



Contents lists available at ScienceDirect

Saudi Pharmaceutical Journal

journal homepage: www.sciencedirect.com

QSAR modelling, molecular docking, molecular dynamic and ADMET prediction of pyrrolopyrimidine derivatives as novel Bruton's tyrosine kinase (BTK) inhibitors

Mourad Aloui^a, Mohammed Er-rajy^a, Hamada Imtara^{b,*}, Amina Goudzal^c, Sara Zarougui^a, Mohamed El fadili^a, David E. Arthur^d, Ramzi A. Mothana^e, Omar M. Noman^e, Mahmoud Tarayrah^f, Elhalaoui Menana^a

^a LIMAS Laboratory, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco

^b Faculty of Sciences, Arab American University Palestine, Jenin 44862, Palestine

^c Engineering Laboratory of Organometallic, Molecular Materials and Environment, Sidi Mohamed Ben Abdellah University, Faculty of Sciences, Fez, Morocco

^d Department of Pure and Applied Chemistry, University of Maiduguri, Nigeria

^e Department of Pharmacognosy, College of Pharmacy, King Saud University, Riyadh 11451, Saudi Arabia

^f Groupe Hospitalier Cochin-Port Royal, Faculty of Medicine, Institut Cochin, Paris University, CNRS, IN-SERM, 75000, Paris, France

ARTICLE INFO

Keywords:

QSAR
Pyrrolopyrimidine
Molecular docking
Molecular dynamic
ADMET propriety
BTK inhibitors

ABSTRACT

In recent years, there has been a focus on developing and discovering novel Bruton's tyrosine kinase (BTK) inhibitors, as they offer an effective treatment strategy for B-cell malignancies. BTK plays a crucial role in B cell receptor (BCR)-mediated activation and proliferation by regulating downstream factors such as the NF- κ B and MAP kinase pathways. To address this challenge and propose potential therapeutic options for B-cell lymphomas, researchers conducted 2D-QSAR and ADMET studies on pyrrolopyrimidine derivatives that act as inhibitors of the BCR site in cytochrome *b*. These studies aim to improve and identify new compounds that could serve as more potent potential BTK inhibitors, which would lead to the identification of new drug candidates in this field.

In our study, we used 2D-QSAR (multiple linear regression, multiple nonlinear regression, and artificial neural networks), molecular docking, molecular dynamics, and ADMET properties to investigate the potential of 35 pyrrolopyrimidine derivatives as BTK inhibitors. A molecular docking study and molecular dynamics simulations of molecule 13 over 10 ns revealed that it establishes multiple hydrogen bonds with several residues and exhibits frequent stability throughout the simulation period. Based on the results obtained by molecular modeling, we proposed six new compounds (Pred1, Pred2, Pred3, Pred4, Pred5, and Pred6) with highly significant predicted activity by MLR models. A study based on the *in silico* evaluation of the predicted ADMET properties of the new candidate molecules is strongly recommended to classify these molecules as promising candidates for new anticancer agents specifically designed to target Bruton's tyrosine kinase (BTK) inhibition.

1. Introduction

In recent years, the development of Bruton's tyrosine kinase (BTK) inhibitors has garnered significant attention due to their efficacy in treating B-cell malignancies (Yang et al., 2022). BTK, a member of the Tec family of tyrosine kinases, is predominantly expressed in B cells, macrophages, and monocytes, while being absent in T cells, NK cells, and plasma cells (Draper et al., 2022). BTK plays a central role in B cells,

facilitating B cell receptor (BCR)-mediated activation and proliferation. It achieves this by modulating downstream factors of the BCR, particularly by activating the NF- κ B and MAP kinase pathways (Satterthwaite et al., 1998). Disruptions in BTK, given its crucial functions in BCR and Fc receptor (FcR) signaling, can lead to the development of severe leukemias and B-cell-related lymphomas. Beyond its involvement in BCR signaling, BTK also plays a critical role in Fc receptor (Fc γ R) signaling. This pathway induces the production of BTK-dependent pro-

Peer review under responsibility of King Saud University.

* Corresponding author.

E-mail addresses: Hamada.tarayrah@gmail.com, hamada.imtara@aaup.edu (H. Imtara).

<https://doi.org/10.1016/j.jsps.2023.101911>

Received 8 September 2023; Accepted 7 December 2023

Available online 12 December 2023

1319-0164/© 2023 The Author(s). Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

inflammatory cytokines, particularly in cells like macrophages (Di Paolo et al., 2011). The discovery of X-linked agammaglobulinemia (XLA), an inherited immune deficiency, by Dr. Ogden Carr Bruton, an American pediatrician, in 1952 underscored the enzyme's significance in B-cell development (Salemi et al., 2015). XLA, an orphan disease, is characterized by diminished immunoglobulin production, rendering individuals more susceptible to bacterial infections (Goyal et al., 2016; Rezaei et al., 2011). This condition also leads to a deficiency in circulating naive B lymphocytes due to impaired B lymphoid maturation and an inability to form secondary lymphoid organs (Tangye et al., 2013). Concerns persist regarding the limited oral bioavailability and selectivity towards the kinase of covalent BTK inhibitors (Burger, 2014). This necessitates a wider range of inhibitors with innovative structures and enhanced target-binding selectivity (Singh et al., 2011). Consequently, BTK has emerged as a highly attractive therapeutic target for a diverse spectrum of diseases characterized by aberrant activation of B lymphocytes and/or macrophages, including B-cell malignancies (Puri et al., 2013; Singh, 2022).

Several approved covalent inhibitors, including evobrutinib, spebrutinib, and branebrutinib, which belong to the category of irreversible inhibitors, have been introduced through clinical trials (Fig. 1). Additionally, numerous non-covalent BTK inhibitors have been reported in the literature. Irreversible BTK inhibitors face challenges related to their low oral bioavailability and limited kinase selectivity, necessitating high clinical doses that may lead to undesirable side effects (Bye et al., 2017; de Vries et al., 2016; Senis et al., 2014). Consequently, there remains a demand for a broader range of inhibitors with innovative structures and precise target binding profiles to overcome current limitations. Therapeutic trials have resulted in the development of numerous approved covalent inhibitors in the field of drug design and discovery. This approach offers several advantages, including reduced experiment duration and quantity, while providing valuable insights for rational drug design. To this end, we conducted an *in silico* study to investigate the biological inhibitory activity of thirty-five novel pyrrolopyrimidine-derived small molecules as novel BTK inhibitors (Yang et al., 2022).

In recent years, the use of molecular modeling techniques has yielded highly impressive results in the drug discovery process [1, 2]. Two-dimensional quantitative structure–activity relationship (2D-QSAR), molecular docking, pharmacokinetic parameters (ADMET), and molecular dynamics (MD) simulation have been carried out to design new molecules capable of inhibiting Bruton's tyrosine kinase (BTK). In this study, 35 BTK inhibitors were thoroughly investigated using a variety of

sophisticated molecular modeling methods, including 2D-QSAR (El fadili et al., 2022; Mrabti et al., 2022), density functional theory (DFT), molecular docking, pharmacokinetic property analysis, drug-like ADMET, as well as molecular dynamics simulations (Er-raji et al., 2022a, 2023), to identify key structural factors affecting inhibitory activity. In addition, the QSAR model, which links molecular descriptors to activity, provides suggestions for the creation of new drugs. To assess their drug potential, all designed compounds were evaluated by calculating ADMET properties. Moreover, molecular dynamics simulations were conducted for 10 ns to estimate the ligand's stability within the protein under normal physiological conditions. The remainder of this article is structured as follows: the second section outlines the materials and methods employed in this study. Subsequently, the third section presents the study's findings. Finally, the concluding section summarizes the discussion of the key results and the overall implications of this research.

2. Materials and methods

2.1. Experimental dataset

The following table (Table 1) presents experimental data on 35 pyrrolopyrimidine-derived molecules as novel inhibitors of BTK.

2.2. Examined compounds

To perform the molecular modeling, we employed experimental data on the Bruton's tyrosine kinase (BTK) inhibitory activities of 35 previously synthesized pyrrolopyrimidine-derived molecules (Yang et al., 2022). The observed activities (IC_{50}) were therefore transformed into (pIC_{50}) by a logarithm of the log IC_{50} scale and are presented in Table 1.

2.3. Calculation of descriptors for the compounds studied

To build a reliable QSAR model, a total of 17 descriptors, including lipophilic, geometric, physicochemical, and steric descriptors, were used. These descriptors were determined using the MM2 technique with ACD/ChemSketch (Österberg and Norinder, 2001) and ChemBioOffice (Milne, 2010). Additionally, the geometry of the molecules concerned was optimized using the density functional theory (DFT) method, with a 6-31G basis set and the B₃LYP functional (Parr and Yang, 1995; Zhang et al., 2009). Electronic descriptors were also determined using the

