Received: 3 September 2024 | Revised: 18 December 2024 | Accepted: 25 December 2024 | Published online: 8 January 2025

RESEARCH ARTICLE

Medinformatics
2025, Vol. 2(1) 22–35
DOI: 10.47852/bonviewMEDIN52024239

BON VIEW PUBLISHING

In Silico Study and Validation of Natural Compounds Derived from *Macleaya cordata* as a Potent Inhibitor for BTK

Ayobami Fidelix[1],*, Tomilola Akingbade[2], Jatin Jangra[3], Babatunde Olabuntu[4], Olutola Adeyemo[5] and Juwon Akingbade[6]

[1]*Department of Neurosurgery, University of Texas, USA*

[2]*Computer-Aided Therapeutic Discovery and Design Platform, Federal University of Technology, Nigeria*

[3]*Department of Pharmaceutical Engineering and Technology, Indian Institute of Technology, India*

[4]*Department of Biochemistry, University of Ibadan, Nigeria*

[5]*Department of Obstetrics and Gynecology, Federal Teaching Hospital, Nigeria*

[6]*Department of Computer Engineering, Federal University Oye Ekiti, Nigeria*

Abstract: Bruton's tyrosine kinase (BTK) is a kinase of the TEC family expressed in B cells and other hematopoietic cells, but it is not expressed in T cells. B-cell malignancies such as multiple myeloma and chronic lymphocytic leukemia have been shown to have a high expression of BTK, thereby displaying oncogenic activities in these diseases, triggering the discovery of BTK inhibitors. The study investigated computationally the phytochemical present in *M cordata* as a novel BTK inhibitor with high efficacy in treating B-cell malignancies. Chelidimerine, Bocconarborine A, and Bocconarborine B show a high binding affinity of −13.7, −13.3, and −12.9, respectively. This study was validated using molecular dynamic stimulation to indicate the stability and interaction of the ligand with the targeted protein. Bocconarborine B has the best binding energy of −30.94 kcal/mol compared to ibrutinib, with a binding energy of −22.46 kcal/mol. The identified hit compounds from this study were subjected to half maximum inhibitory concentration prediction (IC_{50}) using machine learning modeling; the result shows that Bocconarborine B has the best IC_{50} of 48.98 nM. This study is subject to validation via in vivo and in vitro studies.

Keywords: machine learning modeling, B-cell malignancies, Bocconarborine B, ibrutinib, molecular dynamic simulation

1. Background

Bruton's tyrosine kinase (BTK) is a kinase of the TEC family expressed in B cells and other hematopoietic cells, but it is not expressed in T cells [1]. BTK phosphorylates tyrosine and serine upon antigen binding to the B-cell receptor (BCR), where the BCR signaling effect is significantly substantial by playing a crucial role in adaptive immunity [2]. The binding of antigens to the BCR initiates the signaling pathway, leading to the differentiation and proliferation of B lymphocytes (Figure 1) [3]. The BTK comprises 659 amino acids with five different domains, namely TH domains, PH domains, SH3 domains, SH2 domains, and catalytic domains. The domains have different functions and phosphorylation sites. The site for the autophosphorylation of the SH2 and SH3 is Tyrosine 223. At the same time, the catalysis domain gets phosphorylated at Tyrosine 551 and Cysteine 481, and the PH domain functions by regulating the interaction of phospholipids and other proteins [4].

B-cell malignancies such as multiple myeloma and chronic lymphocytic leukemia have been shown to have a high expression of BTK, thereby displaying oncogenic activities in this disease, triggering the discovery of BTK inhibitors [5]. Despite the success of developing BTK inhibitors, the challenge of acquiring resistance mutations to each of the classes of BTK inhibitors has been a significant barrier to treating malignancies associated with BTK mutation [6].

The aberration in the activation of protein kinases is a driver for several malignancies amidst the alterations in cellular motility, proliferation, metabolism, etc. [7, 8]. The central role played by BTK in several B-cell signaling pathways has made it an important therapeutic target for diverse malignancies associated with B cells. Also, the inhibition of BTK is a promising therapy that influences the crucial immune cells in the tumor microenvironments [9]. The use of multidrug combination therapy has lately been a focus for improving clinical response. However, the major setback in this combination therapy is the excessive cytotoxicity [10].

*Corresponding author: Ayobami Fidelix, Department of Neurosurgery, University of Texas, USA. Email: ayobami.o.fidelix@uth.tmc.edu

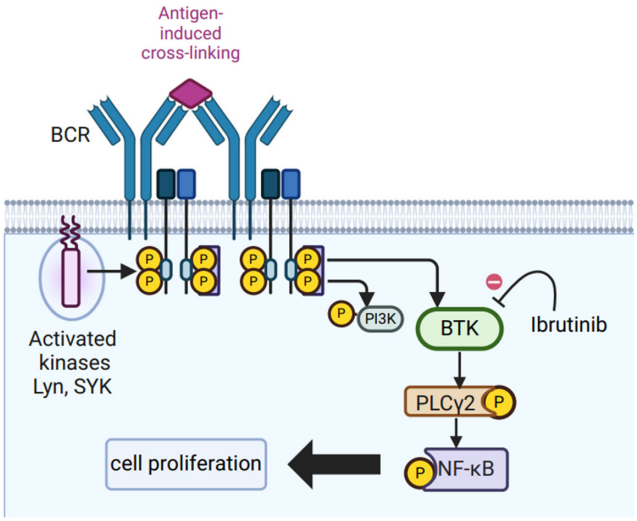


Figure 1. B-cell receptor signaling.

Thus, there is a need to develop a novel BTK inhibitor with high efficacy in treating B-cell malignancies. The development of acquired resistant mutation using synthetic drugs prompted the quest for natural products with potent anticancer abilities [11].

M cordata is a perennial medicinal herb of the family of Papaveraceae, and they are primarily distributed in China [12]. *M cordata*, as a medicinal herb, encompasses diverse biomedical applications such as anti-inflammatory [13], antiviral [14], etc. The phytochemical screening of *M. cordata* reveals the presence of various phytocompounds, such as alkaloids [15].

This study investigated using a computer-aided drug design approach: molecular docking, molecular dynamic simulation, and model training for predicting BTK potential inhibitors.

2. Material and Methods

2.1. Preparation of ligands

The chemicals present in *Macleaya cordata* were obtained from previously published work [16]. The name of each phytochemical was searched in the PubChem database on NCBI, and each chemical structure was obtained in SMILES and SDF formats. The 2D structures of the phytochemicals and the reference drug (ibrutinib) were downloaded in SDF format. Using Discovery Studio, the SDF structures were converted and saved in PDB format.

2.2. Protein preparation

The 3D crystal structure of BTK was retrieved from the RCSB Protein Data Bank (https://www.rcsb.org/) [17] having a PDB ID: 5P9J [18]. The atomic and molecular structure of the crystal was resolved by the X-ray diffraction method and has a resolution value of 1.08 Å. UCSF Chimera Software (v 1.16) was used to minimize and prepare the target protein. The prediction of binding sites of protein was made through BIOVIA Discovery Studio. Its PDB ID was subjected to PPI analysis using the STRING webserver (https://string-db.org/).

2.3. Protein-ligand binding score analysis by molecular docking

Molecular docking studies can reveal the binding interactions between specific proteins and desired ligands. The proteins and ligands were prepared for docking in pdbqt format using Auto Dock Tools 4.2 [19]. For binding to take place, the x, y, and z dimensions were set at 23.2328 × 22.0693 × 23.1420, and the grid box center was set as follows: x, 17.7327; y, 7.3847; z, 4.5447 to obtain favorable docking conformations. The binding site is defined by the residues viz. Cys481, Met477, Thr474, Glu475, Lys430, Val416 and Met449.

All the ligands were docked, and a maximum of 8 exhaustiveness was computed for all of them. All other parameters in the software were in default mode. The binding affinities of compounds for the protein targets were recorded. The compounds were then ranked by their affinity scores. Following docking, PyMol and LigPlot were used to visualize the protein-ligand interactions and Ligand binding residues.

2.4. Drug-likeness and pharmacokinetics prediction

The in silico pharmacokinetics properties of the phytochemical compounds were conducted on the SwissADME web server, available at www.swissadme.ch at default settings [20], by using the SMILES of each of the phytochemicals. These compounds were subjected to Lipinski's rule, Muegge's rule, Ghose's rule, Veber's rule, and Egan's rule to assess their potential as drug candidates.

Furthermore, the absorption, distribution, metabolism, elimination, and toxicity (ADMET) properties of these compounds were evaluated using Deep-PK webserver [21].

