering plastics discovery: a case study of designing polyimides with desired glass-transition temperature. ACS Appl. Mater. Interfaces 15 (31), 37893–37902.

Zhou, M., Luo, J., Chen, M., Yang, H., Learned, R.M., DePaoli, A.M., Tian, H., Ling, L., 2017. Mouse species-specific control of hepatocarcinogenesis and metabolism by FGF19/FGF15. J. Hepatol. 66 (6), 1182–1192. https://doi.org/10.1016/j.jhep.2017.01.027.