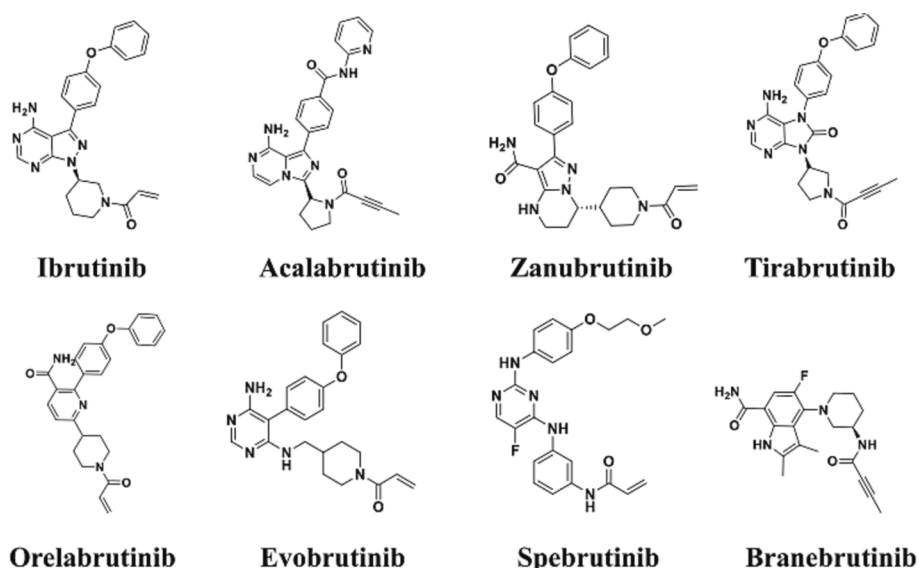


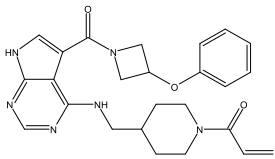
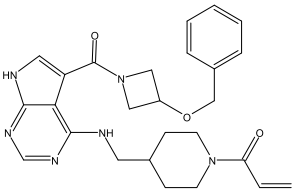
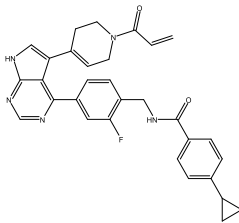
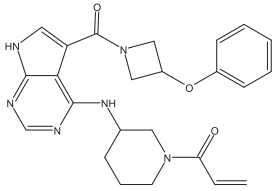
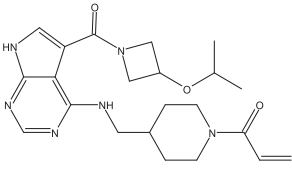
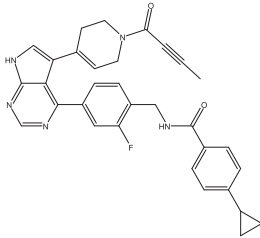
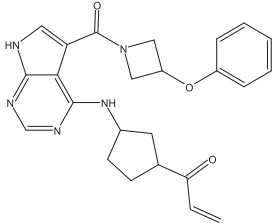
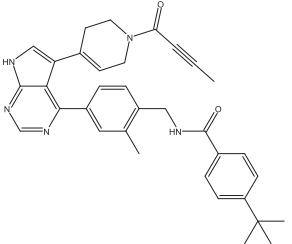
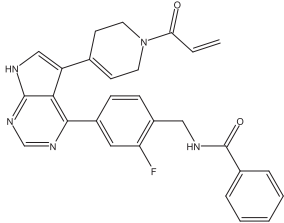
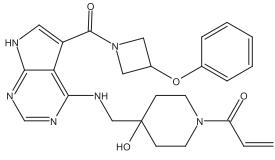
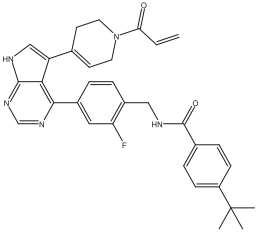
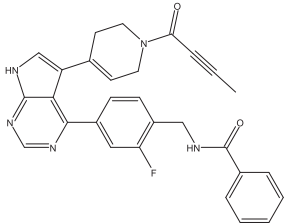
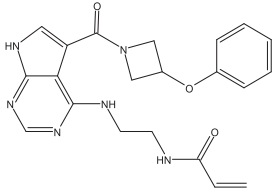
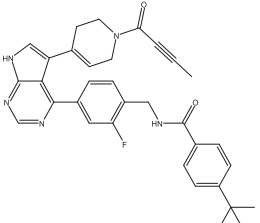
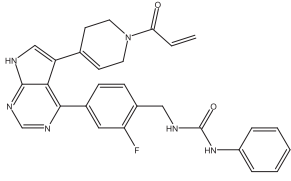
Fig. 1. Approved and representative BTK inhibitors in clinical trials.

Table 1
Structures and pIC₅₀ values of pyrrolopyrimidine derivatives as novel BTK inhibitors.

Comp	Structure	pIC ₅₀	Comp	Structure	pIC ₅₀	Comp	Structure	pIC ₅₀
1*		7.76	13		9.30	25*		8.85
2		8.89	14		9.30	26*		8.72
3		7.28	15		9.10	27		8.92
4*		7.01	16		8.85	28		8.68
5		8.08	17		9.10	29		8.66

(continued on next page)

Table 1 (continued)

Comp	Structure	pIC ₅₀	Comp	Structure	pIC ₅₀	Comp	Structure	pIC ₅₀
6		9.22	18*		8.70	30		8.92
7		8.72	19		7.63	31*		8.72
8		8.35	20		8.96	32*		9.15
9		8.66	21*		8.89	33		9
10		7.03	22		9.10	34		9.10

(continued on next page)

Table 1 (continued)

Comp	Structure	pIC ₅₀	Comp	Structure	pIC ₅₀	Comp	Structure	pIC ₅₀
11		9.15	23		8.60	35		8.85
12*		9.40	24		9.10			

*Refer to test set molecules, pIC₅₀ = 9 - log₁₀ (IC₅₀).

Gaussian 03 quantum chemistry software (Frisch et al., 2004). The results of the descriptor calculations used in this work are presented in additional data (Table S1). The categories of these descriptors grouped together in this work are presented in Table 2.

2.4. Quantitative structure-activity relationship modeling

With a view to developing a QSAR model, we selected a set of 35 molecules from previous work that were shown to have significant activity as novel BTK inhibitors. The set was randomly divided into two subsets: a training set (26 molecules) to build the model, and a test set (9 molecules) to assess the model's validity. Various statistical methods were used to build the QSAR model, namely MLR and ANN (Gupta et al., 2016).

2.4.1. Multiple linear regression (MLR)

The MLR method is extensively utilized in QSAR studies to select molecular descriptors due to its simplicity and reliability (Roy and Mitra, 2011a). MLR is also employed in conjunction with the MNLR (multinomial logistic regression) and ANN (artificial neural network)

Table 2

Description of the descriptors used in this work.

Descriptors	Symbol	Class
Stretch	S	Geometrical
Bend	B	
Stretch-Bend	S-B	
Torsion	Tor	lipophilic
LogP	LogP	
Mol Refractivity	MRef	
Mol weight	MW	Constitutional
Number of HBond Acceptors	NHBA	
Number of HBond Donors	NHBD	topology
Num Rotatable Bonds	NRB	
Total Connectivity	TC	
Total Valence Connectivity	TVC	Quantum (Electronic)
Polarizability	Polariz	
Energy HOMO	E_{HOMO} (eV)	
Energy LUMO	E_{LUMO} (eV)	
total energy	E_T (eV)	
Dipole moment	D_p (D)	

methods to determine suitable descriptors used as input parameters for developing QSAR models. The fundamental premise of MLR is that the dependent variable exhibits a linear relationship with certain independent variables, as expressed by the following equation.

$$Y = a_0 + \sum_{i=1}^n a_i X_i \quad (1)$$

Where Y is the dependent variable, X_i are the independent variables, n is the number of molecular descriptors, a₀ is the constant component of equation (1), a_i represent the coefficients of the molecular descriptors.

2.4.2. Multiple nonlinear regression (MNLR)

The MNLR method is a non-linear concept that consists of selecting the mathematical model that best describes the non-linear variation of a molecular property or a biological activity (Y) in relation to the molecular descriptors (Xi) (Er-raji et al., 2022b). In this context, we use the second-order polynomial model to build the QSAR model via the MNLR technique, based on the descriptors defined by the MLR model. The non-linear relationship between the molecular descriptors and the biological activity is determined by the following equation:

$$Y = a_0 + \sum_{i=1}^n a_i \times X_i + b_i \times X_i^2 \quad (2)$$

In the context of the MLR method for QSAR studies, the equation can be reformulated as follows:

In the model equation (2), Y represents the dependent variable, which corresponds to the biological activity to be predicted. The independent variables, X_i, encompass the molecular descriptors, and their count is denoted by 'n'. The model equation includes a fixed value, a₀, along with factors, a_i and b_i, which represent the coefficients associated with each descriptor in the equation (1).

2.4.3. Artificial neural networks (ANN)

ANNs are employed to enhance the likelihood of characterizing compounds and to construct a predictive model that establishes a relationship between all the quantitative molecular descriptors acquired from the MLR model and the observed biological activity values. Our focus lies in developing an ANN-based QSAR model to validate the

exceptional accuracy of the molecular descriptors selected through the MLR model. Additionally, leveraging the ANN model enables us to achieve highly precise predictions of biological activity for each molecule (Salt et al., 1992). This approach relies on using the sigmoid activation function in the hidden layer and the linear activation function in the output layer. The ANN architecture utilized in this study comprises three layers of neurons, namely, the input layer, the hidden layer, and the output layer, as illustrated in Fig. 2.

In the input layer, the number of neurons should be equal to or less than the total number of descriptors obtained from the multiple linear regression model. The output layer, on the other hand, contains the predicted activity values. Once the number of neurons in the hidden layer is determined, calculating the parameter ρ becomes essential. This parameter, denoted as ρ , is computed using the equation $\rho = (\text{number of weights}) / (\text{number of connections})$. As per the recommendations of certain authors, for the ANN model to be statistically acceptable and to ensure comprehensive contribution from all elements in the database used, the parameter ρ should ideally fall within the range of 1 to 3 (Kůrková, 1992).

2.5. Leave one out cross-validation (LOO-CV)

To assess the accuracy of the QSAR models created through MLR, MNL, and ANN, we utilize an internal validation method known as LOO-CV (Golbraikh and Tropsha, 2002). In the LOO-CV process, we iteratively modify the dataset by removing one molecule at a time, and then reconstruct the QSAR model using the remaining molecules to predict the activity of the deleted molecule. This cycle is repeated for each molecule in the dataset until all molecules have been tested and included in the validation process (Roy and Mitra, 2011a). This validation approach relies on calculating the R_{cv}^2 coefficient of performance, as described in Eq. (3). Following this methodology, we can assess the model's accuracy and determine its predictive capabilities (Golbraikh and Tropsha, 2002). Ideally, the R_{cv}^2 value should exceed 0.5, indicating that the developed model demonstrates robustness concerning internal predictions.

$$R_{cv}^2 = 1 - \frac{\sum (Y_{ob}(trai) - Y_{ca}(trai))^2}{\sum (Y_{ob}(trai) - \bar{Y}_{trai}(trai))^2} \quad (3)$$

$Y_{ob}(train)$ refers to the actual observed response value in the training set. $Y_{ca}(trai)$ represents the response value predicted by the LOO-CV technique applied to the training set. Lastly, $\bar{Y}_{trai}(trai)$ denotes the mean value of both the observed and predicted responses in the training set.

2.6. Y-randomization test

The Y-randomization test is utilized to eliminate the chance of any random correlation between the descriptors and their corresponding biological activities in the MLR-generated model. This test helps ensure that the model's predictive power is not a result of random associations between the variables. Consequently, the presence of any random correlation between the X values (molecular descriptors) and the Y values (biological activity) could significantly affect the performance and reliability of not only the MLR model but also the MNL and ANN models. Therefore, it is of utmost importance to safeguard the efficacy and validity of all these models against any potential influence from such random associations. The randomization test Y involves randomizing the experimental values of properties/activities across the descriptors in the original model. This process generates multiple new models by creating different distributions of the data (Růcker et al., 2007). The Y randomization test deems the QSAR model acceptable and not obtained by chance when the average random correlation coefficient (R_r^2) calculated from the randomly generated models is lower than the correlation coefficient R^2 of the original non-random model. This comparison ensures that the actual model's performance is significantly better than what could be expected from random associations (Roy and Mitra, 2011b).

2.7. Molecular docking simulations

Molecular docking is a precise, fast and efficient technology in the pharmacology sector (Bassani et al., 2022; Er-Rajy et al., 2023). Molecular docking simulation was performed using AutoDock 4.2.3 software to analyze the interaction process and determine binding modes to better understand key structural and geometric requirements (Nour et al., 2022). The docking study is a tool to recognize pharmacophores with the ability to interconnect with an enzyme, which is a function of binding affinity. According to Ritchie et al., 2008, the impulsive nature of the binding relationship between the ligand and the enzyme being studied could be enhanced by lowering the binding energy (Trott and Olson, 2010).