2.5. Protein-ligand stability analysis by molecular dynamics (MD) simulation

To investigate the structural stability of protein–ligand complexes under physiological conditions, comprehensive MD simulations were conducted. Based on their energetically favorable docking conformations, the top-ranked complexes were selected for analysis using GROMACS 2020 in conjunction with the CHARMM-36M force field [22]. Ligand topologies were generated using CHARMM-GUI [23]. Complexes were solvated with TIP3P water model and neutralized with Na+ and Cl− ions. Systems underwent energy minimization and equilibration before a 100 ns MD production run at 310.15 K and 1 bar. MD simulation studies were conducted on the top-selected protein-ligand complexes using GROMACS 2020 with the CHARMM-36M force field. The post-simulation analysis includes RMSD, RMSF, inter-hydrogen bonds, SASA, and Rg calculations using GROMACS modules. Plots were generated with QtGrace tool [24, 25]. In addition, binding free energy (ΔGbind) was calculated using the prime molecular mechanics/generalized Born surface area method [26]. The comprehensive equation for ΔGbind is:

$$\begin{aligned}\Delta G_{bind} &= \left[G_{complex} - \left(G_{receptor} + G_{ligand}\right)\right] \\ &= \left[\left(E_{bond} + E_{angle} + E_{dihedral} + E_{ele} + E_{vdw}\right) + \left(G_{polar} + \gamma_{SASA} + b\right)\right] \\ &\quad - T\Delta S\end{aligned}$$

2.6. Collection and cleansing of data for IC_{50} prediction

The BTK biological activity data (Tyrosine-protein kinase BTK) were acquired from the ChEMBL database [27] using the ChEMBL target ID: 5251, the Chembl client module within the Jupyter computational notebook environment. A total of 3927 compounds with reported half maximum inhibitory concentration ($1C_{50}$) values for BTK were obtained.

The data collected were subjected to cleaning by removing identical compounds without IC_{50} values and available SMILES annotations. A total of 2756 refined compounds were collected after the cleaning process. The IC_{50} values were converted to pIC_{50} values by subjecting the IC_{50} values to the negative logarithm at base 10, thereby providing a better clarification and interpretation of the bioactivity values of each compound. The bioactivity values of the respective compounds were classified as follows: Active compounds (compounds with $pIC_{50} \geq 6$ or $IC_{50} \leq 1000$ nM), intermediate compounds (those with pIC_{50} values between 5 and 6 or IC_{50} between 1000 nM and 10000 nM), and inactive (pIC_{50} values ≤ 5 or IC_{50} above 10000 nM). Based on this classification, a total of 2591 compounds were found active, 153 intermediate compounds, and 12 inactive compounds, as shown in Supplementary Figure 1.

2.7. Exploratory data analysis for IC_{50} prediction

A total of 4 physicochemical descriptors are calculated for exploratory data analysis, as shown in Supplementary Figure 1.1C – 2.1F. The molecular descriptors calculated are as follows: molecular weight, octanol–water partition coefficient (LogP), number of hydrogen bond acceptors, and number of hydrogen bond donors, and 71.2% of the compounds passed the Lipinski rule of 5 while the remaining 28.8% did not pass. A univariate statistical analysis was performed to investigate individual molecular descriptors' distinct patterns and trends across three groups of compounds, as illustrated by the boxplot in Supplementary Figure 1. The analysis utilized the following descriptive statistical parameters: minimum (Min), first quartile (Q1), median, mean, third quartile (Q3), and maximum (Max).

2.8. Machine learning model for IC_{50} prediction

A total of 42 machine learning models were built for BTK. The modeling process involves calculating molecular descriptors, balancing and splitting the data, building machine learning models, comparing the evaluation metrics of the top models, and validating the best model. The process followed for retrieving and preprocessing the data is stated above.

2.9. Molecular descriptors for IC_{50} prediction

In this study, Morgan Fingerprint, a molecular fingerprint with radius = 2 and nBits = 2048 accessed through the Rdkit module [28], was used to convert the SMILE of each compound into 2048 descriptors.

2.10. Data balancing and splitting

The data were divided into two different sets viz: data set and training set. During machine learning modeling, the total dataset of 2756 was used, and the test and training set ratio was set on a scale of 20 and 80, respectively, where the training dataset contains 2204 data while the test dataset contains 552 data. The reproducibility of this model was maintained by setting the splitting and data balancing on a scale of 42 for all the procedures.

2.11. Machine learning model construction

Following data splitting, 41 machine learning regression models were constructed. Then, the top 10 model performances were evaluated and compared, as shown in Supplementary Figure 2, and the algorithm yielding the best performance was used (Random Forest Regressor).

2.12. Model validation

After the best model was chosen, a 10-fold cross-validation on the test set and the evaluation's meaning are presented in Figure 2.

3. Results and Discussions

3.1. Protein-protein interaction

BTK has long been recognized for its critical role in B-cell development and function [29]. However, recent research has unveiled its significance in a broader context, particularly in cancer biology. Supplementary Figure 4 illustrates BTK's intricate network of signaling partners, showing interactions between BTK and BLNK, SYK, and PLCG2, highlighting its crucial role in BCR signaling, elucidating why BTK inhibitors are effective against B-cell malignancies such as chronic lymphocytic leukemia and diffuse large B-cell lymphoma. BTK's interaction with LYN indicates its potential involvement in various blood cancers, while its association with VAV1 suggests possible roles in solid tumors and T-cell malignancies. The interaction with GRB2 indicates BTK's influence on cancer cell growth and survival mediated by growth factors. Further connections with MYD88, PIK3CD, and GTF2I implicate BTK in various oncogenic processes, including lymphoma progression and PI3K pathway abnormalities.

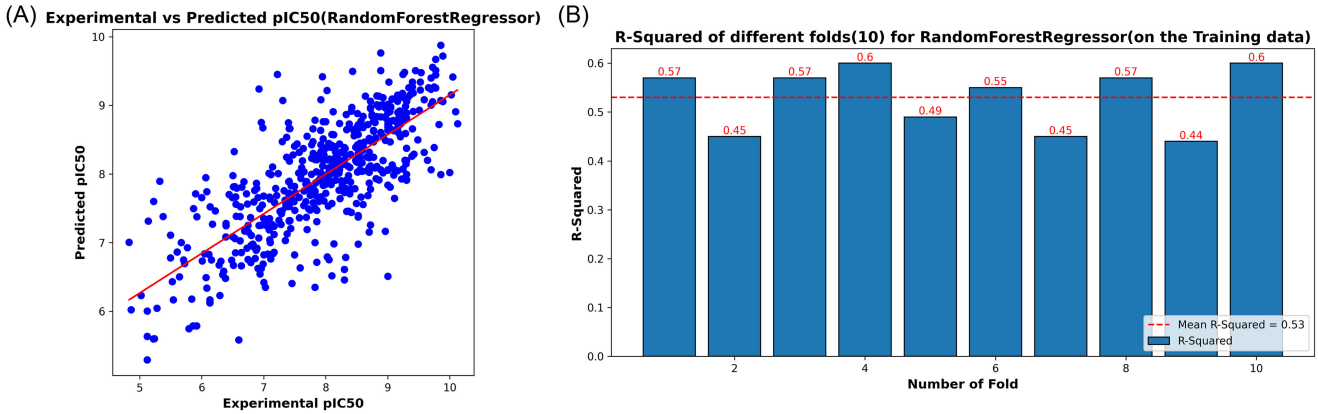


Figure 2. (A) A scatter plot showing the result of the predicted value of the best model (Random Forest Regressor) vs the actual value. (B) A bar plot showing the result of 10-fold cross-validation of the best model on the training set.

Table 1. Molecular docking and drug-likeness analysis of bioactive compounds from *M. cordata* targeting Bruton's tyrosine kinase

No	Compounds	M/W g/mol	Docking scores	Lipinski violation	Ghose violation	Veber violation	Egan violation	Muegge violation	Bioavailability score
1	Ibrutinib	440.5	−10.7	0	1	0	0	0	0.55
2	Chelidimerine	720.7	−13.7	2	4	0	1	3	0.17
3	Bocconarborine_A	736.8	−13.3	2	4	0	1	3	0.17
4	Bocconarborine_B	736.8	−12.9	2	4	0	1	3	0.17
5	Hydroxysanguinarine	347.3	−11.4	0	0	0	0	0	0.55
6	Dihydrosanguinarine	333.34	−11.1	0	0	0	0	0	0.55
7	6_methoxydihydrosanguinarine	363.4	−10.8	0	0	0	0	0	0.55
8	Norsanguinarine	317.3	−10.7	0	0	0	0	0	0.55
9	Corysamine	334.35	−10.7	0	0	0	0	0	0.55
10	6_acetonyldihydrosanguinarine	389.4	−10.6	0	0	0	0	0	0.55
11	Chelirubine	362.4	−10.4	0	0	0	0	0	0.55

3.2. Molecular docking and drug-likeness

The molecular docking study evaluated the binding affinities of compounds from *M. cordata* against BTK, and the results are presented in Table 1. This table also includes the drug-likeness profiles of the hit compounds. To further assess the potential of these compounds as drug candidates, the top-ranking compounds' pharmacokinetic properties and toxicity profiles were analyzed, as shown in Table 2.

3.3. ADMET

Proper evaluation of ADMET properties during pre-clinical and clinical stages could significantly enhance success rates in drug development and high-throughput screening processes [30, 31]. Consequently, to assess the potential of selected compounds as drug candidates, their pharmacokinetic and toxicological properties were evaluated using the Deep-PK webserver [21].

