QSAR-driven screening uncovers and designs novel pyrimidine-4,6-diamine derivatives as potent JAK3 inhibitors

Abdelmoujoud Faris, Ibrahim M. Ibrahim, Radwan Alnajjar, Hanine Hadni, Mashooq Ahmad Bhat, Muhammad Yaseen, Souvik Chakraborty, Nada Alsakhen, Israa M. Shamkh, Fazal Mabood, Ahmed M. Naglah, Ihsan Ullah, Noha Ziedan & Menana Elhallaoui

Published online: 07 Dec 2023.

Submit your article to this journal ⤤

Article views: 411

View related articles ⤤

View Crossmark data ⤤

QSAR-driven screening uncovers and designs novel pyrimidine-4,6-diamine derivatives as potent JAK3 inhibitors

Abdelmoujoud Faris[a] , Ibrahim M. Ibrahim[b] , Radwan Alnajjar[c] , Hanine Hadni[a] , Mashooq Ahmad Bhat[d] , Muhammad Yaseen[e] , Souvik Chakraborty[f] , Nada Alsakhen[g] , Israa M. Shamkh[h] , Fazal Mabood[e] , Ahmed M. Naglah[i] , Ihsan Ullah[e], Noha Ziedan[j] and Menana Elhallaoui[a]

[a]LIMAS, Department of Chemical Sciences, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco; [b]Biophysics Department, Faculty of Science, Cairo University, Cairo, Egypt; [c]Department of Chemistry, Faculty of Science, University of Benghazi, Benghazi, Libya; [d]Department of Pharmaceutical Chemistry, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia; [e]Institute of Chemical Sciences, University of Swat, Main Campus, Charbagh, Swat, Pakistan; [f]Department of Physiology, Bhairab Ganguly College, Belghoria, Kolkata, West Bengal, India; [g]Department of Chemistry, Faculty of Science, The Hashemite University, Zarqa, Jordan; [h]Botany and Microbiology Department, Faculty of Science, Cairo University, Cairo, Egypt; [i]Department of Pharmaceutical Chemistry, College of Pharmacy, King Saud University, Riyadh, Saudi Arabia; [j]Department of Physical, Mathematical and Engineering Science, Faculty of Science, Business and Enterprise, University of Chester, Chester, UK

Communicated by Ramaswamy H. Sarma

ABSTRACT

This study presents a robust and integrated methodology that harnesses a range of computational techniques to facilitate the design and prediction of new inhibitors targeting the JAK3/STAT pathway. This methodology encompasses several strategies, including QSAR analysis, pharmacophore modeling, ADMET prediction, covalent docking, molecular dynamics (MD) simulations, and the calculation of binding free energies (MM/GBSA). An efficacious QSAR model was meticulously crafted through the employment of multiple linear regression (MLR). The initial MLR model underwent further refinement employing an artificial neural network (ANN) methodology aimed at minimizing predictive errors. Notably, both MLR and ANN exhibited commendable performance, showcasing R2 values of 0.89 and 0.95, respectively. The model's precision was assessed via leave-one-out cross-validation (CV) yielding a Q2 value of 0.65, supplemented by rigorous Y-randomization. , The pharmacophore model effectively differentiated between active and inactive drugs, identifying potential JAK3 inhibitors, and demonstrated validity with an ROC value of 0.86. The newly discovered and designed inhibitors exhibited high inhibitory potency, ranging from 6 to 8, as accurately predicted by the QSAR models. Comparative analysis with FDA-approved Tofacitinib revealed that the new compounds exhibited promising ADMET properties and strong covalent docking (CovDock) interactions. The stability of the new discovered and designed inhibitors within the JAK3 binding site was confirmed through 500 ns MD simulations, while MM/GBSA calculations supported their binding affinity. Additionally, a retrosynthetic study was conducted to facilitate the synthesis of these potential JAK3/STAT inhibitors. The overall integrated approach demonstrates the feasibility of designing novel JAK3/STAT inhibitors with robust efficacy and excellent ADMET characteristics that surpass Tofacitinib by a significant margin.

QSAR-Driven Screening Uncovers and Designs Novel Pyrimidine-4,6-diamine Derivatives as Potent JAK3 Inhibitors

Computer-Aided Drug Design
Molecular Docking
ADMET
Molecular Dynamics
Free Binding Energy
Retrosynthesis

Novel Inhibitors JAK3
Identifier and Design

1. Introduction

The regulation of cellular growth, differentiation, and maintenance relies on an important signalling pathway that involves Janus Kinase (JAK) family and STAT transcription factors. Dysregulation of this pathway can lead to autoimmune diseases like rheumatoid arthritis (RA) and cancer (Banerjee

et al., 2017; Xin et al., 2020). One subunit of JAK, known as JAK3, plays a role in signalling interleukin-2 (IL-2) and other cytokines responsible for immune response regulation. Mutations in the Cys909 residue of JAK3 can result in constitutive activation of the kinase, leading to the development of autoimmune diseases (Bryan & Rajapaksa, 2018; Tan et al., 2015). Therefore, inhibiting JAK3 has emerged as a promising therapeutic approach for RA and other autoimmune diseases. JAK3 inhibitors have shown effectiveness in alleviating RA symptoms and slowing down disease progression (Chen, Lu, et al., 2022; Fetter et al., 2020). Extensive research has demonstrated that JAK3 inhibitors can reduce the production of pro-inflammatory cytokines, such as IL-2, IL-6, and IL-17, which are implicated in the development of RA (Barf & Kaptein, 2012; Zhong & Almahmoud, 2023), which are implicated in the development of RA (Wang et al., 2023). Extensive research has demonstrated that JAK3 inhibitors can reduce the production of pro-inflammatory cytokines, such as IL-2, IL-6, and IL-17 (Ding et al., 2023; Faris, Hadni, Ibrahim, et al., 2023; Faris, Hadni, Saleh, et al., 2023). Furthermore, these inhibitors may also decrease the production of growth factors and cytokines involved in cell proliferation and cancer cell survival, potentially slowing down cancer progression (Ghoreschi et al., 2009; Malemud & Pearlman, 2009; Salas et al., 2020). Given the involvement of JAK3 and STAT signaling in various autoimmune diseases and cancer, inhibiting JAK3 shows promise as a therapeutic strategy. However, further research is needed to fully understand the roles of JAK3 and STAT in these diseases and to develop more specific and effective JAK3 inhibitors (Alunno et al., 2019; Banerjee et al., 2017; Hosseini et al., 2020; O'Shea et al., 2013).