The protein molecules and the compound were selected in order to conduct an analysis about their interactions, which assists with the determination of how accurately the target was hit. After that, utilizing the Protein Data Bank (PDB) database (<https://www.rcsb.org/>) (Rose et al., 2021), the PDB file of protein was mined. The process of molecular docking between the acquired protein and ligand was carried out. With the help of the Biovia discovery studio visualizer, both the two-dimensional (2D) and the three-dimensional (3D) images of the interactions between proteins and compound were studied. The active site

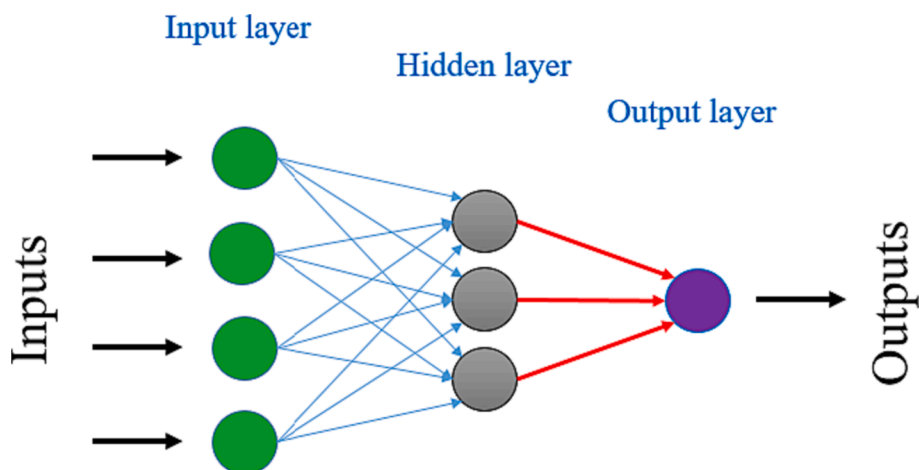


Fig. 2. The architecture of the ANN model used in this study.

is surrounded by the three-dimensional configuration ($X = 22.13$, $Y = 7.91$, $Z = 2.81$), which has a grid with a size of $60 \times 60 \times 60$ points in the x, y, and z directions.

2.8. Molecular dynamics (MD)

The MD simulation was conducted using the Desmond simulation package developed by Schrödinger LLC (Schrödinger, 2017). During the simulation runs, an NPT ensemble with a temperature of 300 K and pressure of 1 bar was employed. The simulations were run for duration of 10 ns, and a relaxation time of 1 picosecond was set for both the lead inhibitors and the optimally designed ligand. The OPLS_2005 force field parameters were used (Shivakumar et al., 2012, 2010) and the long-range electrostatic interactions were calculated using the Particle Mesh Ewald method (Essmann et al., 1995; Petersen, 1995) with a cutoff radius of 9.0 Å. The Simple Point Charge model was employed to explicitly depict the water molecules. To control pressure, a Martyna-Tuckerman-Klein chain coupling approach with a coupling constant of 2.0 ps was utilized, while temperature control was achieved using a Nose-Hoover chain coupling scheme (Jang and Voth, 1997; Martyna et al., 1992; Patra and Bhattacharya, 2014). An r-RESPA integrator was employed to calculate the non-bonded forces, updating short-range forces every step and long-range forces every three steps. Trajectories were saved at 4.8 ps intervals to facilitate analysis. The Desmond MD package implemented the Simulation Interaction Diagram, which was used to examine the behavior and interactions between the ligands and protein. To monitor the stability of the MD simulations, the Root Mean Square Deviation (RMSD) of both the ligand and protein was assessed at various time points. The AMBER 14 (Salomon-Ferrer et al., 2013) package with the AMBER force field FF99 (Spasic et al., 2012) was also used to minimize, add counter ions, solvate, equilibrate, and run periodic box, explicit water (TIP4P) MD simulations for the best inhibitors. The protein-ligand-water system data analysis was carried out with the AMBER Tools distribution program (Kalayan et al., 2023; Kalayan and H. Henchman, 2021).

2.9. In silico pharmacokinetic-pharmacodynamic modeling (ADMET)

The progress in computer technology has played a crucial role in advancing new drug candidate development by reducing the reliance on extensive experimental studies and enhancing overall success rates. Consequently, in the drug discovery process, ADMET pharmacokinetic parameters and drug similarity are now determined early on for preliminary estimation. In silico studies offer a pathway to assess crucial ADMET parameters, including absorption, distribution, metabolism, excretion, and toxicity (Vickers, 2017). The processes involved in drug development include absorption, which refers to the uptake of compounds in the human small intestine; distribution, which involves the movement of compounds throughout the body tissues; metabolism, which pertains to the chemical biotransformation of a compound by the body; excretion, which is the removal of a compound from the body; and the assessment of the compound's toxicity level.

To predict the familiarity of drugs with selected compounds, criteria established by Lipinski, Ghose, Veber, Egan, and Muegge are used. Lipinski, Veber, and Egan's rules are particularly valuable in evaluating the ADME properties of human drugs. These rules are highly beneficial for identifying potential drugs based on the 2D structure of small molecules and the bioavailability of these molecules when administered orally (Hansch et al., 2004). Compounds that fail to meet at least two of the Lipinski, Veber, and Egan rules often encounter various issues with their pharmacokinetic properties related to ADMET. Approximately 10 % of drugs that reach the clinical trial phase do not comply with any of these rules (Lipinski, Veber, and Egan). In addition to the aforementioned rules, we also assess two other factors: the number of rotational bonds (n-ROTB) and the topological polar surface area (TPSA) (Jin et al., 2020). The prediction of these factors allows us to know if the molecule

interacts with the receptor in a flexible or inflexible mode.

3. Results

3.1. Multiple linear regressions

After calculating the molecular descriptors of 35 derivatives (Table S1), several attempts were made to build a reliable model. The best model we obtained was built on four descriptors such as $NHBD$, Tc , E_{HOMO} , and ET . Based on the obtained results, the following molecules (1, 4, 8, 12, 18, 21, 25, 26, 31, and 32) have been chosen for the test set. Meanwhile, the following molecules (2, 3, 5, 6, 7, 9, 10, 11, 13, 14, 15, 16, 17, 19, 20, 22, 23, 24, 27, 28, 29, 30, 33, 34, and 35) are included in the training set. The QSAR model produced by the MLR technique is shown in Eq. (4) below.

$$\begin{aligned} pIC_{50} &= 13.47 - 0.39772 \times NHBD - 12951 \times Tc + 22,76787 \\ &\quad \times EHOMO - 0.00079 \times ET R_{Ajust}^2 \\ &= 0.924; MSE = 0.027; F = 77,3; Pr < 0.0001; R_{cv}^2 = 0.91 \end{aligned} \quad (4)$$

N represents the number of compounds in the training set; MSE represents the root mean square error

Fig. 3 shows the relationship between the observed activity values and predicted values. HOMO orbital energy has a positive effect on activity, but number of H-bond donors, total Connectivity and total energy influence it negatively, as shown in Fig. 4.

3.2. Multiple nonlinear regression

The nonlinear QSAR model obtained using the MNLR technique is shown in Eq. (5) below.

$$\begin{aligned} pIC_{50} &= 37.06650 - 0.36441 \times NHBD - 26410 \times TC + 346.47310 \\ &\quad \times E_{HOMO} - 5198 \times 10^6 \times TC^2 + 759.90603 \times E_{HOMO}^2 - 3.44994 \\ &\quad \times 10^{-6} \times E_T^2 N \\ &= 26; R = 0.97; R^2 = 0.945; MSE = 0.028 \end{aligned} \quad (5)$$

The biological activities observed and estimated by the QSAR model developed on the basis of the training set and the test set for the linear and non-linear models are presented in the additional data (Table S2).

Fig. 5 illustrates the uniform distribution of experimental pIC_{50} values compared to the predicted values obtained through the MNLR technique, confirming the superiority of the developed QSAR model in terms of efficiency.

3.3. Artificial neural networks (ANN)

When using the ANN technique to develop a QSAR model, the architecture used is 4-3-1 with a ρ parameter of 1.368. With a value of ρ between 1 and 3, it is clear that the number 3 in the hidden layer is proportional to the number of descriptors 4 in the input layer, thus predicting the pIC_{50} values expressed by the output layer 1.

Fig. 6 illustrates that the candidate pIC_{50} values are evenly distributed in the training set, ensuring that the ANN model predictions are very similar to the experimentally observed pIC_{50} values.

3.4. Leave-one-out cross-validation (LOO-CV)

Fig. 7 presents the results of the cross-validation performed with the LOO method.

3.5. External validation

We perform an external validation using the Golbraikh and Tropsha

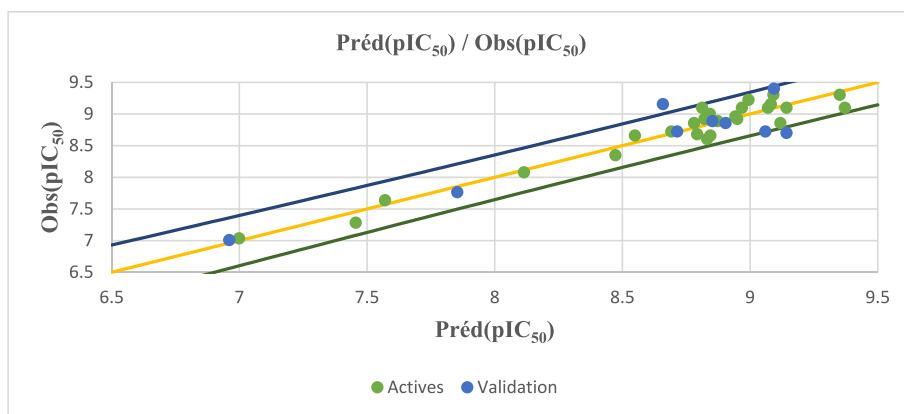


Fig. 3. Correlations between the observed activity values and the predicted ones via the MLR model.

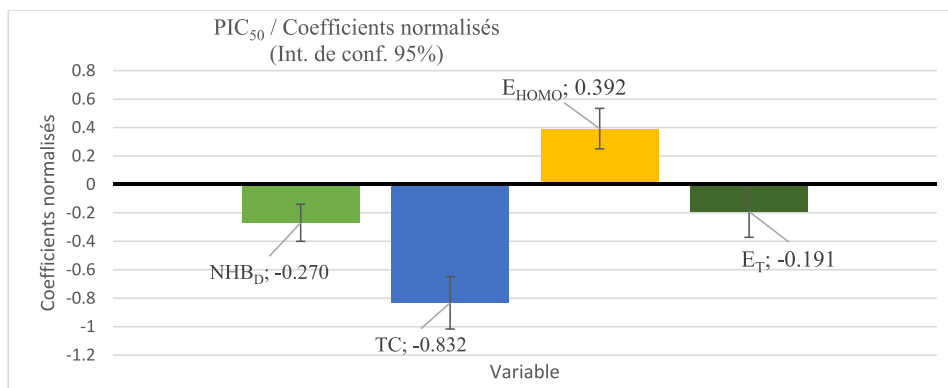


Fig. 4. Analysis of the contribution coefficients of the four molecular descriptors of the MLR model.