Firstly, all six compounds demonstrate similar absorption characteristics. The Caco-2 permeability values range from −5.51 to −5.19, indicating relatively low permeability across this cell line. Nevertheless, all compounds are predicted to be absorbed in the human intestine, suggesting potential oral bioavailability. Furthermore, the Madin-Darby Canine Kidney (MDCK) cell permeability results show a notable difference between the first four compounds (−4.44 to −4.51) and the last two (Chelidimerine and BocconarborineB), which have significantly higher values (205.33 and 226.31, respectively). This suggests that Chelidimerine and BocconarborineB may have enhanced permeability across MDCK cells.

Additionally, all compounds are predicted to be P-glycoprotein substrates, which may affect their absorption and distribution. Interestingly, Chelidimerine and BocconarborineB are also predicted to be P-glycoprotein inhibitors, potentially influencing drug-drug interactions. Moreover, skin permeability predictions show a similar pattern, with the first four compounds having low permeability (−2.37 to −1.29) and the last two having exceptionally high values (27484.11 and 30228.08).

In terms of distribution, the blood-brain barrier permeability predictions suggest limited central nervous system penetration for all compounds, with values ranging from −3.24 to −2.28. Meanwhile, the fraction unbound in human plasma is relatively consistent across all compounds (1.54 to 1.76), indicating moderate protein binding. Similarly, plasma protein binding predictions range from 72.08% to 85.25%, suggesting moderate to high protein binding, which may affect drug distribution and half-life. Furthermore, the steady-state volume of distribution predictions vary from 1.34 to 3.43, with Dihydrosanguinarine, Chelidimerine, and BocconarborineB showing higher values. This suggests these compounds may have a greater tendency to distribute into tissues.

Regarding metabolism, all compounds are predicted to be inhibitors of the Breast Cancer Resistance Protein and CYP1A2, and substrates of CYP1A2. This indicates potential for drug-drug interactions involving these pathways. However, the compounds show varying inhibition and substrate profiles for other cytochrome P450 enzymes. For instance, all compounds except Corysamine are predicted to inhibit CYP2C19, although none are substrates for this enzyme. In contrast, only Dihydrosanguinarine, Chelidimerine, and BocconarborineB are predicted to inhibit CYP2C9. Interestingly, the first four compounds demonstrate both inhibitory and substrate properties for CYP2D6. Furthermore, while all compounds are predicted to inhibit CYP3A4, only the last four serve as substrates for this enzyme. These diverse interactions with cytochrome P450 enzymes suggest complex metabolic profiles that may influence drug-drug interactions and overall pharmacokinetics. As for OATP transporters, only Chelidimerine and BocconarborineB are predicted to inhibit OATP1B3, while none inhibit OATP1B1.

In terms of excretion, clearance predictions range from 8.71 to 13.81 mL/min/kg, suggesting moderate clearance rates. Moreover, all compounds are predicted to have a half-life of less than 3 h, indicating relatively rapid elimination. It is worth noting that none of the compounds are predicted to inhibit the organic cation transporter 2.

Regarding toxicity, all compounds are predicted to be safe concerning human ether-à-go-go-related gene blockade, reducing the risk of cardiac arrhythmias. However, the first four compounds (Norsanguinarine, Hydroxysanguinarine, Dihydrosanguinarine, and Corysamine) are predicted to be AMES toxic and skin sensitizers, while Chelidimerine and BocconarborineB are not. Notably, Corysamine is the only compound predicted to be carcinogenic.

Additionally, all compounds except Corysamine are predicted to cause liver injury type I, and all compounds are predicted to cause liver injury type II. The maximum tolerated dose predictions range from 0.85 to 1.42 (mg/kg/day,), with Chelidimerine and BocconarborineB having the lowest values. On a positive note, all compounds are predicted to be biodegradable.

Lastly, the compounds show a consistent pattern in endocrine disruption predictions. All are predicted to be safe regarding the androgen receptor, estrogen receptor, aromatase, and PPARγ. However, all compounds are predicted to be toxic with respect to

Table 2. ADMET profiling of hit *M.cordata* bioactive compounds after molecular docking against BTK protein

Compounds	Norsanguinarine	Hydroxysanguinarine	Dihydrosanguinarine	Corysamine	Chelidimerine	Bocconarborine B
ABSORPTION						
Caco-2 Permeability	−5.26	−5.19	−5.19	−5.3	−5.48	−5.51
Human intestinal absorption (HIA)	Absorbed	Absorbed	Absorbed	Absorbed	Absorbed	Absorbed
Madin-Darby Canine Kidney	−4.44	−4.43	−4.49	−4.51	205.33	226.31
P-Glycoprotein Inhibitor	Non-Inhibitor	Non-Inhibitor	Non-Inhibitor	Non-Inhibitor	Inhibitor	Inhibitor
P-Glycoprotein Substrate	Substrate	Substrate	Substrate	Substrate	Substrate	Substrate
Skin Permeability	−1.76	−1.29	−2.37	−2.03	27484.11	30228.08
DISTRIBUTION						
Blood-Brain Barrier (Central Nervous System)	−2.28	−2.41	−2.44	−3.09	−3.24	−3.04
Fraction Unbound (Human)	1.74	1.76	1.55	1.54	1.73	1.7
Plasma protein binding (PPB)	82.8	85.25	75.73	77.81	73.05	72.08
Steady State Volume of Distribution	1.51	1.34	2.83	1.51	3.43	2.96
METABOLISM						
Breast Cancer Resistance Protein	Inhibitor	Inhibitor	Inhibitor	Inhibitor	Inhibitor	Inhibitor
CYP1A2 inhibitor	Inhibitor	Inhibitor	Inhibitor	Inhibitor	Inhibitor	Inhibitor
CYP1A2 substrate	Substrate	Substrate	Substrate	Substrate	Substrate	Substrate
CYP2C19 inhibitor	Inhibitor	Inhibitor	Inhibitor	Non-inhibitor	Inhibitor	Inhibitor
CYP2C19 substrate	Non-substrate	Non-substrate	Non-substrate	Non-substrate	Non-substrate	Non-substrate
CYP2C9 inhibitor	Non-Inhibitor	Non-Inhibitor	Inhibitor	Non-inhibitor	Inhibitor	Inhibitor
CYP2C9 substrate	Substrate	Substrate	Substrate	Substrate	Non-substrate	Non-substrate
CYP2D6 inhibitor	Inhibitor	Inhibitor	Inhibitor	Inhibitor	Non-inhibitor	Non-inhibitor
CYP2D6 substrate	Substrate	Substrate	Substrate	Substrate	Non-substrate	Non-substrate
CYP3A4 inhibitor	Inhibitor	Inhibitor	Inhibitor	Inhibitor	Inhibitor	Inhibitor
CYP3A4 substrate	Non-substrate	Non-substrate	Substrate	Substrate	Substrate	Substrate
OATP1B1	Non-inhibitor	Non-inhibitor	Non-inhibitor	Non-inhibitor	Non-inhibitor	Non-inhibitor
OATP1B3	Non-inhibitor	Non-inhibitor	Non-inhibitor	Non-inhibitor	Inhibitor	Inhibitor
EXCRETION						
Clearance	9.49	8.71	13.81	10.51	11.82	11.51
T1/2	Half-Life < 3 hs	Half-Life < 3 hs	Half-Life < 3 hs	Half-Life < 3 hs	Half-Life < 3 hs	Half-Life < 3 hs
Organic Cation Transporter 2	Non-inhibitor	Non-inhibitor	Non-inhibitor	Non-inhibitor	Non-inhibitor	Non-inhibitor
TOXICITY						
hERG Blockers	Safe	Safe	Safe	Safe	Safe	Safe
AMES Toxicity	Toxic	Toxic	Toxic	Toxic	Safe	Safe
Skin Sensitization	Toxic	Toxic	Toxic	Toxic	Safe	Safe
Carcinogenicity	Safe	Safe	Safe	Toxic	Safe	Safe
Liver Injury I	Toxic	Toxic	Toxic	Safe	Toxic	Toxic
Liver Injury II	Toxic	Toxic	Toxic	Toxic	Toxic	Toxic
Maximum Tolerated Dose	1.42	1.27	1.41	1.14	0.92	0.85
Biodegradation	Safe	Safe	Safe	Safe	Safe	Safe
ENDOCRINE DISRUPTION						

Androgen receptor	Safe	Safe	Safe	Safe	Safe	Safe
Estrogen receptor	Safe	Safe	Safe	Safe	Safe	Safe
Aryl hydrocarbon Receptor	Toxic	Toxic	Toxic	Toxic	Toxic	Toxic
Glucocorticoid receptor (GR)	Toxic	Toxic	Toxic	Toxic	Safe	Safe
Aromatase	Safe	Safe	Safe	Safe	Safe	Safe
Peroxisome Proliferator-Activated Receptor gamma (PPARγ)	Safe	Safe	Safe	Safe	Safe	Safe
Androgen Receptor (AR) Ligand-Binding Domain (LBD)	Safe	Safe	Safe	Safe	Toxic	Toxic
Estrogen receptor (ER) Ligand-Binding Domain (LBD)	Safe	Safe	Safe	Safe	Safe	Safe

the aryl hydrocarbon receptor. The first four compounds are predicted to be toxic for the glucocorticoid receptor, while Chelidimerine and BocconarborineB are predicted to be toxic for the androgen receptor ligand binding domain.