Inhibiting distinct cytokine pathways offers a promising approach to addressing various medical conditions and has demonstrated notable efficacy. For instance, treatments targeting anti-IL12/23 or more selective anti-IL23 agents have received approval for managing psoriasis, psoriatic arthritis, and have recently been granted approval in both Europe and the United States for treating patients with Crohn's disease, exhibiting clear effectiveness and a favorable safety profile (Lai & Dong, 2016). Nonetheless, it is important to note that therapeutic antibodies, despite these therapeutic advancements, may still be associated with immunogenicity concerns and are currently limited to parenteral administration. JAK inhibitors represent a class of small molecules that are currently available in the market or are in the development pipeline for the treatment of various immune-related diseases, including psoriasis, RA, and inflammatory bowel disease (IBD) (Gadina et al., 2018). The JAK family encompasses four tyrosine kinases, namely, JAK1, JAK2, JAK3, and TYK2, which interact with the intracellular domain of a wide array of cytokines and certain growth factor receptor chains. Activation of JAKs is initiated through conformational changes induced by ligands in receptor complexes, triggering a phosphorylation cascade that leads to the activation of members of the signal transducer and activator of transcription (STAT) family (Figure 1) (Kurdi & Booz, 2009). Subsequently, phosphorylated STATs translocate to the nucleus, modulating gene expression in a ligand-dependent manner (Figure 1).

Research has established that different cytokine receptors are linked to specific JAK enzyme heterodimers or homodimers, with each JAK being associated with more than one receptor (Figure 2). Numerous studies have indicated that JAK1/JAK3 plays a role in γ-common chain cytokines, the JAK1/JAK2 complex is connected to INFγ, IL-6, and other gp130 cytokines, while the JAK1/TYK2 heterodimer binds to type I interferons and the IL-10 family of cytokines. JAK2, in contrast, forms a homodimer on EPO and leptin receptors.

Several JAK inhibitors have made it to the market or are currently in clinical development, each with its unique profile (Figure 3). For instance, Ruxolitinib, the first JAK inhibitor to reach the market, is a JAK1/JAK2 inhibitor used to treat polycythemia vera or myelofibrosis (McKeage, 2015). Tofacitini, the first JAK inhibitor used to combat autoimmune diseases, is a PAN JAK inhibitor, displaying some selectivity toward
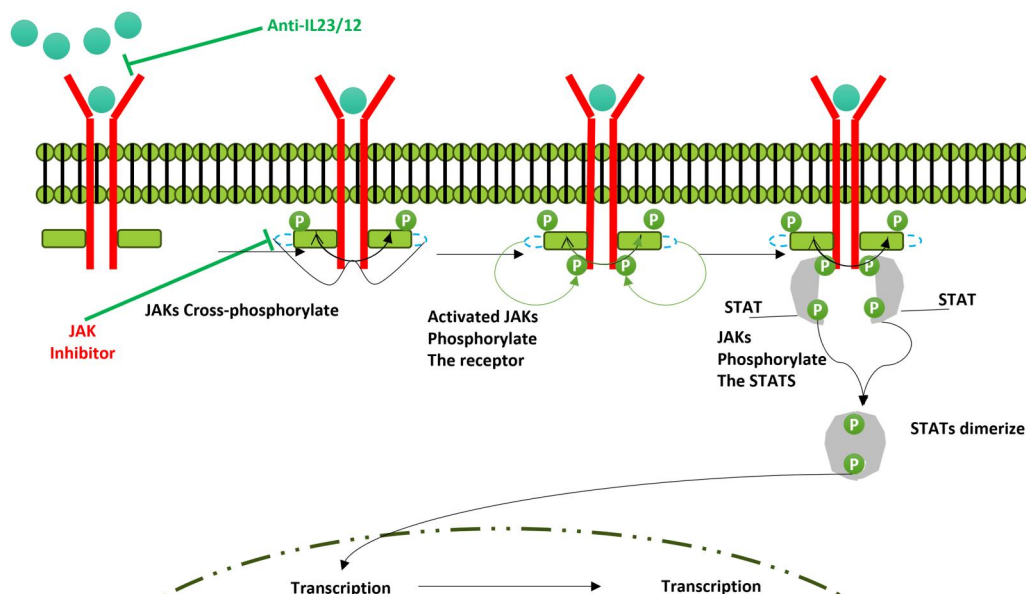
Figure 1. Signaling through cytokines.
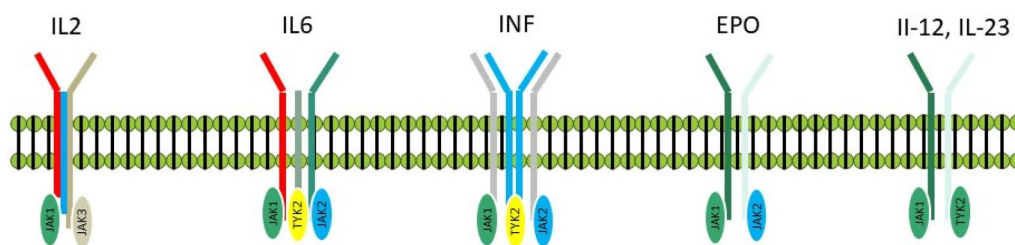
Figure 2. JAK kinases form heterodimers with cytokine receptors.
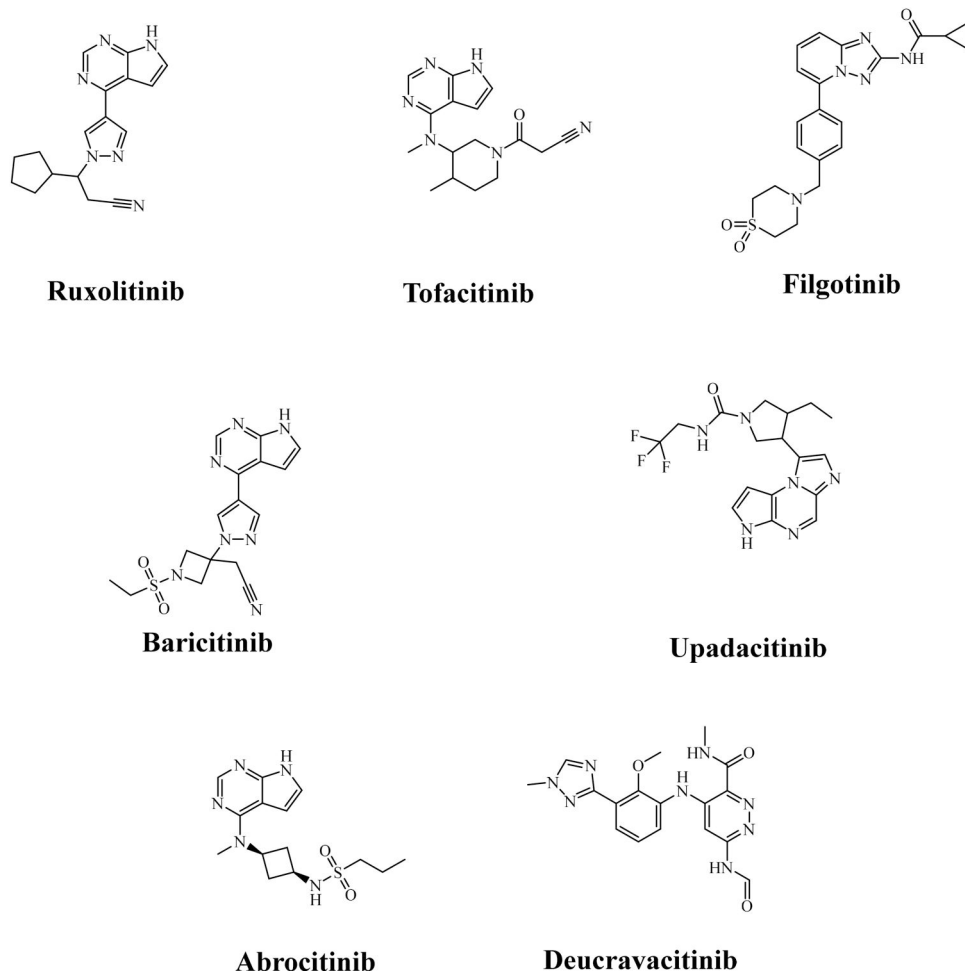
Figure 3. Depictions of JAK inhibitors or JAK-targeted drugs.