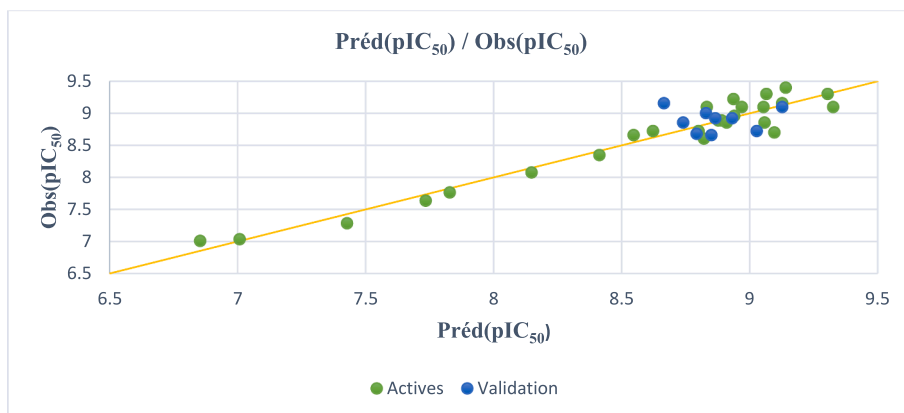


Fig. 5. Correlations between the observed/ predicted activity by the MNLR model.

criteria (Golbraikh et al., 2003) to evaluate the QSAR models' capability in predicting pIC_{50} activity values for molecules in the test set (Table 3 and 4). This assessment involves calculating the correlation coefficient R^2 , a significant criterion in evaluating how well externally validated models can predict the activities of molecules that were not part of the model development process (Fig. 8).

3.6. Y-randomization test

Randomization is a commonly employed technique in QSAR studies to ensure the reliability and robustness of the obtained models. After selecting a regression model, randomization is performed to validate the

model by assessing potential correlations. In this specific case, Y-randomization is utilized to assess the effectiveness of the model.

The activity values of the compound series are randomly shuffled and redistributed multiple times, and each time the model is reconstructed using the original descriptors and procedure. This process is repeated 100 times to obtain a comprehensive evaluation (Table S3).

3.7. Molecular docking simulations

Table 5 shows the binding energy for three compounds (12, 13 and 14). In addition, the docking analysis found that the chemical has a high affinity for interacting with the protein correspondingly (Figs. 9, 10 and

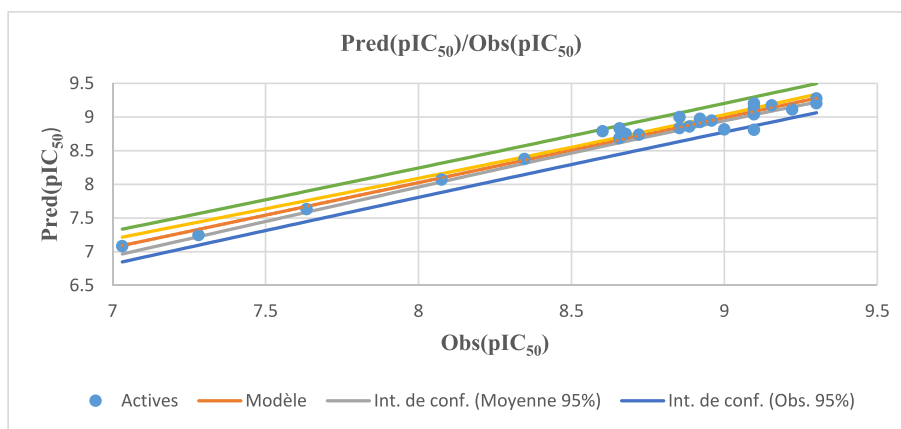


Fig. 6. Correlation between the observed and the predicted activities calculated by ANN.

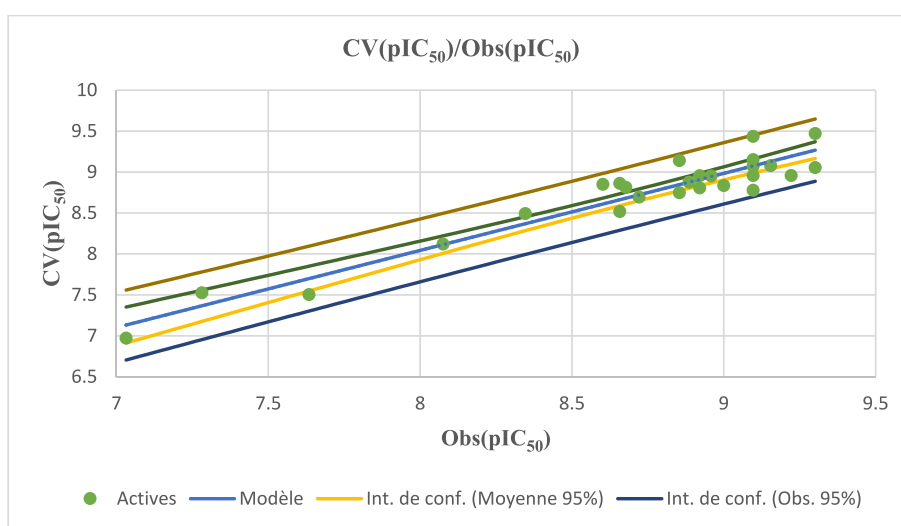


Fig. 7. Correlation of observed and predicted activities calculated using LOO-CV.

Table 3

Criteria of Golbraikh and Tropsha's for external validation.

Parameter	Threshold	Modelscore
Q^2_{training}	$Q^2_{\text{training}} > 0.5$	0.91
r^2	$r^2 > 0.6$	0.85
r_0^2		0.84
r_0^2		0.848
$ r_0^2 - r_0^2 $	$ r_0^2 - r_0^2 < 0.3$	0.008
K	$0.85 < k < 1.15$	1
$\frac{r^2 - r_0^2}{r^2}$	$\frac{r^2 - r_0^2}{r^2} < 0.1$	0.01
k'	$0.85 < k' < 1.15$	0.99
$\frac{r^2 - r_0^2}{r^2}$	$\frac{r^2 - r_0^2}{r^2} < 0.1$	0.002

11).

3.8. Molecular dynamics simulations studies

The molecular docking analyses have revealed specific information regarding the binding modes of the protein-inhibitor complex. However, to identify even the smallest discrepancies, we conducted molecular dynamics simulations. By utilizing the most favorable interactions and energy-optimized conformation obtained from the molecular docking results, we examined the atomic details of compound 13 within the

Table 4

The results of the prediction by MLR methods for the test set.

Compounds	pIC ₅₀ obs	Pred (pIC ₅₀) MLR	Residual
1	7.762	7.854	-0.092
4	7.006	6.961	0.045
20	9.398	9.094	0.304
27	8.699	9.143	-0.444
30	8.886	8.853	0.033
35	8.854	8.904	-0.050
36	8.721	8.715	0.006
41	8.721	9.060	-0.339
42	9.155	8.659	0.496

solvent system.

To commence the MD study, we selected the most active conformation based on hydrogen bond interactions, root-mean-square deviation (RMSD), and energy values. Subsequently, we performed a 10 ns simulation to assess the stability of the protein-inhibitor complex. Moreover, we extensively investigated the results obtained from the simulation to gain further insights into the system (Fig. 12).

The RMSF attribute enables the identification of specific modifications in the protein chain. As depicted in Fig. 13, the catalytic domain exhibits significant fluctuations in the RMSF values of the protein's backbone amino acids, particularly at the N- and C-terminals, when compared to other regions of the Protein.

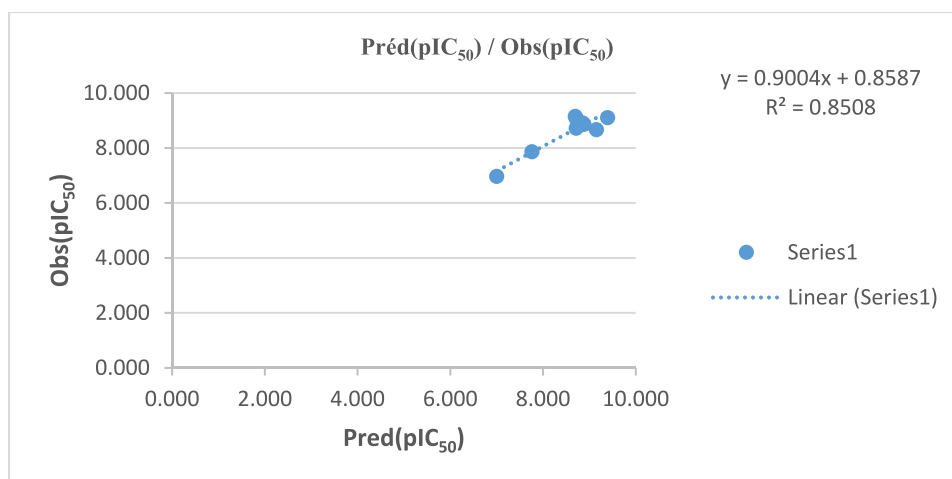


Fig. 8. Correlations between the observed / predicted activity by the MLR model for the test set.

Table 5

The results of molecular docking for the three most active compounds of the studied series.

Compound	Binding energy	NOHB	HBAA	Distance
12	−8.12	3	GLU 599	2.284
			TYR 617	2.728
			HIS620	2.889
13	−9.26	3	ASN526	3.017
			ARG525	2.204
			LYS430	2.854
14	−8.08	2	GLU59	2.150
			TYR61	2.798

The MD results show that the RMSD value for the protein backbone atom, relative to its initial position, increased to 2.20 Å over the first 5 ns of the trajectory, but then decreased to 1.75 Å at 10 ns. The average RMSD value for Compound 13 backbone atoms was 1.95 Å for the heavy atoms, indicating excellent structural reconstruction throughout the simulation of the compound 13_5P9J complex. Throughout the simulation, the protein's RMSF (root mean square fluctuation) remained stable at approximately 1.50 Å, providing a solid foundation for further investigation (Fig. 14 and Fig. 15).

3.9. Design of new compounds

The main objective of this study is to design new BTK inhibitors derived from pyrrolopyrimidine, based on the recommendations we extracted from the 2D-QSAR studies. In this study, six pyrrolopyrimidine derivatives were designed (Pred1, Pred2, Pred3, Pred4, Pred5 and Pred6) to enhance the inhibitory activity of Bruton's tyrosine kinase inhibitor Table 6 and Table 7.

3.10. Lipinski's rule

All the proposed new compounds comply with Lipinski's five rules (Table 8).

3.11. ADMET properties

To ensure the potential suitability of the designed molecules as medications, we employed pharmacokinetic parameters such as ADMET. In silico ADMET properties were predicted using the pkCSM online tool (Pires et al., 2015), and the details are provided in the corresponding table (Table 9).