Following molecular docking analysis, the binding interactions of the identified hit compounds were evaluated in comparison to the reference drug, ibrutinib, using PyMol and LigPlot software tools. This analysis unveiled the robust binding of the investigated compounds within the active sites of BTK (Figure 3). The study examined each ligand cluster, identifying specific amino acid residues interacting with each ligand, characterizing hydrogen bonds (H-bonds) formed, including their lengths in Angstroms (Å), and recognizing hydrophobic interactions between the ligands and the protein. A detailed breakdown of these interactions for each compound is provided in Table 3.

This approach elucidated the molecular basis of the observed binding affinities for the hit compounds relative to ibrutinib, offering insights into their potential effectiveness as BTK inhibitors.

3.4. Half maximal inhibitory concentration (IC_{50}) prediction

A compound's half-maximal inhibitory concentration ($1C_{50}$) is the potency to block the potassium human ether-à-go-go-related gene channels [32]. The half-maximal inhibitory concentration is shown in Table 4, with ibrutinib having the best IC_{50} of 0.68 nM. Other compounds, such as Bocconarborine B, Bocconarborine A, and 6-acetonyldihydrochelerythrine, has an IC_{50} values of 48.98, 48.98, and 38.90 nM, respectively. Other lead compounds from Table 4 have an IC_{50} value of more than 100 nM.

3.5. Molecular dynamics (MD) simulations study

3.5.1. Root mean square deviation (RMSD) and root mean square fluctuation (RMSF)

The RMSD calculation from the atoms in the complex backbone is crucial for analyzing molecular dynamic trajectory equilibration [33]. RMSD backbone measurement for two complexes provides conformational stability information. The RMSD values of the protein-ligand complex of Bocconarborine B, Chelidimerine, Corysamine, Dihydrosanguinarine, Hydroxysanguinarine, Norsanguinarine, and ibrutinib are 0.13, 0.15, 0.2, 0.14, 0.18, 0.15, 0.24 nm, respectively as shown in Figure 4. Corysamine had a spread of data when compared to other bioactive compounds. This result validates the stability of Dihydrosanguinarine and Bocconarborine B with an acceptable RMSD value of 0.5 nm and 0.39 nm, respectively.

The RMSF is calculated to estimate the average fluctuations of protein residues during MD simulation [34]; thus, it allows the thorough examination of the flexibility of the protein's backbones. The average value of the RMSF of the ligand-protein complexes Bocconarborine B, Chelidimerine, Corysamine, Dihydrosanguinarine, Hydroxysanguinarine, Norsanguinarine, and ibrutinib is 0.28, 0.44, 0.21, 0.35, 0.22, 0.23, and 0.25 nm respectively. The average of the ligand-protein complex depicts the stability of the respective complexes formed.

3.5.2. Hydrogen bond

The intramolecular hydrogen bonds are a crucial factor in determining the stability of protein-ligand complexes [35]. The intramolecular hydrogen bond interaction shows the binding of the hit compounds and the reference drug within the binding pocket of BTK. The average number of hydrogen bonds formed in the BTK-ibrutinib was found to be 4, while other compounds,

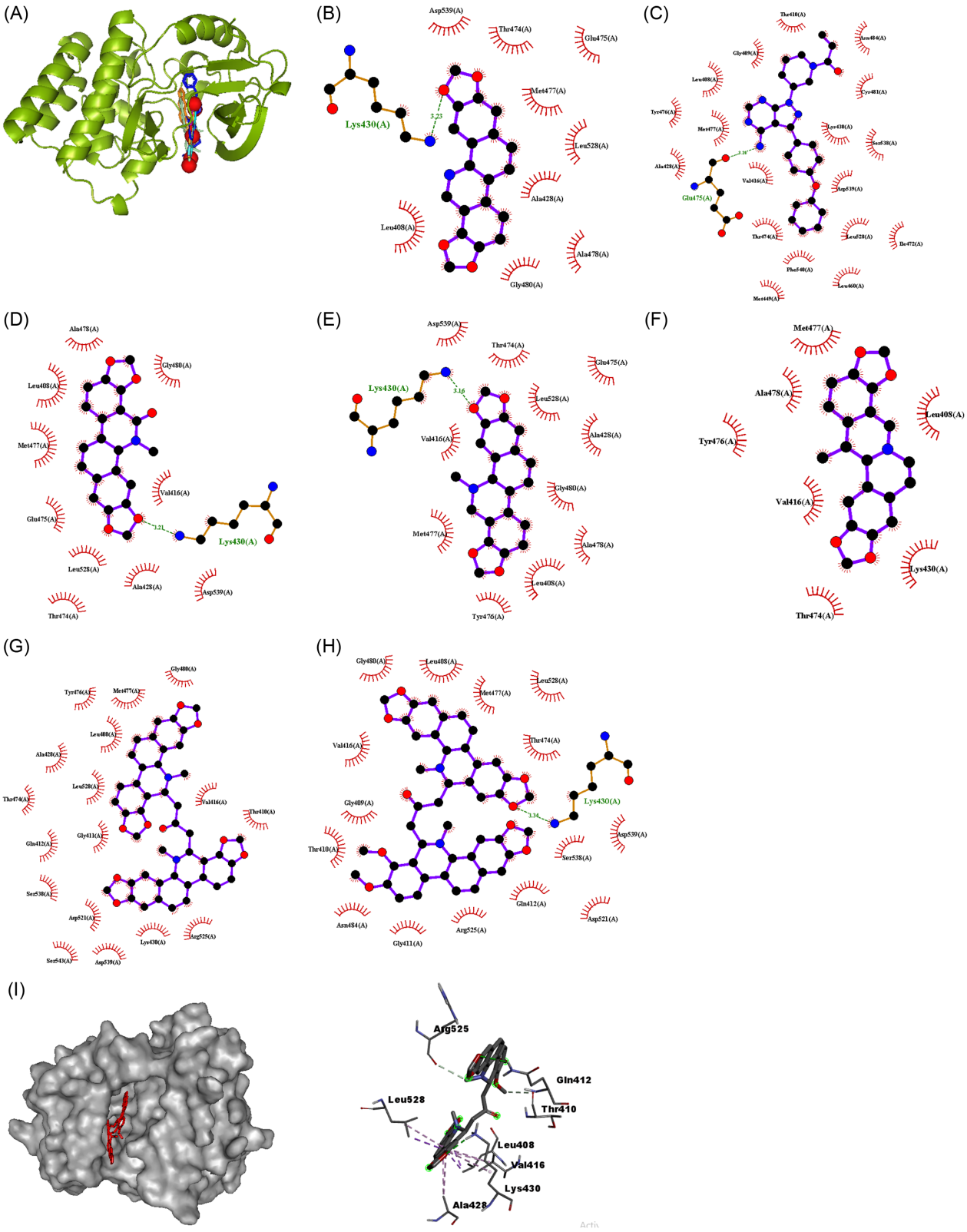


Figure 3. Molecular docking analysis of *M. Cordata* phytocompounds and ibrutinib with BTK. (A) Amino acid interactions of phytocompounds in the substrate binding cavity of BTK, shown in a solvent-accessible surface view (B–H) 2D molecular docking interaction analyses of the following compounds against BTK: (B) Norsanguinarine. (C) Ibrutinib. (D) Hydroxysanguinarine. (E) Dihydrosanguinarine. (F) Corysamine. (G) Chelidimerine. (H) Bocconarborine B. (I) 3D structure interaction of Boconarborine B on the surface of the active site of the BTK protein.