JAK1, JAK2, and JAK3 (Emery et al., 2018). With an enhanced understanding of JAKs and their involvement in cytokine receptors, pharmaceutical companies have been diligently working over the past decade to develop compounds with more refined selectivity profiles capable of modulating different cytokine signaling pathways. Several JAK1 inhibitors, such as filgotinib, Upadacitinib, and Abrocitinib, or TYK2 inhibitors like Deucravacitinib, are currently in clinical development for autoimmune diseases such as RA, psoriasis, or Crohn's disease (Hosking et al., 2018; White et al., 2018).

For a favorable comparison in this study, the comparison with the new inhibitors is made with Tofacitinib drug, as Tofacitinib predominantly inhibits JAK3, with minor effects on JAK1 and JAK2. Tofacitinib lowers inflammation and slows disease development by inhibiting JAK3. Tofacitinib, a JAK3 inhibitor, effectively heals rheumatoid arthritis (Hu et al.,

2022), a key player in immune cell differentiation and proliferation. Tofacitinib has been associated to several serious side effects in addition to its effectiveness. As a result, new inhibitors with unique designs are being investigated, with the potential to give more efficacy while having fewer side effects. The creation of novel pharmaceuticals has become more reliant on computer-aided drug development (CADD). It speeds up the process by immediately identifying potential drug candidates from enormous libraries of chemical substances. CADD lowers clinical trial costs and risks by selecting the most promising medications more quickly and precisely than earlier techniques. Scientists may use CADD to create and improve pharmaceuticals in a rational and effective manner, perhaps leading to the creation of innovative therapies for previously incurable ailments (Kapetanovic, 2008; Nascimento et al., 2021, 2022).

Several computational techniques were utilized, encompassing Quantitative Structure-Activity Relationship (QSAR) analysis, pharmacophore modeling, ADMET prediction, CovDock, MD simulation, Molecular Electrostatic Potential (MEP), and Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) calculations. All these techniques were assigned distinct roles in the design and prediction of JAK3/STAT inhibitors. In addition, MLR and ANN methodologies were applied to predict the properties of the newly designed compounds (Hadni & Elhallaoui, 2019). Pharmacophore models were constructed to identify essential chemical features required for compounds to interact effectively with the JAK3 protein. These models aided in the design and prediction of new inhibitors by serving as a framework for the generation of potential drug candidates (Faris, Ibrahim, Al kamaly, et al., 2023; Faris, Ibrahim, Hadni, et al., 2023). ADMET properties are crucial in drug development. Computational tools were used to predict the absorption, distribution, metabolism, excretion, and toxicity characteristics of the newly designed compounds, ensuring they have favourable pharmacokinetic profiles (Faris, Cacciatore, Ibrahim, et al., 2023; Ferreira & Andricopulo, 2019). CovDock techniques were applied to study the binding interactions between the designed inhibitors and the JAK3 protein, providing insights into the formation of covalent bonds and the strength of binding contacts (Zhong & Almahmoud, 2023). MEP Is a crucial tool in designing new molecules to target JAK3. By analysing the MEP of existing molecules and the active sites of JAK3, we can understand potential electrostatic interactions and more precisely design new inhibitors. This MEP-based approach allows us to rationalize the design of potent molecules, optimizing their electrostatic properties for improved binding and greater efficacy against JAK3, thereby paving the way for new advancements in drug development (Daoui et al., 2023). MD simulations were employed to assess the stability and dynamic behaviour of the newly proposed inhibitors within the JAK3 binding site. This technique helped evaluate the compounds' ability to maintain stable interactions over time (En-Nahli et al., 2022; Kukol, 2014). Free binding Energy calculations were used to estimate the binding affinity of the inhibitors. This approach provided a quantitative measure of the strength of binding between the compounds and the target protein (Faris, Cacciatore, Ibrahim, et al., 2023; Faris, Ibrahim, Al kamaly, et al., 2023; Ylilauri & Pentikäinen, 2013). The combined use of these computational techniques led to the identification and design of new compounds with potential JAK3 inhibitory activity. The study also used QSAR models to estimate the pIC50 values of these compounds, demonstrating their efficacy in inhibiting JAK3. The findings suggested that these newly anticipated ligands could be promising candidates for JAK3 inhibition and potential alternatives to the medication tofacitinib-drug. Additionally, a retrosynthetic study was conducted to simplify the synthesis of these prospective JAK3/STAT inhibitors. All approaches used in this study played specific roles in the design and prediction of newly designed JAK3/STAT inhibitors, making them favorable for *in vitro* studies.