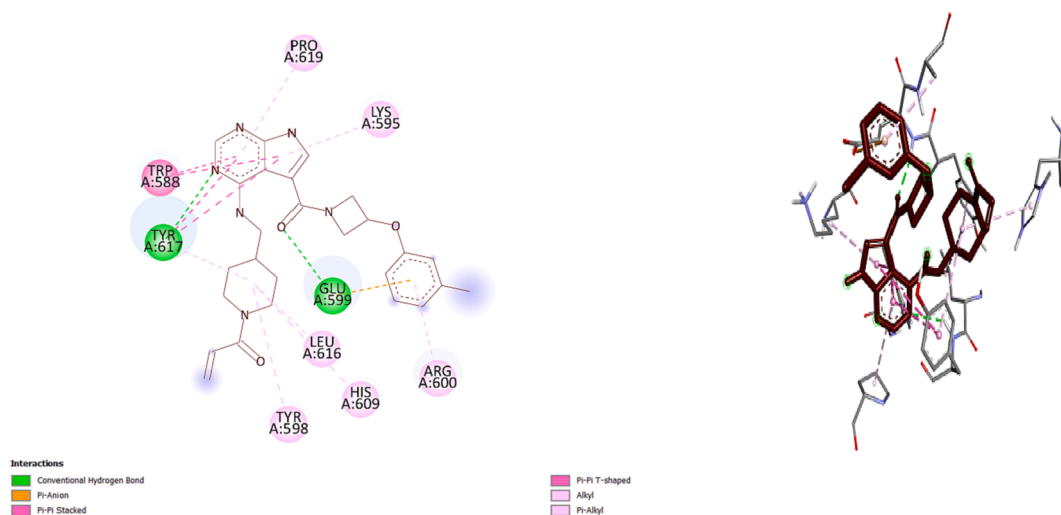


Fig. 9. 2D/3D -predicted binding modes of molecule 14 and the active site of human 5p9g.

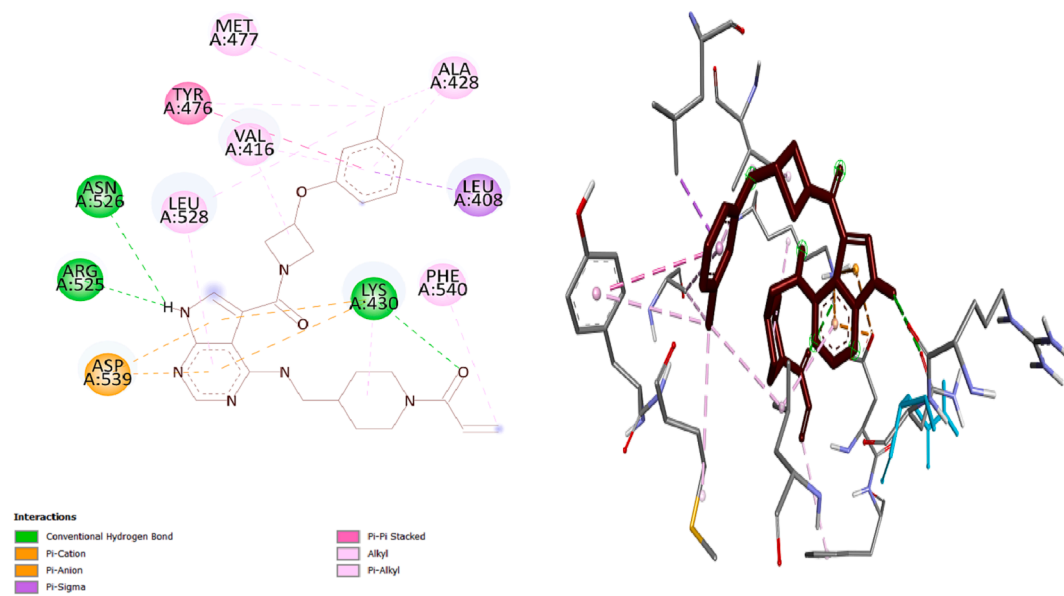


Fig. 10. 2D/3D-predicted binding modes of molecule 13 and the active site of human 5p9g.

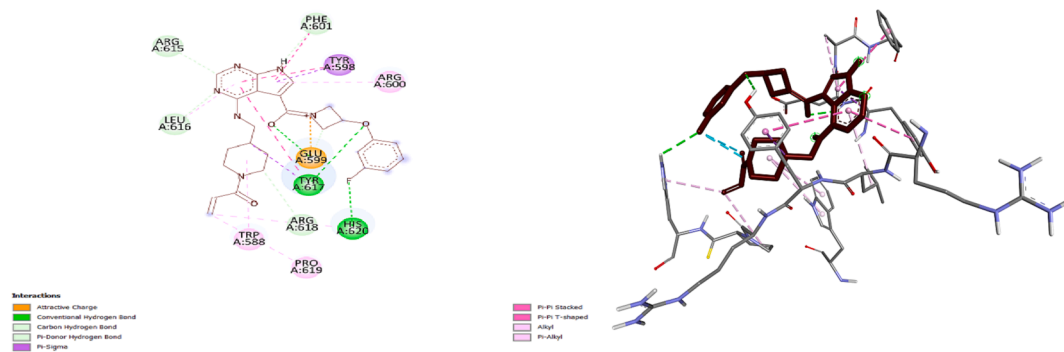


Fig. 11. 2D/3D-predicted binding modes of molecule 12 and the active site of human 5p9g.

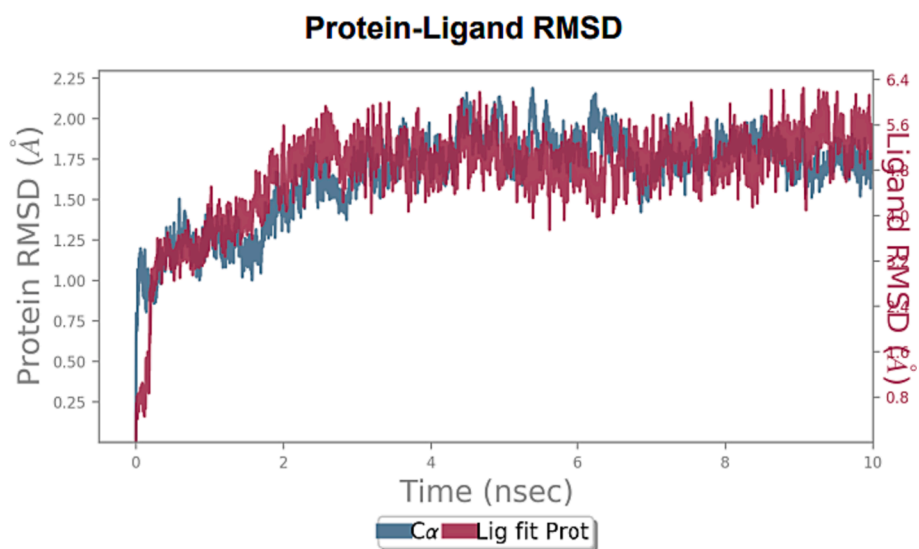


Fig. 12. The RMSD of protein 5P9J relative to the starting complexes during 10 ns MD trajectory for compound 13.

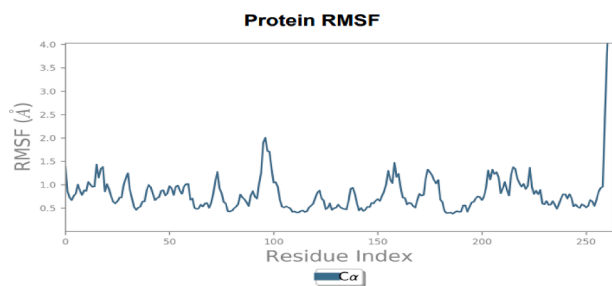


Fig. 13. The RMSF of 5P9J protein during 10 ns MD, representing local changes along the rotein chain for compound 13.

4. Discussion

RML results clearly indicate that the four selected descriptors are linearly correlated with BTK inhibitory biological activity (pIC_{50}) values Eq. (4). The QSAR model obtained by this technique is evaluated according to the following parameters: R^2 , F, MSE, P-value and R_{cv}^2 .

Coefficients of determination with a higher value ($R^2 = 0.94$), a lower MSE value ($MSE = 0.027$), and a high value of the statistical confidence level ($F = 77.37$), certifying that the QSAR model presented in Eq. (4) is statistically acceptable. In addition, the P-value obtained of less than 0.05 ($Pr < 0.0001$) shows that the QSAR model equation is statistically significant at a level greater than 95%. Otherwise, the value of cross-validation correlation coefficient ($R_{cv}^2 = 0.91$), much greater than 0.5, indicates the correctness of the QSAR model obtained by the MLR technique. The R_{cv}^2 value less than R^2 indicate the lightness and weakness of the model when excluding an item from the training set. The MLR-based QSAR model for the test and training set molecules acquires the latter. Fig. 3 shows that the significant correlation between the observed and predicted pIC_{50} values is explained by the low value of the MSE. It appears that the four descriptors in equation (4) are strongly linearly correlated with the biological activity of pIC_{50} . To improve the correlation between the activities predicted by the MLR-based QSAR model and the four molecular descriptors (NHB_D, TC, EHOMO and ET), a novel QSAR model is created using two non-linear methods, namely the MNLR and ANN techniques.

The performance indicators of the obtained non-linear QSAR model, namely $R = 0.97$, $R^2 = 0.945$ and $MSE = 0.028$, clearly demonstrate that this model is statistically valid (Fig. 5). Furthermore, the value of the

L00-CV coefficient ($R_{cv}^2 = 0.64$) confirms that the nonlinear model is internally validated, emphasizing that the efficiency and reliability of this model is attributed to the contribution of all elements ($N = 27$) of the training set in its construction.

The QSAR model created using the ANN technique displays a very high coefficient of determination R^2 of 0.97, indicating an excellent fit to the data. In addition, the MSE is low, with a value of 0.18, suggesting high accuracy in the model predictions. The results obtained show that the QSAR model possesses significant statistical quality when it comes to predicting BTK inhibitory activity. Therefore, it is relevant to use the four descriptors (NHB_D, TC, E_{HOMO} and E_T) in the prediction of pIC_{50} values. These descriptors were selected with relevance to this task.

The obtained parameters, $R^2 = 0.91$ and $RMSE = 0.17$, indicate that the constructed QSAR model is not significantly influenced by the cross-validation method (Fig. 7). These clear results demonstrate the stability and robustness of the proposed QSAR model. However, it should be noted that cross-validation is not an adequate measure to fully evaluate the capabilities of QSAR models.

The R^2 test yielded a value of 0.85, and all values fell within the acceptable range, confirming the successful validation of the Golbraikh and Tropsha criteria. These results indicate that the QSAR model created is highly effective. Additionally, the external validation of the QSAR models further demonstrates their strong capability to accurately predict pIC_{50} values for the experimental inhibitory activity of BTK inhibitors (Table 4 and Fig. 8).

The average correlation coefficients from these randomized models yield an R value of 0.38, R^2 value of 0.16, and Q_{cv}^2 value of 0.86. Comparing these averages to the values of the model, it is evident that the random target values generate significantly lower average R^2 and Q_{cv}^2 values. These results demonstrate that the correlations observed in model MLR between the descriptors and activities are not coincidental, confirming the robustness of the model. Therefore, the model's predictive power and reliability are validated through this randomization test.