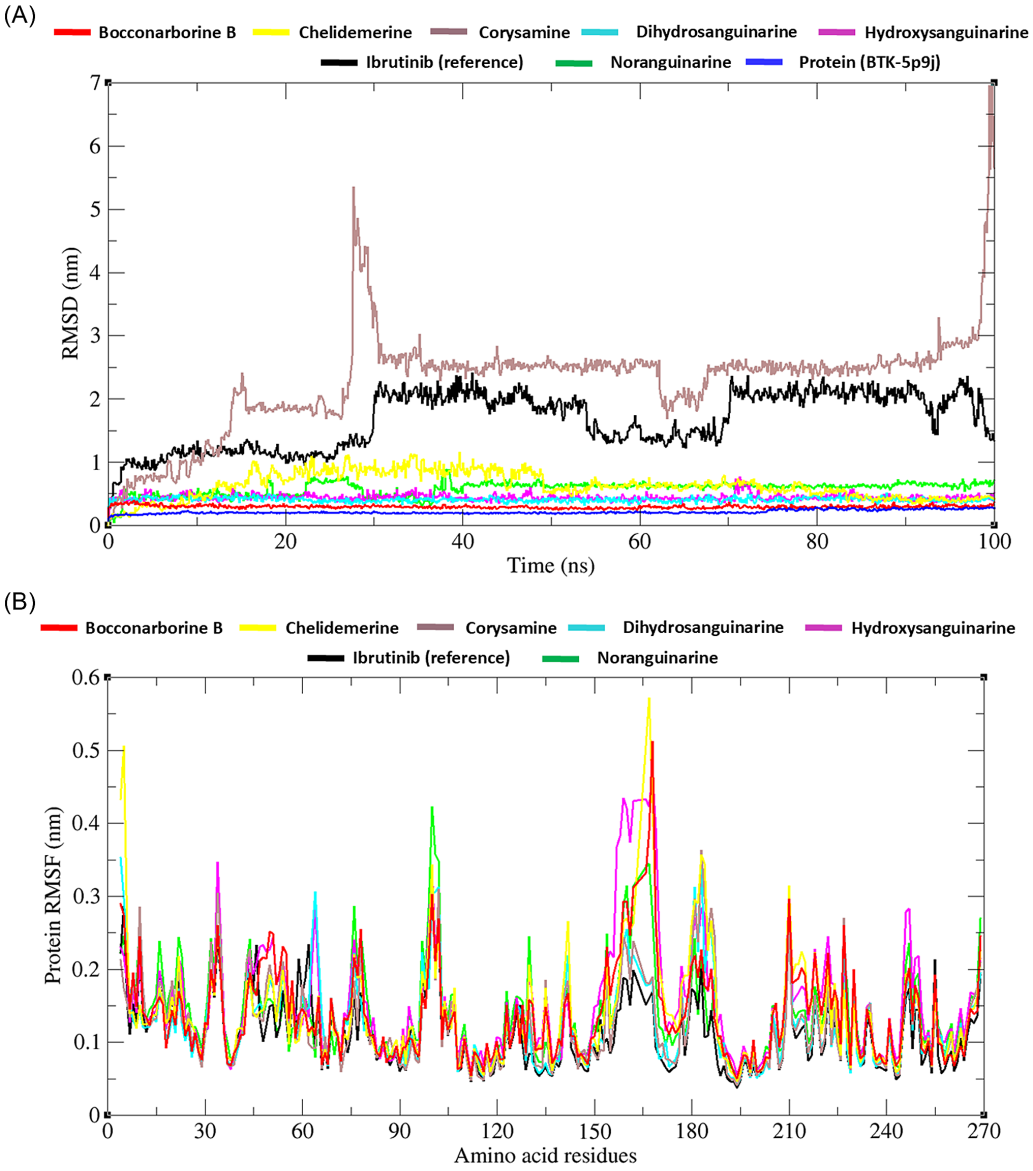


Figure 4. RMSD and RMSF of the selected ligands in complex with the protein. (A) RMSD plot of the selected ligands in complex with the protein (BTK). (B) Protein RMSF plot for the selected ligands in comparison to the reference (black).

such as Noranguinarine and Bocconarborine B, were found to be three as shown in Figure 5 (information that supports Figure 5 could be available in the Supplementary Figure 3). The minimum fluctuations in the intramolecular hydrogen bond of Noranguinarine suggest a more consistent and stable interaction when compared to the standard drug.

3.5.3. Radius of gyration (ROG) and solvent-accessible surface area (SASA)

The structural compactness was calculated using a dynamic approach to ascertain the binding and unbinding process when running the molecular dynamic simulation. The radius of gyration was used to determine the structural compactness of the protein-ligand complexes. The average radius of gyration value for Bocconarborine B, Chelidimerine, Corysamine, Dihydrosanguinarine, Hydroxysanguinarine, Norsanguinarine, and ibrutinib is 1.87, 1.86, 1.88, 1.9, 1.91. 1.91, and 1.9 nm, respectively as shown in Figure 6. Bocconarborine B and Chelidimerine have the lowest average values, indicating they are the most rigid ligands.

SASA involves measuring the exposed surface area of protein-ligand molecules, achieved by surface area measurement of the interaction between the involved molecules and the solvent molecules [36]. Also, SASA works by analyzing the exposure of various interactions between ligands and protein with the time changes involved. The complexes have the SASA analysis with an average of 133–135 nm^2 for all the ligands, except for ibrutinib and Noranguinarine, with an average of 136 and 137 nm^2, respectively.

3.6. MMPBSA

The binding free energy of BTK and ibrutinib with other hits compounds was assessed using the MMPBSA. The result as shown in Table 5 reveals that all the hit compounds have considerable binding energy however, Bocconarborine B has the best binding energy of −30.94 kcal/mol compared to ibrutinib, with a binding energy of −22.46 kcal/mol. Other hit compounds viz: Chelidemerine, Corysamine, Dihydrosanguinarine, hydroxysanguinarine, and Nonsanguinarine have the binding

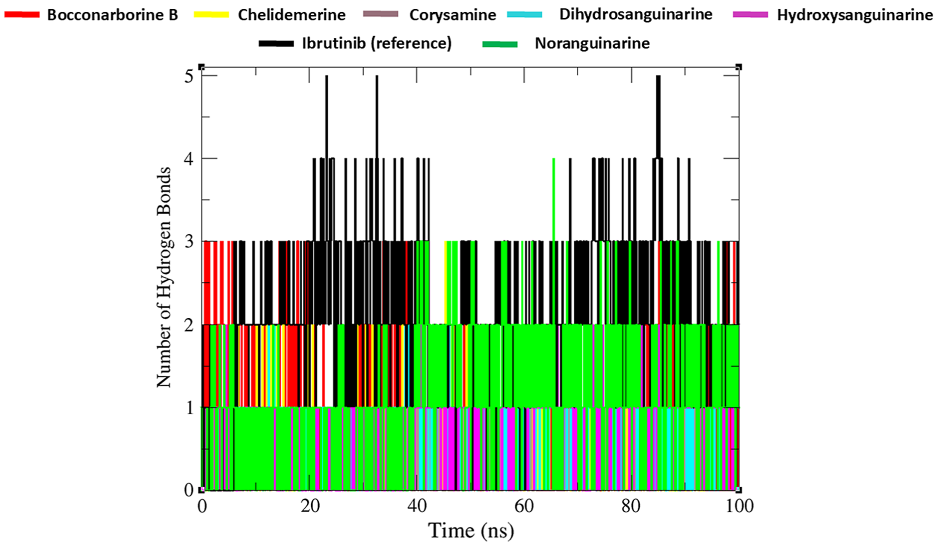


Figure 5. Graph depicting the number of intermolecular H-bonds formed between protein-ligand complexes of selected ligands and the reference drug (black).

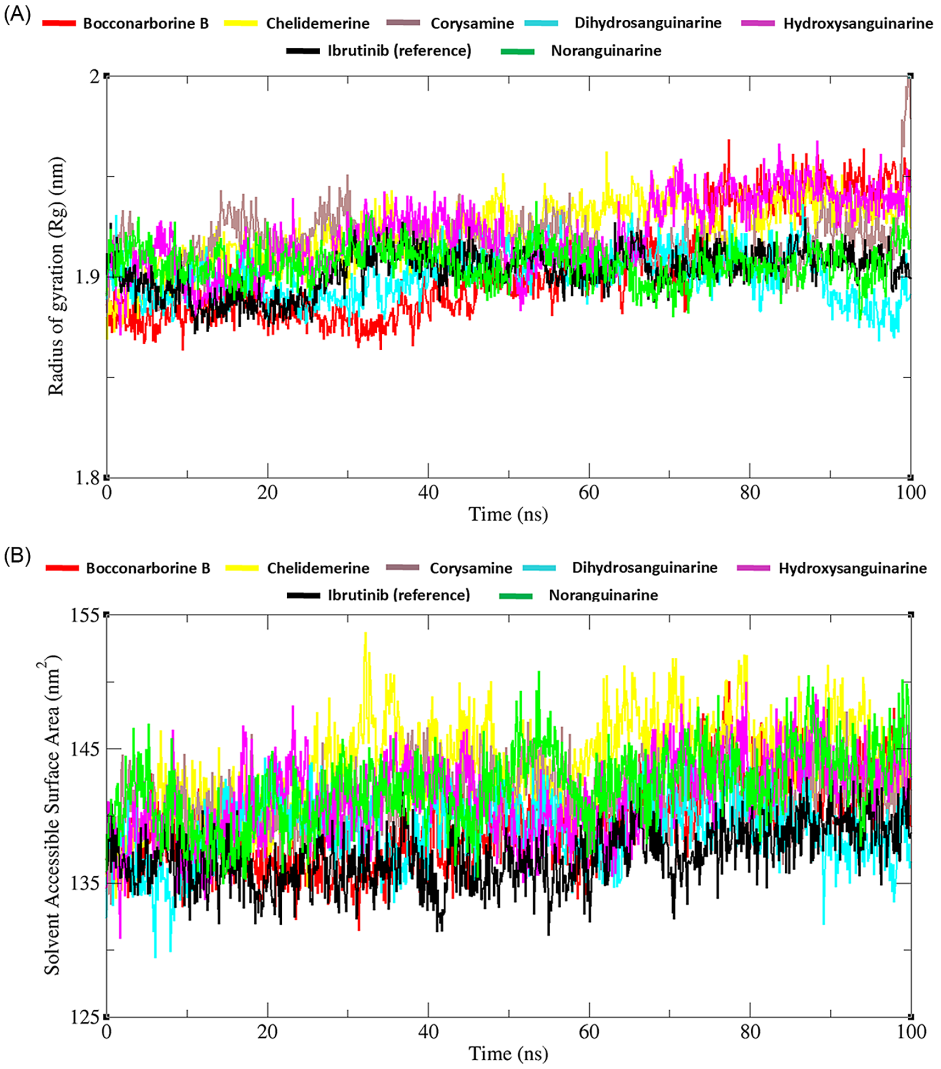


Figure 6. The radius of gyration plot and the solvent-accessible surface area. (A) Radius of gyration plot. (B) Solvent-accessible surface area.