2. Methods and materials

2.1. Data set

Thirty pyrimidine-4,6-diamines were obtained from earlier studies were study (Yu et al., 2019) for JAK3/STAT inhibiting efficacy. A dataset of 21 molecules was used for training, while 9 molecules were saved for testing. To facilitate compound analysis and comparison, the IC_{50} values (reported in nM) were converted into pIC_{50} (Exp) values using the formula $pIC50 = -\log (IC_{50})$. As shown by the pIC_{50}(Exp) values in Table 1, the obtained results exhibited a variety of inhibitory actions among the pyrimidine-4,6-diamine derivatives.

2.2. Software

Various software programs were used to conduct various analyses over the course of this investigation. Xlstat and JMPpro17 software were used for statistical analyses such as MLR, ANN, and LOO (Cox & Gaudard, 2013). Maestro's (Schrödinger Release, 2021-1; *Maestro, Schrödinger, LLC: New York, NY, USA, 2021*) CovDock module was used to conduct covalent docking. The pkCSM web server was used to construct ADMET predictions (Pires et al., 2015), with descriptor calculations executed using the PaDEL-Descriptor software as detailed in the supplementary file (Yap, 2011). For the visualization of molecular structures in both 2D and 3D, BIOVIA Studio software was employed (Systèmes, 2020), while ChemSketch software was utilized for molecule design (*Free Chemical Drawing Software for Students | ChemSketch | ACD/Labs*, n.d.). Furthermore, to investigate molecular interactions and assess binding free energy, MD simulations and Molecular Mechanics Generalized Born Surface Area (MM/GBSA) calculations were carried out employing Gromacs software (Abraham et al., 2023).

2.3. 2D-QSAR model

2.3.1. Calculations of 2D-QSAR descriptors
As detailed in the supplemental file, the moe and PaDEL-Descriptor programs (Moe, 2014; Yap, 2011), were used to construct a total of 380 descriptors for a collection of molecules. Based on four descriptors, the best-performing models were found among these descriptors.

2.3.2. Statistical analysis
The stepwise strategy, in conjunction with MLR and ANN methodologies, was used in the construction of the QSAR model. To discover the ideal collection of variables, the stepwise technique, a statistical methodology, was used repeatedly to add or delete variables from the model depending on their statistical significance (Chowdhury & Turin, 2020; Faris, Hadni, Saleh, et al., 2023).
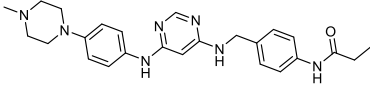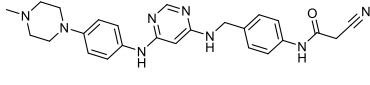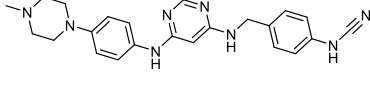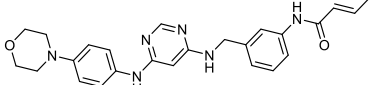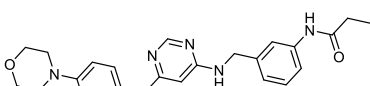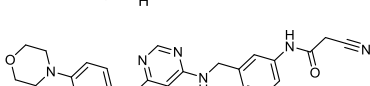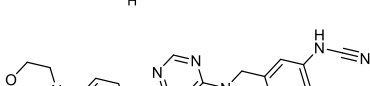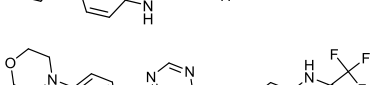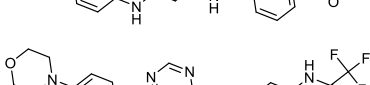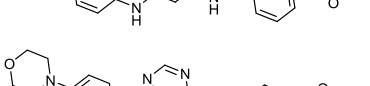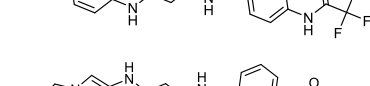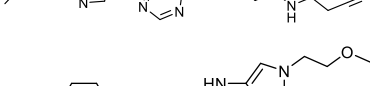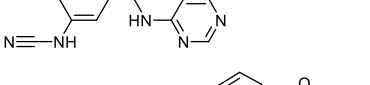
The stepwise regression are as follows: Forward selection: This approach begins with an empty model and evaluates the inclusion of each variable based on a chosen model fit criterion. The variable that results in the most statistically significant improvement in fit is added to the model. The

Table 1. The experimental values and pIC$_{50}$ prediction.