Among three compound; compound 13 had the highest binding energy (-9.26 kcal/mol), in comparison to compound 12 (-8.12 kcal/mol), 14(-8.08 kcal/mol) (Table 5). Compound 13 interacted with ASN526, ARG525, LYS430 hydrogen bonding amino acid with 3 hydrogen bond (Fig. 10), while compound 12 interacted with 3 hydrogen bond with 3 amino acid GLU 599, TYR 617 and HIS620 (Fig. 11). Compound 14 interacted with amino acids GLU59 and TYR61 connected linked by two hydrogen bonds (Fig. 9). The Biovia discovery visualization tool demonstrated many interactions of ligand with protein through a variety

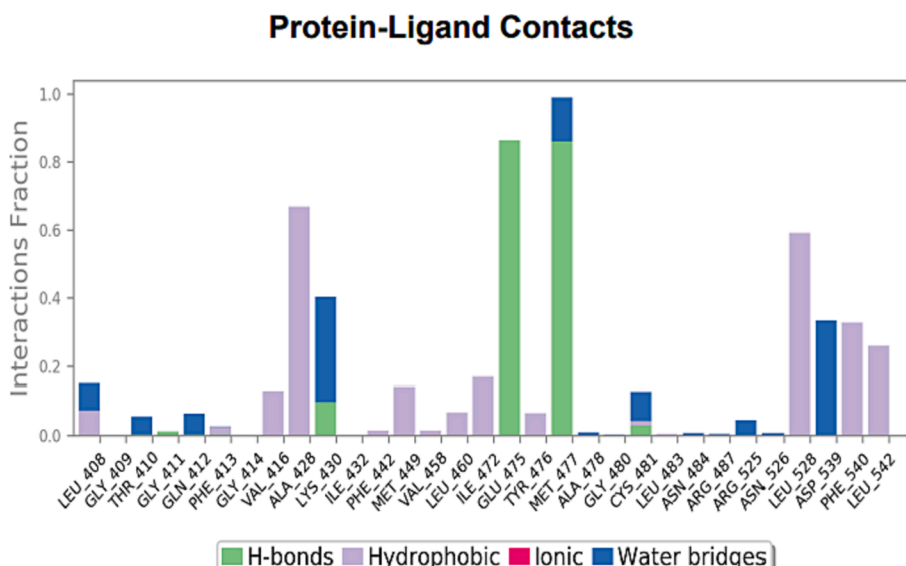


Fig. 14. The plot represents the hydrogen bonding interactions of compound 13 with respect to residues of 5P9J.

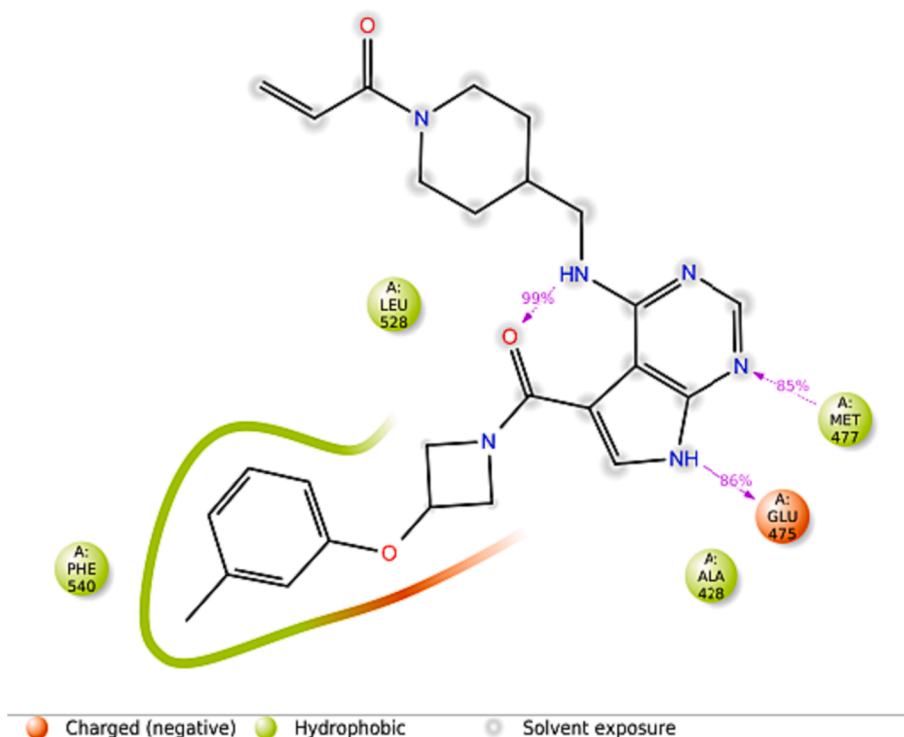


Fig. 15. Two-dimensional diagram of hit compound 13_5P9J interaction during 10 ns MD simulation.

Table 6

The values of parameters calculated for the news molecules and this predict activity.

Comp	NHB _D	Tc	E _{HOMO}	E _T	pIC ₅₀ _pred
Pred1	1	2.19×10^{-6}	-0.20472	-1982.461592	9.69
Pred2	1	2.19×10^{-6}	-0.20688	-1982.457269	9.64
Pred3	2	1.89×10^{-6}	-0.20042	-1638.500400	9.16
Pred4	2	1.89×10^{-6}	-0.20512	-1638.482827	9.05
Pred5	1	2.19×10^{-6}	-0.20166	-1639.612407	9.49
Pred6	1	2.19×10^{-6}	-0.20224	-1982.459187	9.75

of amino acid residues. These interactions included hydrogen bonds, alkyl bonds, sigma bonds, ionic bonds, sulfate bonds, and van der Waals bonds. Table 5 provides a rundown of the binding energy, the hydrogen bond, as well as 2D and 3D images of proteins that have been shown to interact with ligands.

The MD simulation of compound 13 (Figs. 12, 13 and 14) revealed the presence of water-bridge, hydrogen bonding, and hydrophobic interactions in various stable regions, namely Glu475–Met477, Lys430–Asp539, and Leu408–Leu542. Throughout the MD simulation, Glu475 formed hydrogen bonds, while Lys430, Met477, and Cys481 interacted with water bridges through hydrogen bonding. Water bridges were exclusively observed with Thr410, Gln412, and Asp539. Additionally, Ala428, Leu528, Phe540, and Leu542 exhibited both hydrophobic and ionic interactions, while Leu408 displayed interactions with both hydrophobic and water bridges. These specific residues played a significant role in the active binding site, contributing to the stabilization of compound 13 within the cavity.

During the simulation, numerous hydrogen bonds were detected at favorable frequencies. Moreover, it was observed that Met477 (–C N–H–O–Met) adopted a stable conformation of compound 13 at the active site for approximately 85 % of the simulation duration. The presence of hydrogen bond–water bridge networks was also demonstrated:

1. In 35 % of the molecular dynamics (MD) simulation, a nitrogen atom attached to a terminal keto oxygen atom formed a hydrogen bond with a conserved water molecule and residue Lys430.
2. In 30 % of the MD simulation, a nitrogen atom attached to compound 13 formed a hydrogen bond with a conserved water molecule and residue Cys481.

Throughout the entire simulation, the hydrogen bonding was systematically assessed, indicating that compound 13 possesses inhibitory potential due to the presence of additional water bridges (refer to Fig. 15).

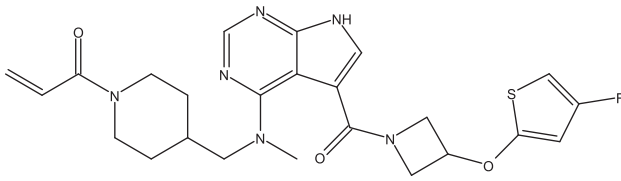
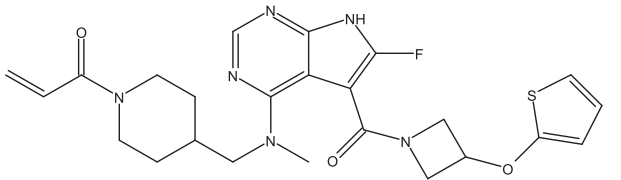
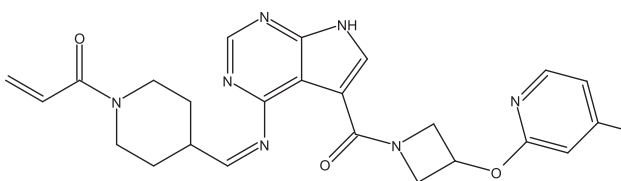
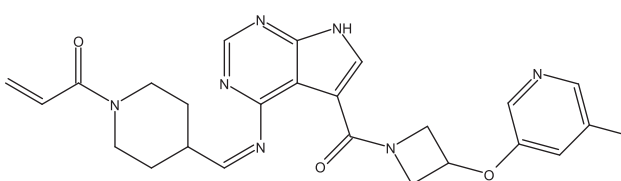
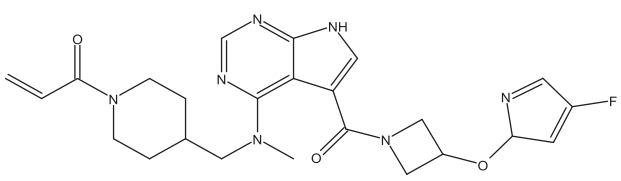
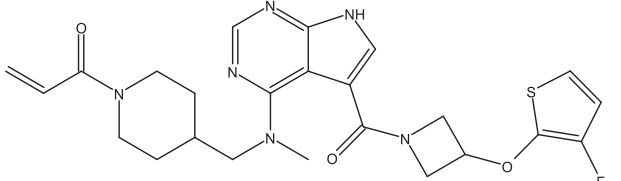
The ROG (radius of gyration) was used to estimate the rigidity of the protein structure. The impact of the Compound13_5P9J complex on the overall tightness of the protein (5P9J) was investigated, revealing a mild stretching effect within a range of 1.95–2.20 Å during the simulation. Throughout a 10 ns MD simulation, Compound 13 displayed a consistently stable gyration radius, averaging 5.8 Å. The reference molecule exhibited a solvent-accessible surface area (SASA) ranging from 120.22 to 180.30 Å², along with a significant polar surface area (PSA) between 135.10 and 150.30 Å², indicating its stability during the entire 10 ns MD simulation.

The descriptors of these new compounds were calculated in the same way as the molecules in the series studied. The predicted activity of the newly designed compounds was obtained by the RLM model developed Table 6. From the table we can see that the new candidate compounds have an inhibitory activity close to and greater than that of the most active compounds in the series studied.

This compliance indicates that these compounds (Table 7) have favorable physicochemical properties, suggesting a higher likelihood of good oral bioavailability and potential as effective drug candidates (Table 8). However, while adherence to Lipinski's Rule is a positive sign, it's important to remember that these rules serve as guidelines rather than definitive determinants. Additional assessments and evaluations, such as further preclinical studies, are still essential to fully assess the compounds viability as potential medications.

Details of the in-silico study are provided in Table 9. An absorbance

Table 7
Structures of new compounds and their pIC_{50} predicted on the basis of the 2D-QSAR model.