Table 3. Molecular interaction analysis of BTK with norsanguinarine, ibrutinib, hydroxysanguinarine, dihydrosanguinarine, corysamine, chelidimerine, and bocconarborine B

Ligand-Protein complexes	Residues forming H-bonds	Residues involved in hydrogen bond formation (Bond distances Å)	Hydrophobic interacting residues
BTK_Norsanguinarine	1	Lys430(3.23)	Leu 408, Gly 480, Ala478, Ala428, Leu528, Met477, Glu475, Thr474, Asp539
BTK_ Ibrutinib	1	Glu475(3.20)	Thr410, Gly409, Leu408, Met477, Ala428, Tyr476, Val416, Thr474, Phe540, Met449, Lys430,Cys481, Asn484
BTK_ Hydroxysanguinarine	1	Lys430(3.21)	Val416, Gly480, Ala478, Leu408, Met477, Glu475, Leu528,Ala428, Asp539
BTK_ Dihydrosanguinarine	1	Lys430(3.16)	Asp539, Thr474, Leu528, Ala428, Gly480, Ala478, Leu408, Tyr476, Met477, Val416
BTK_Corysamine	0	None	Met477, Ala478, Tyr476, Val416, Thr474, Lys430, Leu408
BTK_Chelidimerine	0	None	Val416, Thr410, Arg525, Lys430, Gly480, Met477, Leu408, Leu528, Gly411, Asp521, Asp539, Tyr476, Ala428, Thr474
BTK_ Bocconarborine B	1	Lys430(3.34)	Gly480, Val416, Gly409, Thr410, Asn484, Gly411, Arg525, Gln412,Asp521, Ser538, Asp539,

Table 4. Inhibitory concentration prediction of ibrutinib and other hit compounds

S/N	Compound name	pIC_{50}	IC_{50}(nM)	MW	HBA	HBD	LogP
1	Bocconarborine B	7.31	48.98	736.777	11	0	8.56160
2	Chelidimerine	6.97	107.15	720.734	11	0	8.27310
3	Corysamine	6.76	173.78	334.351	4	0	3.11622
4	Dihydrosanguinarine	6.94	114.82	333.343	5	0	3.91400
5	Hydroxysanguinarine	7.00	100.00	347.326	6	0	3.30230
6	Norsanguinarine	6.74	181.97	317.300	5	0	3.99860
7	Ibrutinib (Reference)	9.17	0.68	440.507	7	1	4.21730
8	Bocconarborine A	7.31	48.98	736.777	11	0	8.56160
9	6-acetonyldihydrochelerythrine	7.41	38.90	405.450	6	0	4.72270
10	Chelirubine	6.64	229.09	362.361	5	0	3.43670
11	6-methoxydihydrosanguinarine	6.68	208.93	363.369	6	0	

Table 5. Energy contributions of different protein-ligand complexes in MMPBSA assay

S.No.	Compound ID	$\Delta G_{VDWaals}$ (kcal/mol)	ΔG_{EEL} (kcal/mol)	ΔG_{EPB} (kcal/mol)	$\Delta G_{ENPolar}$ (kcal/mol)	ΔG_{gas} (kcal/mol)	ΔG_{solv} (kcal/mol)	ΔG_{total} (kcal/mol)
1.	Bocconarborine B	−63.76	−30.08	68.77	−5.86	−93.84	62.90	−30.94
2.	Chelidemerine	−46.80	−16.75	47.59	−5.33	−63.55	42.26	−21.28
3.	Corysamine	−17.63	−14.89	26.13	−2.12	−32.52	24.01	−8.50
4.	Dihydrosanguinarine	−34.50	−15.76	37.14	−3.36	−50.26	33.78	−16.49
5.	Hydroxysanguinarine	−33.29	−10.08	32.50	−3.14	−43.37	29.36	−14.01
6.	Norsanguinarine	−35.31	−13.72	35.01	−3.59	−49.03	31.42	−17.61
7.	Ibrutinib (reference)	−6.53	−424.50	410.97	−2.41	−431.02	408.57	−22.46

energy of −21.24 kcal/mol, −8.50 kcal/mol, −16.49 kcal/mol, −14.01 kcal/mol, and −17.61 kcal/mol respectively.

4. Conclusion

The invention of ibrutinib greatly improved the BTK inhibitors' selectivity. However, it has been reported that ibrutinib had atrial fibrillation as a side effect, which is one of the crucial reasons for the development of better BTK inhibitors. These adverse reactions have been attributed to its off-target effects when ibrutinib is combined with other kinases, such as those in the EGFR and TEC families. Therefore, finding small molecule inhibitors of BTK with low cytotoxicity and high efficacy is crucial to developing the cure for various intricacies associated with BTK, including B-cell malignant tumors. Thus, this study identified promising hit compounds from *Macleaya cordata* that can inhibit BTK, considering the challenge of acquiring resistance mutation to each of the classes of BTK inhibitors, thus breaking a significant barrier to treating malignancies associated with BTK mutation. We show that Bocconarborine B has a high binding affinity (−30.94 kcal/mol) for BTK compared to ibrutinib (−22.46 kcal/mol). This study is subject to further validation via in vivo and in vitro studies.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support this work are available upon reasonable request to the corresponding author.

Author Contribution Statement

Ayobami Fidelix: Conceptualization, Methodology, Formal analysis, Resources, Writing – original draft, Writing – review & editing, Supervision. **Tomilola Akingbade:** Conceptualization, Methodology, Formal analysis, Resources, Writing – original draft, Writing – review & editing, Project administration. **Jatin Jangra:** Software, Validation, Resources. **Babatunde Olabuntu:** Software, Validation, Resources. **Olutola Adeyemo:** Writing – review & editing, Visualization. **Juwon Akingbade:** Investigation, Data curation, Writing – review & editing.

References

[1] Buske, C., Jurczak, W., Salem, J. E., & Dimopoulos, M. A. (2023). Managing Waldenström's macroglobulinemia with BTK inhibitors. *Leukemia*, *37*(1), 35–46. https://doi.org/10.1038/s41375-022-01732-9

[2] Skånland, S. S., Karlsen, L., & Taskén, K. (2020). B cell signalling pathways—New targets for precision medicine in chronic lymphocytic leukaemia. *Scandinavian Journal of Immunology*, *92*(5), e12931. https://doi.org/10.1111/sji.12931

[3] Cool, A., Nong, T., Montoya, S., & Taylor, J. (2024). BTK inhibitors: Past, present, and future. *Trends in Pharmacological Sciences*, *45*(8), 691–707. https://doi.org/10.1016/j.tips.2024.06.006

[4] Satterthwaite, A. B. (2018). Bruton's tyrosine kinase, a component of B cell signaling pathways, has multiple roles in the pathogenesis of lupus. *Frontiers in Immunology*, *8*, 1986. https://doi.org/10.3389/fimmu.2017.01986

[5] Rada, M., Barlev, N., & Macip, S. (2018). BTK: A two-faced effector in cancer and tumour suppression. *Cell Death & Disease*, *9*(11), 1064. https://doi.org/10.1038/s41419-018-1122-8

[6] Nawaratne, V., Sondhi, A. K., Abdel-Wahab, O., & Taylor, J. (2024). New means and challenges in the targeting of BTK. *Clinical Cancer Research*, *30*(11), 2333–2341. https://doi.org/10.1158/1078-0432.CCR-23-0409

[7] Woyach, J. A., Furman, R. R., Liu, T. M., Ozer, H. G., Zapatka, M., Ruppert, A. S., . . . , & Byrd, J. C. (2014). Resistance mechanisms for the Bruton's tyrosine kinase inhibitor ibrutinib. *New England Journal of Medicine*, *370*(24), 2286–2294. https://doi.org/10.1056/NEJMoa1400029

[8] Wang, E., Mi, X., Thompson, M. C., Montoya, S., Notti, R. Q., Afaghani, J., . . . , & Abdel-Wahab, O. (2022). Mechanisms of resistance to noncovalent Bruton's tyrosine kinase inhibitors. *New England Journal of Medicine*, *386*(8), 735–743. https://doi.org/10.1056/NEJMoa2114110

[9] Pal Singh, S., Dammeijer, F., & Hendriks, R. W. (2018). Role of Bruton's tyrosine kinase in B cells and malignancies. *Molecular Cancer*, *17*(1), 57. https://doi.org/10.1186/s12943-018-0779-z

[10] Jin, H., Wang, L., & Bernards, R. (2023). Rational combinations of targeted cancer therapies: Background, advances and challenges. *Nature Reviews Drug Discovery*, *22*(3), 213–234. https://doi.org/10.1038/s41573-022-00615-z

[11] Fidelix, A., Victor, A. T., Jangra, J., Adeyemo, O., & Blessing, A. D. (2024). Flavonoids from piper betle: A computational approach to target NRAS-driven melanoma. *World Journal of Pharmacy and Pharmaceutical Sciences*, *13*(8), 43–57. https://www.wjpps.com/Wjpps_controller/abstract_id/21238