Compounds	Structure 2D	IC$_{50}$ (Exp)	pIC$_{50}$	pIC$_{50}$ MLR	Residual	pIC$_{50}$ LOO (CV)	Residual	pIC$_{50}$ ANN	Residual
1		2,36	6.14	5.56	0.58	5.39	0.56	6.19	−0.047
2		1,93	6.23	5.63	0.60	5.47	0.57	6.16	0.073
3		0,126	8.91	8.14	0.78	7.98	0.86	8.27	0.641
4		0,222	7.62	8.14	−0.52	8.24	0.38	8.27	−0.649
5		0,139	8.14	8.07	0.07	8.04	0.01	8.08	0.059
6		0,160	8.03	8.07	−0.04	8.08	0.00	8.08	−0.051
7		0,137	8.15	7.61	0.55	7.52	0.39	8.15	−0.003
8		0,135	8.16	7.61	0.56	7.52	0.41	8.15	0.007
9		0,042	9.62	8.54	1.08	7.62	2.00	7.08	2.540
10*		0,559	7.08	7.62	−0.54	7.75	0.45	7.08	0
11*		0,087	8.68	8.89	−0.21	8.15	0.53	8.10	0.578
12		0,155	8.10	8.15	−0.05	8.17	0.01	8.10	−0.002
13*		2,41	6.05	6.92	−0.87	6.31	0.26	5.96	0.089
14*		0,796	6.67	5.82	0.86	6.23	0.44	5.95	0.721
15		0,107	8.46	8.29	0.17	8.19	0.07	8.53	−0.065
16		290	5.52	5.75	−0.23	5.82	0.09	5.55	−0.028
17		77,5	5.99	6.31	−0.32	6.37	0.14	5.96	0.029

(continued)

Table 1. Continued.

Compounds	Structure 2D	IC$_{50}$ (Exp)	pIC$_{50}$	pIC$_{50}$ MLR	Residual	pIC$_{50}$ LOO (CV)	Residual	pIC$_{50}$ ANN	Residual
18		252	5.54	6.23	−0.69	6.33	0.62	5.95	−0.409
19*		0,394	7.81	8.23	−0.42	8.29	0.48	8.53	−0.715
20		290	5.52	5.75	−0.23	5.82	0.09	5.55	−0.028
21		100	6.00	5.78	0.22	5.73	0.07	6.10	−0.099
22*		228	5.56	5.80	−0.24	5.70	0.14	5.61	−0.049
23*		0,512	7.32	7.52	−0.20	7.76	0.44	8.37	−1.049
24		290	5.52	5.22	0.30	5.13	0.15	5.52	0.003
25*		290	5.52	5.52	0.00	5.33	0.19	5.52	−0.002
26		290	5.52	5.33	0.19	5.05	0.22	5.52	−0.002
27*		0,139	8.14	8.14	0.00	7.74	0.40	8.02	0.123
28		290	5.52	5.20	0.32	4.95	0.32	5.52	−0.002
29		344	5.46	6.29	−0.83	6.49	1.06	6.10	−0.643
30		378	5.41	6.29	−0.88	6.50	1.19	6.10	−0.693

*Test.

process is then repeated until no additional variables improve the model to a statistically significant extent. Backward elimination: In this method, all candidate variables are initially included in the model. The deletion of each variable is tested using a selected model fit criterion. The variable that causes the least statistically significant deterioration in model fit is removed. The process is iterated until no further variables can be deleted without a statistically significant loss of fit. Bidirectional elimination: This technique combines elements of forward selection and backward elimination. It involves testing for the inclusion or exclusion of variables at each step. The process evaluates variables for inclusion based on the most statistically significant improvement in fit and evaluates variables for exclusion based on the least statistically significant deterioration in fit. This bidirectional process continues until no further variables can be added or removed without a statistically significant impact on the model fit.

Figure 4. Structure diagram of a multilayer perceptron.

Previously, a comprehensive model was created that included all accessible descriptors. Following that, variables were successively added or eliminated based on their statistical significance, using a p-value threshold of 0.05. MLR was used as a statistical approach to investigate the connection between the dependent variable (Property) and several independent factors (Descriptors). The primary aim of this strategy is the mitigation of disparities existing between empiri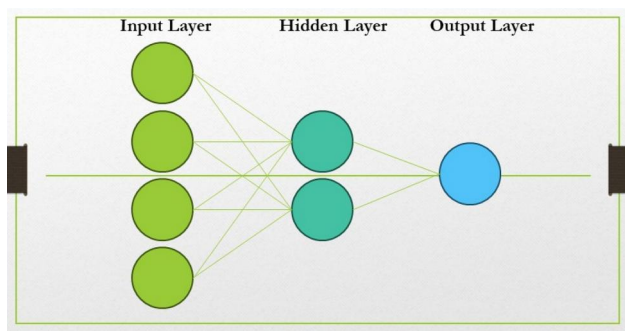cally observed and projected values. Artificial neural networks (ANNs) manifest as intricate networks of interconnected cells engaged in the transmission and reception of data, thereby mirroring the structural and functional characteristics of biological neurons as found in the human nervous system. These networks' mathematical representations are made up of numerous layers of neurons. A multilayer perceptron (MLP) network was used in this investigation, with neurons grouped into layers.

The artificial neural network was built with a 4-2-1 design, which included four neurons in the input layer, two neurons in the hidden layer, and one neuron in the output layer. A sigmoid function was used in the output layer. Concurrently, the ultimate layer comprises a singular neuron, assigned the responsibility of encapsulating the logarithmic IC50 values. In the intermediate, or concealed, layer, there exist a pair of neurons, a count determined by the parameter ρ, which is formulated as the ratio between the aggregate weight count and the overall connection count within the neural network. It is worth emphasizing that in the context of this specific investigation, the ρ value is constrained within the interval of $1 < \rho < 3$, as elucidated within pertinent scholarly works. Figure 4 provides a visual representation of the architecture characterizing the utilized Artificial Neural Network (ANN) models (Faris, Hadni, Saleh, et al., 2023; Hadni et al., 2018).