Compounds	Structure	pIC_{50_pred}
Pred1		9.69
Pred2		9.64
Pred3		9.16
Pred4		9.05
Pred5		9.49
Pred6		9.75

value below 30 % indicates inadequate absorption, while values exceeding 72 % suggest favorable absorption in the human intestines. Notably, a VD_{ss} (Volume of distribution) above 0.45 is considered substantial, indicating a wide distribution within the body. The enzymatic metabolism, which involves the biochemical transformation of drugs in the body, plays a critical role in altering drug compounds. This process leads to the formation of various enzymatic metabolites, which can influence drug reactions at different concentrations (Ahmed, 2015).

An in-depth examination of drug metabolism is crucial, considering the possible differences in physicochemical and pharmacological characteristics. Of particular importance in this regard is the participation of cytochrome P450 (CYP450), notably the CYP1, CYP2, CYP3, and CYP4 families, responsible for over 90 % of phase I metabolism (Ahmed, 2015; Šrejber et al., 2018). Among these, CYP3A4 holds particular significance in our study (Chandrasekaran et al., 2018), as the newly designed compounds display properties of both substrates and inhibitors for this

Table 8

Lipinski's role of newly designed compounds and their drug-like properties.

Rule	TPSA	n- Rotatable Bonds	MW	LogP	n-HBA	n-HBD	Violations				Synthetic accessibility	Toxicity
							Lipinski	Veber	Egan	Muegge		
	< 140 A ²	< 10	< 500 Da	≤ 5	< 10	< 5	≤ 2	≤ 2	≤ 2	≤ 2	0 < S.A < 10	AMES toxicity
Pred1	122.90	7	498.57	2.9227	7	1	Yes	Yes	Yes	Yes	3.66	Categorical (yes/no) No
Pred2	122.90	7	498.57	2.9227	7	1	Yes	Yes	Yes	Yes	3.67	No
Pred3	116.34	7	479.51	2.2319	7	2	Yes	Yes	Yes	Yes	3.53	No
Pred4	116.34	7	479.51	2.2319	7	2	Yes	Yes	Yes	Yes	3.49	No
Pred5	107.02	7	481.52	1.9236	7	1	Yes	Yes	Yes	Yes	4.48	No
Pred6	122.90	7	498.57	2.9227	7	1	Yes	Yes	Yes	Yes	3.61	No

Table 9

ADMET properties.

models	Properties														
	Absorption			Distribution			Metabolism						Excretion	Toxicity	
	Intestinal absorption (human)		P-Gp substrate	VDss (human)		CYP450						Total clearance	AMES toxicity		
	Numeric (%absorbed)			Categorical (yes/no)	Numeric (Log L kg ⁻¹)		Substrate		Inhibitor						
Unity	Numeric (%absorbed)		Categorical (yes/no)		Numeric (Log L kg ⁻¹)		Categorical (YES/NO)						Numeric (log mL min ⁻¹ kg ⁻¹)	Categorical (yes/no)	
Predicted values															
Pred1		80.952	Yes		0.756	No	Yes	No	Yes	Yes	Yes	Yes	Yes	0.773	No
Pred2		79.489	Yes		0.182	No	Yes	No	Yes	Yes	Yes	Yes	Yes	0.529	No
Pred3		72.649	Yes		0.639	No	Yes	No	Yes	Yes	Yes	Yes	Yes	0.389	No
Pred4		74.817	Yes		0.651	No	Yes	No	Yes	Yes	Yes	Yes	Yes	0.251	No
Pred5		76.617	yes		0.757	No	Yes	No	No	Yes	No	Yes	Yes	0.496	No
Pred6		83.922	Yes		0.793	No	Yes	No	Yes	Yes	Yes	Yes	Yes	0.652	No

enzyme.

Clearance, a parameter representing the relationship between drug concentration and elimination rate, determines the drug's duration in the body. As a result, the recently developed compounds show notably high clearance values, ensuring a favorable retention of the drug. Moreover, evaluating the non-toxic nature of the predicted compounds is crucial in the process of drug selection. Encouragingly, all the compounds we designed exhibit non-toxic characteristics, which significantly bolster their promise as potential candidates for medication.

5. Conclusion

The study aimed to create new pyrrolopyrimidine-derived drugs as BTK inhibitors. A 2D-QSAR study evaluated 35 analogues' structural components, constructing a QSAR model with predictive power. The molecule N°13 showed significant hydrogen bonding and frequent stability, indicating its potential in BTK inhibitory activity. In addition, six new molecules with potent BCR inhibitory activity were developed using various methods. These molecules have acceptable pharmacokinetic properties and favorable ADMET properties, making them potential agents for cancer treatment. Future research plans include developing 3D-QSAR models and synthesizing the predicted molecules.

CRediT authorship contribution statement

Mourad Aloui: Conceptualization, Methodology. **Mohammed Er-rajy:** Conceptualization, Methodology, Formal analysis. **Hamada Imtara:** Conceptualization, Resources, Data curation, Writing – review & editing. **Amina Goudzal:** Methodology. **Sara Zarougui:** Formal analysis. **Mohamed El fadili:** Investigation, Formal analysis. **David E. Arthur:** . **Ramzi A. Mothana:** Formal analysis, Writing – review & editing. **Omar M. Noman:** Resources, Writing – review & editing. **Mahmoud Tarayrah:** Data curation, Writing – review & editing.

Elhalaoui Menana: Writing – review & editing, Supervision.

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.

Acknowledgement

The authors extend their appreciation to Researchers Supporting Project number (RSP2023R119), King Saud University, Riyadh, Saudi Arabia for funding this work.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jsps.2023.101911>.

References

- T.A. Ahmed, 2015. Basic Pharmacokinetic Concepts and Some Clinical Applications. BoD – Books on Demand.
- Bassani, D., Pavan, M., Bolcato, G., Sturlese, M., Moro, S., 2022. Re-exploring the ability of common docking programs to correctly reproduce the binding modes of non-covalent inhibitors of SARS-CoV-2 protease Mpro. *Pharmaceuticals* 15, 180. <https://doi.org/10.3390/ph15020180>.
- Burger, J.A., 2014. Bruton's tyrosine kinase (BTK) inhibitors in clinical trials. *Curr. Hematol. Malig. Rep.* 9, 44–49. <https://doi.org/10.1007/s11899-013-0188-8>.
- Bye, A.P., Unsworth, A.J., Desborough, M.J., Hildyard, C.A.T., Appleby, N., Bruce, D., Kriek, N., Nock, S.H., Sage, T., Hughes, C.E., Gibbins, J.M., 2017. Severe platelet dysfunction in NHL patients receiving ibrutinib is absent in patients receiving acalabrutinib. *Blood Adv.* 1, 2610–2623. <https://doi.org/10.1182/bloodadvances.2017011999>.
- B. Chandrasekaran, S.N. Abed, O. Al-Attraqchi, K. Kuche, R.K. Tekade, 2018. Chapter 21 - Computer-Aided Prediction of Pharmacokinetic (ADMET) Properties, in: Tekade, R. K. (Ed.), Dosage Form Design Parameters, Advances in Pharmaceutical Product