[12] Zhang, H., Zhou, W., Chen, Y., Xu, H., Hou, D., Lv, S., . . . , & Yang, L. (2022). The tolerance, absorption, and transport characteristics of *Macleaya cordata* in relation to lead, zinc, cadmium, and copper under hydroponic conditions. *Applied Sciences*, *12*(19), 9598. https://doi.org/10.3390/app12199598

[13] Sun, M., Zhong, X., Zhou, L., Xu, Z., Huang, P., & Zeng, J. (2022). Plant–microbe hybrid synthesis provides new insights for the efficient use of *Macleaya cordata*. *World Journal of Microbiology and Biotechnology*, *38*(6), 110. https://doi.org/10.1007/s11274-022-03295-4

[14] Jiang, C., Yang, H., Chen, X., Qiu, S., Wu, C., Zhang, B., & Jin, L. (2020). *Macleaya cordata* extracts exert antiviral effects in newborn mice with rotavirus-induced diarrhea via inhibiting the JAK2/STAT3 signaling pathway. *Experimental and Therapeutic Medicine*, *20*(2), 1137–1144. https://doi.org/10.3892/etm.2020.8766

[15] Zhang, H., Hu, L., Du, X., Sun, X., Wang, T., & Mu, Z. (2023). Physiological and molecular response and tolerance of *Macleaya cordata* to lead toxicity. *BMC Genomics*, *24*(1), 277. https://doi.org/10.1186/s12864-023-09378-2

[16] Lin, L., Liu, Y. C., Huang, J. L., Liu, X. B., Qing, Z. X., Zeng, J. G., & Liu, Z. Y. (2018). Medicinal plants of the genus *Macleaya* (*Macleaya cordata, Macleaya microcarpa*): A review of their phytochemistry, pharmacology, and toxicology. *Phytotherapy Research*, *32*(1), 19–48. https://doi.org/10.1002/ptr.5952

[17] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ..., & Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, *28*(1), 235–242. https://doi.org/10.1093/nar/28.1.235

[18] Bender, A. T., Gardberg, A., Pereira, A., Johnson, T., Wu, Y., Grenningloh, R., ..., & Liu-Bujalski, L. (2017). Ability of Bruton's tyrosine kinase inhibitors to sequester Y551 and prevent phosphorylation determines potency for inhibition of Fc receptor but not B-cell receptor signaling. *Molecular Pharmacology*, *91*(3), 208–219. http://dx.doi.org/10.1124/mol.116.107037

[19] Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785–2791. https://doi.org/10.1002/jcc.21256

[20] Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, *7*(1), 42717. https://doi.org/10.1038/srep42717

[21] Myung, Y., de Sá, A. G., & Ascher, D. B. (2024). Deep-PK: Deep learning for small molecule pharmacokinetic and toxicity prediction. *Nucleic Acids Research*, *52*(W1), W469–W475. https://doi.org/10.1093/nar/gkae254

[22] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

[23] Jo, S., Kim, T., Iyer, V. G., & Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, *29*(11), 1859–1865. https://doi.org/10.1002/jcc.20945

[24] Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5

[25] Sourceforce. (n.d.). *QtGrace files*. Retrieved from: https://sourceforge.net/projects/qtgrace/files/

[26] Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z., & Hou, T. (2019). End-point binding free energy calculation with MM/PBSA and MM/GBSA: Strategies and applications in drug design. *Chemical Reviews*, *119*(16), 9478–9508. https://doi.org/10.1021/acs.chemrev.9b00055

[27] Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., ..., & Leach, A. R. (2017). The ChEMBL database in 2017. *Nucleic Acids Research*, *45*(D1), D945–D954. https://doi.org/10.1093/nar/gkw1074

[28] Read the Docs. (n.d.). *The RDKit documentation*. Retrieved from: https://readthedocs.org/projects/rdkit/

[29] van Norman, G. A. (2020). Limitations of animal studies for predicting toxicity in clinical trials: Part 2: Potential alternatives to the use of animals in preclinical trials. *JACC: Basic to Translational Science*, *5*(4), 387–397. https://doi.org/10.1016/j.jacbts.2020.03.010

[30] Li, D., Artursson, P., Avdeef, A., Benet, L. Z., Houston, J. B., Kansy, M., ..., & Sugano, K. (2020). The critical role of passive permeability in designing successful drugs. *ChemMedChem*, *15*(20), 1862–1874. https://doi.org/10.1002/cmdc.202000419

[31] Nündel, K., Busto, P., Debatis, M., & Marshak-Rothstein, A. (2013). The role of Bruton's tyrosine kinase in the development and BCR/TLR-dependent activation of AM14 rheumatoid factor B cells. *Journal of Leukocyte Biology*, *94*(5), 865–875. https://doi.org/10.1189/jlb.0313126

[32] Gomis-Tena, J., Brown, B. M., Cano, J., Trenor, B., Yang, P. C., Saiz, J., ..., & Romero, L. (2020). When does the IC_{50} accurately assess the blocking potency of a drug? *Journal of Chemical Information and Modeling*, *60*(3), 1779–1790. https://doi.org/10.1021/acs.jcim.9b01085

[33] Panda, S. M., & Tripathy, U. (2024). *In silico* screening and identifying phytoconstituents of *Withania somnifera* as potent inhibitors of BRCA1 mutants: A therapeutic against breast cancer. *International Journal of Biological Macromolecules*, *282*, 136977. https://doi.org/10.1016/j.ijbiomac.2024.136977

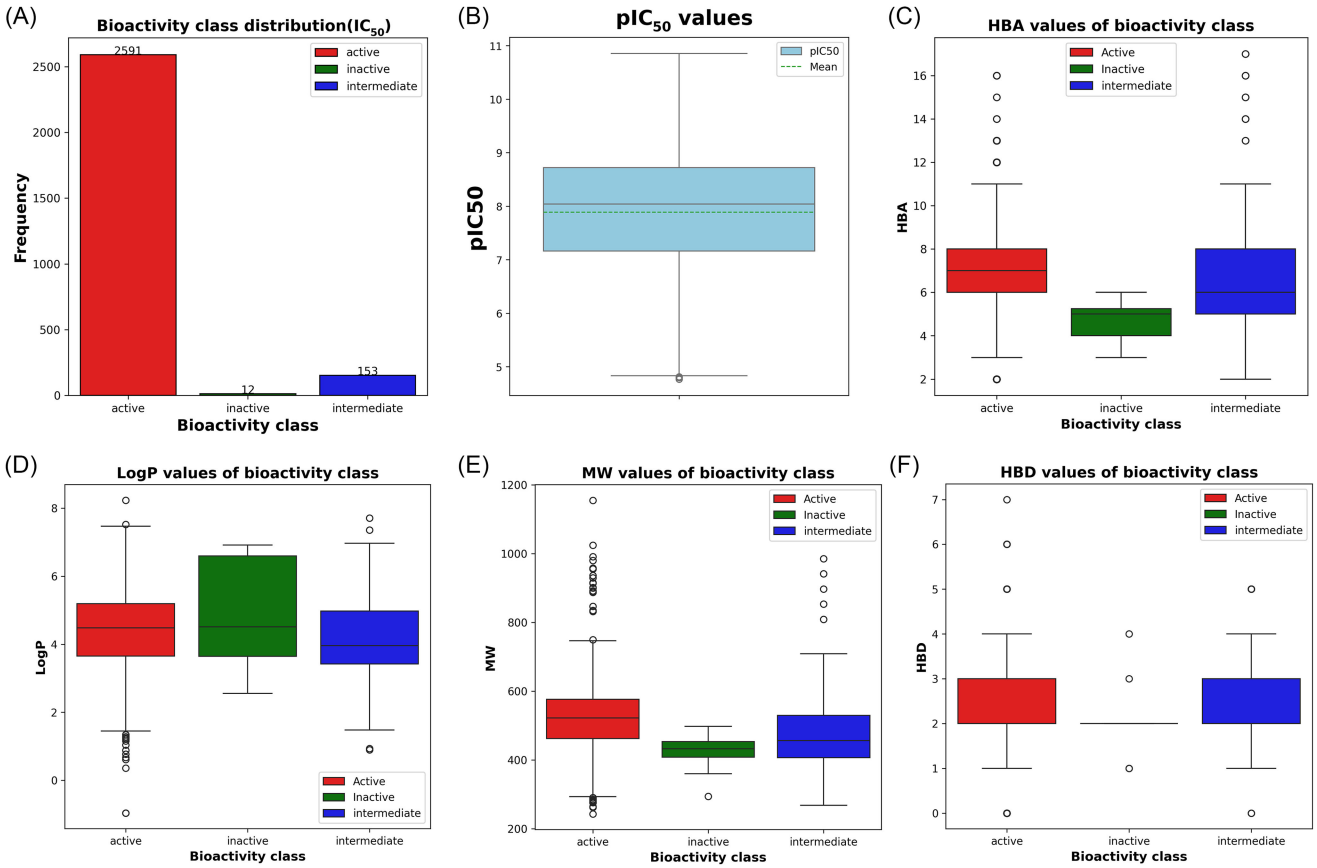
[34] Boroujeni, M. B., Dastjerdeh, M. S., Shokrgozar, M., Rahimi, H., & Omidinia, E. (2021). Computational driven molecular dynamics simulation of keratinocyte growth factor behavior at different pH conditions. *Informatics in Medicine Unlocked*, *23*, 100514. https://doi.org/10.1016/j.imu.2021.100514