2.3.3. Assessment of predictive ability of a QSAR model
Various metrics, including the coefficient of determination, are used to evaluate the quality of a model (R^2), standard deviation (S) or root mean square error (RMSE), cross-validation correlation coefficients (Q^2_{CV}), and Fischer (F). The R^2, S, and F values indicate the degree of correspondence between simulated and experimental data, signifying the model's predictive capacity within its constraints and enabling estimation of the accuracy of the test set values (Snedecor & Cochran, 1967). The Q^2_{CV} coefficient, which is referred to as internal since it is computed from the structures used to

generate the model, provides information on the model's predictive capabilities. The R^2 value measures the difference between theoretical and experimental values, with higher modelling quality observed when the points are closer to the fit line (Esposito et al., 2004). The coefficient of determination may be employed to evaluate the fit of points to this line.

$$R^2 = 1 - \frac{\sum \left(y_{i.exp} - \hat{y}_{i, theo}\right)^2}{\sum \left(y_{i.exp} - \overline{y}_{i, exp}\right)^2}, \tag{1}$$

where, the experimental value of the anti-JAK3 activity is represented by $y_{i, exp}$, while the theoretical value of the anti-JAK3 activity is denoted by $y_{i, theo}$. The mean value of experimental values of anti-JAK3 activity is indicated by $\overline{y}_{i, exp}$. A higher R^2 value indicates a stronger correlation between theoretical and experimental values. Another statistical measure is the RMSE, which facilitates evaluation. The equation for R^2 is as follows:

$$R^2 = \sqrt{\frac{\sum \left(y_{i.exp} - y_{i, theo}\right)^2}{n - k - 1}}, \tag{2}$$

where: n is the number of observations and k is the number of independent variables.

The Fisher F test is a method utilized to determine the statistical significance of the model, which assesses the quality of the chosen descriptors from the model. The equation for the Fisher F test is given by:

$$F = \frac{\sum \left(y_{i.theo} - y_{i, exp}\right)^2}{\sum \left(y_{i.exp} - y_{i, theo}\right)^2} * \frac{n - k - 1}{k}, \tag{3}$$

A cross-validation coefficient of determination is employed to evaluate the prediction accuracy of the training set. It is calculated using the following formula:

$$Q^2_{CV} = \frac{\sum \left(y_{i.theo} - \overline{Y}_{i, exp}\right)^2 - \sum \left(y_{i.theo} - y_{i, exp}\right)^2}{\sum \left(y_{i.theo} - y_{i, exp}\right)^2}, \tag{4}$$

2.3.4. Criteria for accepting a model
Eriksson et al. (2003) define the performance of a mathematical model as satisfactory when it achieves a value of 0.5 and exceptional when it reaches a value of 0.9. If the acceptance criteria are met, the model is expected to perform well during testing. Golbraikh and Tropsha (2002), Ouattara et al. (2017), and Tropsha et al. (2003) propose five criteria, including the external validation set, to determine the predictive capability of a model (Table 6).

2.4. Validation of the QSAR model

2.4.1. Y-randomization
The Y-randomization test is a popular approach for determining a model's resilience. The dependent variable (biological activity) is randomly given values in this technique, while the independent variable (chosen descriptors) remains constant (Faris, Hadni, Saleh, et al., 2023). After then, the randomized data is utilized to build a new QSAR model. To confirm the

robustness of this new model, the randomized model's squared correlation coefficient (R^2) should be lower than the non-randomized model's squares correlation coefficient (R^2) (Hadni & Elhallaoui, 2019). Furthermore, the mismatch between the correlation values of the non-randomized and randomized models must be examined and highlighted.

2.4.2. Cross-validation (LOOCV)

Cross-validation with the leave-one-out approach was used to assess the prediction power of the QSAR model. This is an important stage since a high correlation value indicates a better fit of the data. According to Golbraikh and Tropsha's research (Golbraikh & Tropsha, 2002), cross-validation is required but inadequate to prove the prediction capabilities of the suggested QSAR model. External validation with a test set is required instead to guarantee the model's predictive capability. The external validation should precisely match the requirements of Golbraikh and Tropsha.

2.5. Docking molecular

2.5.1. Docking molecular (reversible) of Tofacitinib

Prior to conducting molecular docking, we subjected the ligands intended for docking to optimization using Avogadro software 2.0. Subsequently, we obtained the JAK3 structure from the RCSB database (PDB ID: 4Z16). The crystal complex of 4Z16 contains co-crystallized ligand 4LH and water molecules. To prepare the protein, we eliminated the water molecules and the co-crystallized ligand from the protein structure. Additionally, we introduced polar hydrogens to the JAK3 protein structure using Discovery Studio software 2021. The active site of 4Z16 was defined by the inclusion of the co-crystallized ligand (4LH) within a specified sphere. Once the ligand and protein were appropriately prepared, we conducted molecular docking utilizing AD4 and AutoVina methods. The three-dimensional grid was created using the AUTOGRID method, which calculates ligand binding energy with their receptor (Moriwaki et al., 2018). The initial grid size used in the docking process was set to $x = 60$, $y = 60$, and $z = 60$, with a grid point spacing of 0.375 Å. The active site of the receptor, located at coordinates ($x = -6.68875$ Å, $y = -14.7757$ Å, and $z = 1.89597$ Å), served as the center of the grid. The docking results obtained from the AD4 and Vina algorithms were visualized using Discovery Studio software 2021 (Systèmes, 2020; Trott & Olson, 2010). CovDock, the guide for the procedure, provides detailed instructions on how to carry out the CovDock approach, specifically utilizing AD4 with flexible side chains.

2.5.2. CovDock (irrevesible) of new compound predicted

The covalent docking approach is utilized in computational methods to predict the binding affinity and mode of ligands that can form covalent bonds with reactive residues in the target protein. Schrodinger's software suite is commonly employed for covalent docking (Schrödinger Release, 2021-1, 2021).

Ligands undergo preparation using LigPrep before covalent docking, which involves generating a three-dimensional structure, assigning bond orders and charges accurately, and limiting the ligand size to a maximum of 500 atoms. The ligand is neutralized, and energy calculations are performed using the OPLS_2005 force field. Stereoisomers are generated, resulting in up to 32 stereoisomers per ligand. In the protein preparation workflow, the user specifies the protein and small molecules for processing. The ligand to be docked is identified, and the presence of metal ions and other molecules in the binding site is determined.