- Development and Research. Academic Press, pp. 731–755. <https://doi.org/10.1016/B978-0-12-814421-3.00021-X>.
- de Vries, R., Smit, J.W., Hellemans, P., Jiao, J., Murphy, J., Skee, D., Snoeys, J., Sukuntherrg, J., Vliegen, M., de Zwart, L., Mannaert, E., de Jong, J., 2016. Stable isotope-labelled intravenous microdose for absolute bioavailability and effect of grapefruit juice on ibuprofen in healthy adults. *Br. J. Clin. Pharmacol.* 81, 235–245. <https://doi.org/10.1111/bcp.12787>.
- Di Paolo, J.A., Huang, T., Balazs, M., Barbosa, J., Barck, K.H., Bravo, B.J., Carano, R.A.D., Darrow, J., Davies, D.R., DeForge, L.E., Diehl, L., Ferrando, R., Gallion, S.L., Giannetti, A.M., Gribling, P., Hurez, V., Hymowitz, S.G., Jones, R., Kropf, J.E., Lee, W.P., Maciejewski, P.M., Mitchell, S.A., Rong, H., Staker, B.L., Whitney, J.A., Yeh, S., Young, W.B., Yu, C., Zhang, J., Reif, K., Currie, K.S., 2011. Specific Btk inhibition suppresses B cell- and myeloid cell-mediated arthritis. *Nat. Chem. Biol.* 7, 41–50. <https://doi.org/10.1038/nchembio.481>.
- Draper, B., Yee, W.L., Pedrana, A., Kyi, K.P., Qureshi, H., Htay, H., Naing, W., Thompson, A.J., Hellard, M., Howell, J., 2022. Reducing liver disease-related deaths in the Asia-Pacific: the important role of decentralised and non-specialist led hepatitis C treatment for cirrhotic patients. *Lancet Reg. Health – West. Pac.* 20. <https://doi.org/10.1016/j.lanwpc.2021.100359>.
- M. El Fadili, M. Er-Rajjy, M. Kara, A. Assouguem, A. Belhassan, A. Alotaibi, N.N. Mrabti, H. Fidan, R. Ullah, S. Ercisli, S. Zarougui, M. Elhallaoui, 2022. QSAR, ADMET In Silico Pharmacokinetics, Molecular Docking and Molecular Dynamics Studies of Novel Bicyclo (Aryl Methyl) Benzamides as Potent GlyT1 Inhibitors for the Treatment of Schizophrenia. *Pharmaceuticals* 15, 670. <https://doi.org/10.3390/ph15060670>.
- Er-rajy, M., El Fadili, M., Hadni, H., Mrabti, N.N., Zarougui, S., Elhallaoui, M., 2022a. 2D-QSAR modeling, drug-likeness studies, ADMET prediction, and molecular docking for anti-lung cancer activity of 3-substituted-5-(phenylamino) indolone derivatives. *Struct. Chem.* 33, 973–986. <https://doi.org/10.1007/s11224-022-01913-3>.
- Er-Rajy, M., El Fadili, M., Mujwar, S., Zarougui, S., Elhallaoui, M., 2023. Design of novel anti-cancer drugs targeting TRKs inhibitors based 3D QSAR, molecular docking and molecular dynamics simulation. *J. Biomol. Struct. Dyn.* 1–14. <https://doi.org/10.1080/07391102.2023.2170471>.
- M. Er-rajy, M. El Fadili, N.N. Mrabti, S. Zarougui, M. Elhallaoui, 2022b. QSAR, molecular docking, ADMET properties in silico studies for a series of 7-propanamide benzoxaboroles as potent anti-cancer agents. *Chin. J. Anal. Chem.* 50, 100163. <https://doi.org/10.1016/j.cjac.2022.100163>.
- M. Er-rajy, M. El Fadili, S. Mujwar, H. Imtara, O. Al kamaly, S. Zuhair Alshawwa, F.A. Nasr, S. Zarougui, M. Elhallaoui, 2023. Design of novel anti-cancer agents targeting COX-2 inhibitors based on computational studies. *Arab. J. Chem.* 105193. <https://doi.org/10.1016/j.arabj.2023.105193>.
- Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L.G., 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103, 8577–8593. <https://doi.org/10.1063/1.470117>.
- M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr, T. Vreven, K.N. Kudin, J.C. Burant, 2004. Gaussian 03, Revision C. 02. Wallingford, CT: Gaussian, IncGoogle Sch.
- Golbraikh, A., Shen, M., Xiao, Z., Xiao, Y.-D., Lee, K.-H., Tropsha, A., 2003. Rational selection of training and test sets for the development of validated QSAR models. *J. Comput. Aided Mol. Des.* 17, 241–253. <https://doi.org/10.1023/A:1025386326946>.
- Golbraikh, A., Tropsha, A., 2002. Beware of q²! *J. Mol. Graph. Model. QSAR in Vivo* 20, 269–276. [https://doi.org/10.1016/S1093-3263\(01\)00123-1](https://doi.org/10.1016/S1093-3263(01)00123-1).
- Goyal, V., Grimwood, K., Marchant, J., Masters, I.B., Chang, A.B., 2016. Pediatric bronchiectasis: No longer an orphan disease. *Pediatr. Pulmonol.* 51, 450–469. <https://doi.org/10.1002/ppul.23380>.
- M.K. Gupta, S. Gupta, R.K. Rawal, 2016. Chapter 8 - Impact of Artificial Neural Networks in QSAR and Computational Modeling, in: Puri, M., Pathak, Y., Sutariya, V.K., Tippetarij, S., Moreno, W. (Eds.), *Artificial Neural Network for Drug Design, Delivery and Disposition*. Academic Press, Boston, pp. 153–179. <https://doi.org/10.1016/B978-0-12-801559-9.00008-9>.
- Hansch, C., Leo, A., Mekapati, S.B., Kurrup, A., 2004. QSAR and ADME. *Bioorg. Med. Chem.* 12, 3391–3400. <https://doi.org/10.1016/j.bmc.2003.11.037>.
- Jang, S., Voth, G.A., 1997. Simple reversible molecular dynamics algorithms for Nosé-Hoover chain dynamics. *J. Chem. Phys.* 107, 9514–9526. <https://doi.org/10.1063/1.475247>.
- Jin, Z., Wang, Y., Yu, X.-F., Tan, Q.-Q., Liang, S.-S., Li, T., Zhang, H., Shaw, P.-C., Wang, J., Hu, C., 2020. Structure-based virtual screening of influenza virus RNA polymerase inhibitors from natural compounds: Molecular dynamics simulation and MM-GBSA calculation. *Comput. Biol. Chem.* 85, 107241. <https://doi.org/10.1016/j.compbiolchem.2020.107241>.
- Kalayan, J., Chakravorty, A., Warwicker, J., Henchman, R.H., 2023. Total free energy analysis of fully hydrated proteins. *Proteins Struct. Funct. Bioinforma.* 91, 74–90. <https://doi.org/10.1002/prot.26411>.
- Kalayan, J.H., Henchman, R., 2021. Convergence behaviour of solvation shells in simulated liquids. *Phys. Chem. Chem. Phys.* 23, 4892–4900. <https://doi.org/10.1039/D0CP05903J>.
- Kürková, V., 1992. Kolmogorov's theorem and multilayer neural networks. *Neural Netw.* 5, 501–506. [https://doi.org/10.1016/0893-6080\(92\)90012-8](https://doi.org/10.1016/0893-6080(92)90012-8).
- Martyna, G.J., Klein, M.L., Tuckerman, M., 1992. Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* 97, 2635–2643. <https://doi.org/10.1063/1.463940>.
- Milne, G.W.A., 2010. Software review of ChemBioDraw 12.0. *J. Chem. Inf. Model.* 50, 2053. <https://doi.org/10.1021/ci100385n>.
- Mrabti, N.N., Mrabti, H.N., Khalil, Z., Bouyahya, A., Mohammed, E.R., Dguigui, K., Doudach, L., Zengin, G., Elhallaoui, M., 2022. Molecular docking and QSAR studies for modeling the inhibitory activity of pyrazole-benzimidazolone Hybrids as Novel Inhibitors of Human 4-hydroxyphenylpyruvate dioxygenase against type I tyrosinemia disease. *Biointerface Res Appl Chem* 13, 38.
- Nour, H., Abdou, A., Belaidi, S., Jamal, J., Elmakssoudi, A., Dakir, M., Chtita, S., 2022. Discovery of promising cholinesterase inhibitors for Alzheimer's disease treatment through DFT, docking, and molecular dynamics studies of eugenol derivatives. *J. Chin. Chem. Soc.* 69, 1534–1551. <https://doi.org/10.1002/jccs.202200195>.
- Österberg, T., Norinder, U., 2001. Prediction of drug transport processes using simple descriptors and PLS statistics The use of ACD/LogP and ACD/ChemSketch descriptors. *Eur. J. Pharm. Sci.* 12, 327–337. [https://doi.org/10.1016/S0928-0987\(00\)00189-5](https://doi.org/10.1016/S0928-0987(00)00189-5).
- Parr, R.G., Yang, W., 1995. Density-functional theory of the electronic structure of molecules. *Annu. Rev. Phys. Chem.* 46, 701–728. <https://doi.org/10.1146/annurev.pc.46.100195.003413>.
- Patra, P.K., Bhattacharya, B., 2014. Nonergodicity of the Nose-Hoover chain thermostat in computationally achievable time. *Phys. Rev. E* 90, 043304. <https://doi.org/10.1103/PhysRevE.90.043304>.
- Petersen, H.G., 1995. Accuracy and efficiency of the particle mesh Ewald method. *J. Chem. Phys.* 103, 3668–3679. <https://doi.org/10.1063/1.470043>.
- Pires, D.E.V., Blundell, T.L., Ascher, D.B., 2015. pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.* 58, 4066–4072. <https://doi.org/10.1021/acs.jmedchem.5b00104>.
- Puri, K.D., Di Paolo, J.A., Gold, M.R., 2013. B-cell receptor signaling inhibitors for treatment of autoimmune inflammatory diseases and B-cell malignancies. *Int. Rev. Immunol.* 32, 397–427. <https://doi.org/10.3109/08830185.2013.818140>.
- Rezaei, N., Hedayat, M., Aghamohammadi, A., Nichols, K.E., 2011. Primary immunodeficiency diseases associated with increased susceptibility to viral infections and malignancies. *J. Allergy Clin. Immunol.* 127, 1329–1341.e2. <https://doi.org/10.1016/j.jaci.2011.02.047>.
- Rose, Y., Duarte, J.M., Lowe, R., Segura, J., Bi, C., Bhikadiya, C., Chen, L., Rose, A.S., Bittrich, S., Burley, S.K., Westbrook, J.D., 2021. RCSB protein data bank: Architectural advances towards integrated searching and efficient access to macromolecular structure data from the PDB archive. *J. Mol. Biol., Comput. Resour. Mol. Biol.* 433, 166704. <https://doi.org/10.1016/j.jmb.2020.11.003>.
- Roy, K., Mitra, I., 2011a. On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. *Comb. Chem. High Throughput Screen.* 14, 450–474. <https://doi.org/10.2174/138620711795767893>.
- Roy, K., Mitra, I., 2011b. On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. *Comb. Chem. High Throughput Screen.* 14, 450–474. <https://doi.org/10.2174/138620711795767893>.
- Rücker, C., Rücker, G., Meringer, M., 2007. y-Randomization and its variants in QSAR/QSAR. *J. Chem. Inf. Model.* 47, 2345–2357. <https://doi.org/10.1021/ci700157b>.
- Salemi, S., Markovic, M., Martini, G., D'Amelio, R., 2015. The expanding role of therapeutic antibodies. *Int. Rev. Immunol.* 34, 202–264. <https://doi.org/10.3109/08830185.2013.863304>.
- Salomon-Ferrer, R., Case, D.A., Walker, R.C., 2013. An overview of the Amber biomolecular simulation package. *Wires Comput. Mol. Sci.* 3, 198–210. <https://doi.org/10.1002/wcms.1121>.
- Salt, D.W., Yildiz, N., Livingstone, D.J., Tinsley, C.J., 1992. The use of artificial neural networks in QSAR. *Pestic. Sci.* 36, 161–170. <https://doi.org/10.1002/ps.2780360212>.
- Satterthwaite, A.B., Li, Z., Witte, O.N., 1998. Btk function in B cell development and response. *Semin. Immunol.* 10, 309–316. <https://doi.org/10.1006/smim.1998.0123>.
- Senis, Y.A., Mazharian, A., Mori, J., 2014. Src family kinases: at the forefront of platelet activation. *Blood* 124, 2013–2024. <https://doi.org/10.1182/blood-2014-01-453134>.
- Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., Sherman, W., 2010. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theory Comput.* 6, 1509–1519. <https://doi.org/10.1021/ct900587b>.
- Shivakumar, D., Harder, E., Damm, W., Friesner, R.A., Sherman, W., 2012. Improving the prediction of absolute solvation free energies using the next generation OPLS force field. *J. Chem. Theory Comput.* 8, 2553–2558. <https://doi.org/10.1021/ct300203w>.
- Singh, J., 2022. The Ascension of Targeted Covalent Inhibitors. *J. Med. Chem.* 65, 5886–5901. <https://doi.org/10.1021/acs.jmedchem.1c02134>.
- Singh, J., Petter, R.C., Baillie, T.A., Whitty, A., 2011. The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* 10, 307–317. <https://doi.org/10.1038/nrd3410>.
- Spasic, A., Serafini, J., Mathews, D.H., 2012. The amber ff99 force field predicts relative free energy changes for RNA helix formation. *J. Chem. Theory Comput.* 8, 2497–2505. <https://doi.org/10.1021/ct300240k>.
- Šrejber, M., Navrátilová, V., Paloncýová, M., Bazgier, V., Berka, K., Anzenbacher, P., Otyepka, M., 2018. Membrane-attached mammalian cytochromes P450: An overview of the membrane's effects on structure, drug binding, and interactions with redox partners. *J. Inorg. Biochem.* 183, 117–136. <https://doi.org/10.1016/j.jinorgbio.2018.03.002>.
- Tangye, S.G., Ma, C.S., Brink, R., Deenick, E.K., 2013. The good, the bad and the ugly — TFH cells in human health and disease. *Nat. Rev. Immunol.* 13, 412–426. <https://doi.org/10.1038/nri3447>.
- Trott, O., Olson, A.J., 2010. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. <https://doi.org/10.1002/jcc.21334>.

- Vickers, N.J., 2017. Animal communication: When I'm calling you, will you answer too? *Curr. Biol.* 27, R713–R715. <https://doi.org/10.1016/j.cub.2017.05.064>.
- Yang, M., Jiang, H., Yang, Z., Liu, X., Sun, H., Hao, M., Hu, J., Chen, X., Jin, J., Wang, X., 2022. Design, synthesis, and biological evaluation of pyrrolopyrimidine derivatives as novel Bruton's tyrosine kinase (BTK) inhibitors. *Eur. J. Med. Chem.* 241, 114611 <https://doi.org/10.1016/j.ejmech.2022.114611>.
- Zhang, Y., Xu, X., Goddard, W.A., 2009. Doubly hybrid density functional for accurate descriptions of nonbond interactions, thermochemistry, and thermochemical kinetics. *Proc. Natl. Acad. Sci.* 106, 4963–4968. <https://doi.org/10.1073/pnas.0901093106>.