[35] Weiss, M. S., Brandl, M., Sühnel, J., Pal, D., & Hilgenfeld, R. (2001). More hydrogen bonds for the (structural) biologist. *Trends in Biochemical Sciences*, *26*(9), 521–523. https://doi.org/10.1016/S0968-0004(01)01935-1

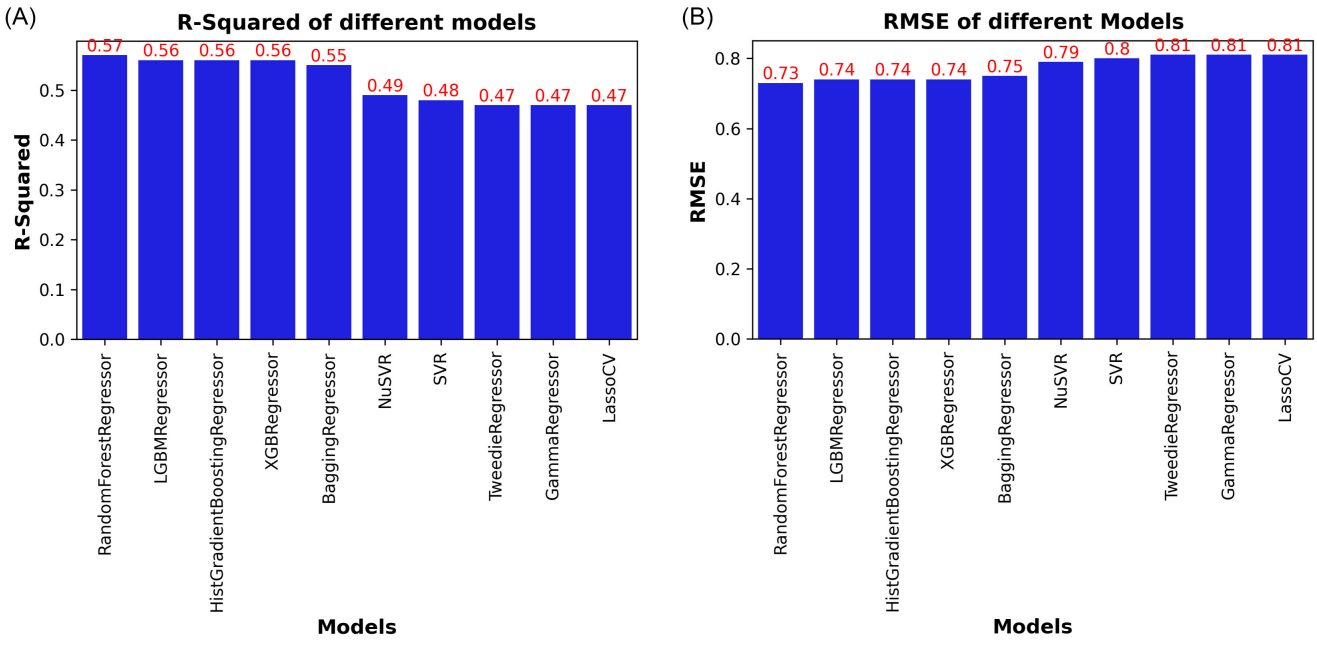
[36] Dhiman, A., & Purohit, R. (2023). Identification of potential mutational hotspots in serratiopeptidase to address its poor pH tolerance issue. *Journal of Biomolecular Structure and Dynamics*, *41*(18), 8831–8843. https://doi.org/10.1080/07391102.2022.2137699

How to Cite: Fidelix, A., Akingbade, T., Jangra, J., Olabuntu, B., Adeyemo, O., & Akingbade, J. (2025). In Silico Study and Validation of Natural Compounds Derived from *Macleaya cordata* as a Potent Inhibitor for BTK. *Medinformatics*, *2*(1), 22–35. https://doi.org/10.47852/bonviewMEDIN52024239

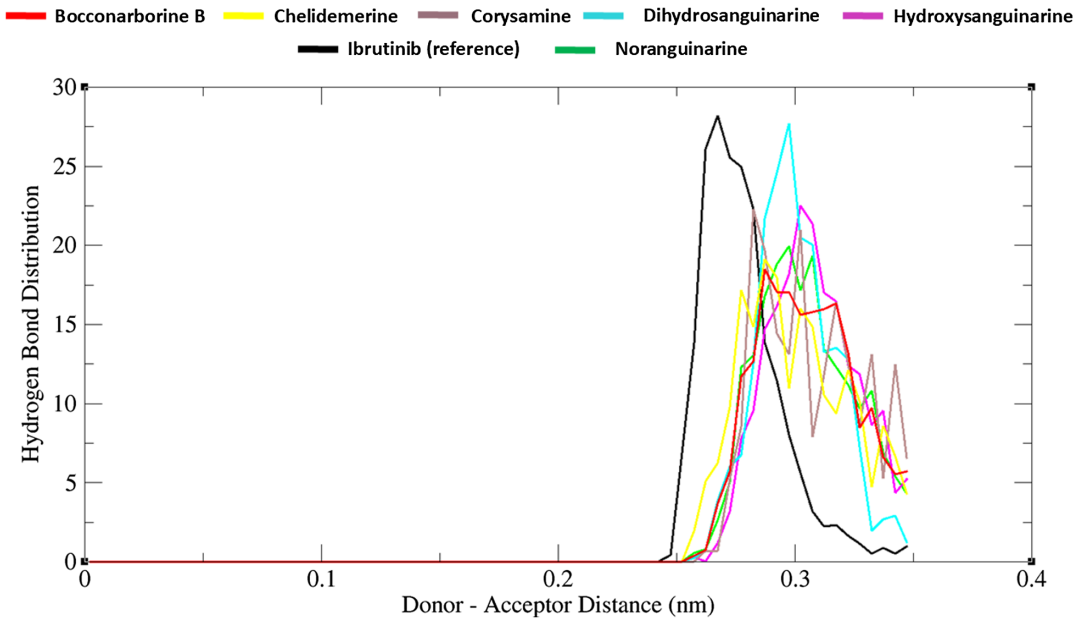
Supplementary Information



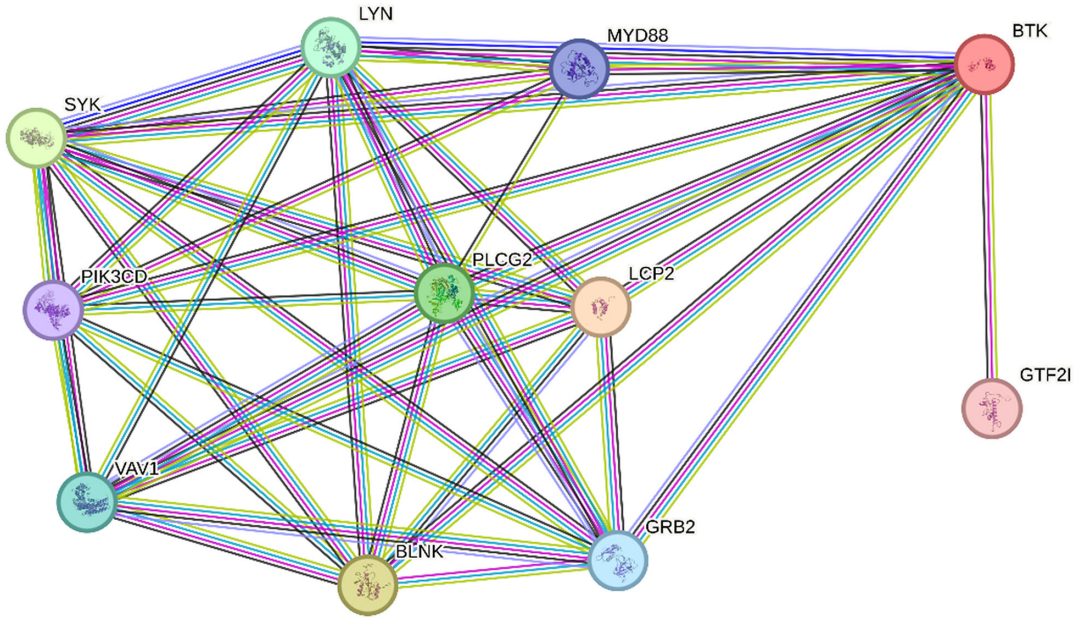
Supplementary Figure 1. (A) A count plot showing the bioactivity class of the dataset with each class count. (B) The pIC_{50} boxplot of the dataset distribution. (C) The molecular weight boxplot of the dataset distribution. (D) The LogP value boxplot of the dataset distribution. (E) The HBA value boxplot of the dataset distribution. (F) The HBD value boxplot of the dataset distribution.



Supplementary Figure 2. (A) A bar plot showing the R-square result of the top 10 models among the 41 models trained and (B) a bar plot showing the RMSE result of the top 10 models among the 41 models trained.



Supplementary Figure 3. H-bond distribution (within 0.3 nm distance) for protein-ligand complex of the selected ligands in comparison to the reference drug (black).



Supplementary Figure 4. Comprehensive protein-protein interaction (PPI) analysis of Bruton's tyrosine kinase (BTK).