A reactive residue in the protein, such as Cys909, is selected, and a reaction type favoring covalent bond formation is chosen. Covalent docking involves docking the ligand to the protein using a search algorithm. The docking box is positioned based on the centroid of the workspace ligand and co-crystallized ligand, with its size adjusted to accommodate ligands of similar size. A docking cutoff of 2.5 kcal/mol is set to retain poses for subsequent refinement, allowing a maximum of 200 poses for optimization. After docking, MM-GBSA scoring is performed to calculate the binding free energy of the ligand-protein complex. The output includes ligand poses bound to the reactive site, with a maximum of 1000 top-scoring ligands reported. Covalent docking, utilizing Schrodinger's tools such as LigPrep and MM-GBSA scoring, is a powerful computational approach for predicting the binding mode and affinity of ligands that can form covalent bonds with reactive residues in the target protein.

The choice of the JAK3 4Z16 structure as the main point for investigating covalent inhibitors of JAK3 was motivated by several factors, emphasizing its advantages over other protein structures within the scope of this work. The 4Z16 structure, in particular, has a commendably high resolution, allowing for a detailed investigation of JAK3 molecular interactions and active sites (Chen, Yin, et al., 2022). This element is critical in understanding the processes driving JAK3 covalent inhibition and developing tailored therapeutics. Furthermore, the 4Z16 structure's distinguishing features, such as the inclusion of a cysteine residue (Cys909), provide exciting potential for the development of covalent JAK3 inhibitors (Gholamhoseinnia & Asadollahi-Baboli, 2022; Su et al., 2022; Wang et al., 2023; Zhong & Almahmoud, 2023, 2023; Zhu et al., 2020). Furthermore, the 4Z16 structure is preferred due to its ease of use and the availability of additional data. Researchers may have previously worked with this structure in unrelated studies or have access to relevant information that substantiates its application in this specific research endeavor, particularly in the development of innovative compounds for treating conditions such as rheumatoid arthritis, immune diseases, inflammation, and haematopoiesis (Bank, n.d.; Tan et al., 2015).

2.6. Pharmacophore hypothesis

Pharmacophore hypothesis generation plays a crucial role in the field of drug discovery and design, as it facilitates the identification of critical chemical attributes necessary for molecular binding to a particular biological target. The

HypoGen algorithm, which is extensively utilized in Maestro software, is a commonly employed method for generating pharmacophore hypotheses (Suma et al., 2020). This approach involves feeding a collection of diverse compounds that are known to bind to the target of interest into Maestro. The software then analyzes these compounds to generate a pharmacophore model that encompasses shared chemical features among all the compounds. This model serves as a valuable guide for designing new compounds with predicted high affinity and specificity for the target. The HypoGen algorithm, implemented within Maestro, is a powerful tool in the field of drug discovery and design. It enables the rapid generation of pharmacophore hypotheses that can inform the synthesis of novel compounds (Dhiman et al., 2023).

These pharmacophore models can then be employed to design more effective and selective compounds, ultimately resulting in the discovery of novel drugs to treat serious diseases. This approach enables researchers to optimize their time and resources during the drug development process while enhancing the chances of identifying potent and selective compounds (Zhu et al., 2023).

2.7. ADMET predictions

The pharmaceutical industry has experienced significant benefits from the application of molecular modeling techniques, which have played a crucial role in the discovery of new drug candidates. These techniques have reduced the heavy reliance on experimental testing and have improved overall success rates. Using *In* silico studies, it becomes possible to explore and evaluate the pharmacokinetic parameters related to ADMET. This allows researchers to gain valuable insights into the behaviour and properties of potential drug compounds, facilitating the identification of promising candidates for further development (Chen, Yin, et al., 2022; Ferreira & Andricopulo, 2019). These parameters encompass a range of factors, such as drug absorption in the gastrointestinal tract, drug penetration through the blood-brain barrier and into the central nervous system, drug metabolism through chemical biotransformation in the body, drug elimination through excretion pathways, and the evaluation of drug toxicity levels.

To evaluate the potential of the designed compounds as viable drugs, the ADMET pharmacokinetic parameters are assessed using the online tool pkCSM. This tool provides valuable insights into the pharmacokinetic properties of the compounds, aiding in the prediction and optimization of their absorption, distribution, metabolism, excretion, and toxicity characteristics. (Pires et al., 2015), alongside predictions concerning the challenges posed by synthetic accessibility (Villoutreix & Taboureau, 2015).

2.8. Molecular dynamics (MD)

The GROMACS simulation tool was used to run MD simulations (Kumari et al., 2014). To generate the GROMACS input files, the CHARMM-GUI web server (*CHARMM-GUI*, n.d.) was used to produce the GROMACS input files, and the CHARMM36 force field was accessed *via* CHARMM-GUI. The TIP3P water model was used to solve the system within a cubic box of $85 \times 85 \times 85$ angstroms. The system was neutralized by adding $Na+$ and Cl^- ions at a concentration of 0.154 M, and the ion locations were calculated using the Monte Carlo technique (Jo et al., 2008). The system was exposed to a total of 52 $Na+$ ions and 50 Cl^- ions. With a total of 5,000,000 steps, the energy reduction procedure was carried out utilizing the steepest descent approach. The system then reached equilibrium at 310 K for 10 ns in an ensemble with constant atom number, volume, and temperature (NVT). Following that, unconstrained MD simulations were run for 500 ns in an ensemble with a constant number of atoms, pressure, and temperature (NPT), at 310 K reference temperature and 1 ATM pressure (*CHARMM-GUI*, n.d.). MD trajectory analysis was performed using the Visual Molecular Dynamics (VMD) software to examine the system's stability e (Humphrey et al., 1996). Key parameters such as protein solvent accessible surface area (SASA), the radius of gyration (RoG), root mean square deviation (RMSD), and root mean square fluctuation (RMSF) were computed (Boyenle et al., 2022).

In the comprehensive analysis of protein structures, crucial parameters such as protein solvent accessible surface area (SASA), the radius of gyration (RoG), root mean square deviation (RMSD), and root mean square fluctuation (RMSF) were computed. Furthermore, the analysis was enriched by incorporating define secondary structure of proteins (DSSP) to discern secondary structural elements, free energy landscape (FEL) exploration to unveil the energy landscape of the protein's conformational space, and principal component analysis (PCA) to gain insights into the underlying structure and dynamics of the protein (Faris, Ibrahim, Al kamaly, et al., 2023).

2.9. Binding free energy

The analysis of binding affinity between receptors and small ligands can be determined by examining the binding free energy. In this study, the binding free energy was calculated using the molecular mechanics/generalized Born surface area (MM/GBSA) method with the Gromacs (Kumari et al., 2014; Valdés-Tresanco et al., 2021). The calculation Equations (1)–(7) used in this study are presented below [28]: $\Delta_{VDWAALS}$, ΔE_{EL}, ΔE_{GB}, ΔE_{SURF}, ΔG_{GAS}, ΔG_{SOLV} and Δ_{TOTAL}.

$\Delta_{VDWAALS}$: This term represents the van der Waals interaction energy between the protein-ligand complex and its surroundings (Aqvist, 1990).

$$\Delta_{VDWAALS} = \sum{}^i \sum{}^j 6\varepsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right], \quad (5)$$

where ε is the energy scaling factor, σ_{ij} is the distance at which the potential energy of the interaction between atoms i and j is zero, r_{ij} is the distance between atoms i and j, and the summations are over all pairs of atoms i and j. The equation is based on the Lennard–Jones potential.

ΔE_{EL}: This term represents the electrostatic interaction energy between the protein-ligand complex and its surroundings (Calculations were performed using Gaussian, 2009; En-Nahli et al., 2022; Frisch et al., 2009).

$$\Delta E_{EL} = \sum{}^{i}\sum{}^{j} \frac{q_i * q_j}{\varepsilon r}, \qquad (6)$$

where q_i and q_j are the partial charges on atoms i and j, r is the distance between atoms i and j, ε is the dielectric constant of the solvent, and the summations are over all pairs of atoms i and j. The equation is based on Coulomb's law.

ΔE_{GB}: This term represents the energy change associated with the molecule's transfer from a vacuum to a solution (Kim et al., 2013).

$$\Delta E_{GB} = \gamma \sum{}^{i} \frac{q_i^2}{r_i} + k \sum{}^{i}\sum{}^{j} \frac{q_i * q_j}{r_{ij}} + \sum{}^{i} \sigma_i, \qquad (7)$$

where γ and κ are constants that depend on the solvent dielectric constant and ionic strength, q_i is the partial charge on atom i, r_i is the distance from atom i to the center of the solvent-accessible surface, r_{ij} is the distance between atoms i and j, and σ_i is a surface tension term that penalizes the creation of a solvent-accessible surface. The equation is based on the Generalized Born model.

ΔE_{SURF}: This term refers to the energy change associated with the protein-ligand complex's surface area (Dong et al., 2013; Izadyar et al., 2015; Jaffar, 2015; Palacios-Prado et al., 2022).

$$\Delta E_{SURF} = \gamma \sum{}^{i} \frac{1}{r_i}, \qquad (8)$$

where γ is a constant that depends on the solvent dielectric constant and ionic strength, and r_i is the distance from atom i to the centre of the solvent-accessible surface. The equation is based on the solvent-accessible surface area (SASA) model.

ΔG_{GAS}: This term represents the Gibbs free energy change associated with the gas phase (Atkins et al., 2014).

$$\Delta G_{GAS} = H - TS, \qquad (9)$$

where H is the enthalpy, T is the temperature, and S is the entropy. The equation is derived from the Gibbs–Helmholtz equation.

ΔG_{SOLV}: This term represents the Gibbs free energy change associated with the solvation of the protein-ligand complex (Friesner et al., 2006).

$$\Delta G_{SOLV} = \Delta H_{SOLV} - T\Delta S_{SOLV}, \qquad (10)$$

where ΔH_{SOLV} is the enthalpy change associated with the solvation process, ΔS_{SOLV} is the entropy change associated with the solvation process, and T is the temperature. The equation is based on the thermodynamic definition of Gibbs free energy.

Δ_{TOTAL}: This term represents the total energy change associated with the interaction of the protein-ligand complex with its environment (Abel et al., 2017; Friesner et al., 2006).

$$\Delta_{TOTAL} = \Delta_{VDWAALS} + \Delta E_{EL} + \Delta E_{GB} + \Delta_{ESURF} + \Delta G_{SOLV}, \qquad (11)$$

Each term is determined using the corresponding equation as explained previously. The total energy change shows the protein-ligand complex's overall stability or instability in its environment.

2.10. Retrosynthesis

Retrosynthesis, a potent organic chemistry method, is used to build efficient synthetic pathways for complicated compounds. Synthetic Pathway Assembler (SPAYA), a recently created program, uses artificial intelligence and machine learning techniques to automate the retrosynthesis process. The retrosynthesis approach entails rapidly generating synthetic routes for target molecules by analyzing a large database of known reactions and suggesting paths that meet established criteria (Spaya.Ai, n.d.). Time and resources can be saved by using SPAYA during the synthesis of complicated compounds. Chemists may utilize the program to quickly discover prospective starting materials and reaction pathways, allowing them to focus their efforts on the most promising synthetic routes. Furthermore, SPAYA assists in recognizing possible difficulties in synthesis plans, such as the creation of undesirable by-products, allowing for process modification and improvement.

2.11. Molecular electrostatic potential (MEP)

Using visualizations of the MEP (Daoui et al., 2023), we identified reactive sites favourable for electrophilic and nucleophilic actions on the surface of the model compound's molecular structure, the most active molecule (6d). The analysis of the 3D contour map of the MEP simulation results was performed using GaussView 5.0 software. Prior to the

Table 2. Chemical properties of compounds.

Compound	A_Ar	B_Dou	P.V_P	Q.V_N	pIC50 (Exp)
1	18	2	12.95	184.657	6.14
2	18	2	12.95	184.31	6.23
3	17	2	12.95	149.549	8.91
4	17	2	12.95	149.549	7.62
5	17	2	23.274	157.317	8.14
6	17	2	23.274	157.317	8.03
7	17	2	12.95	152.053	8.15
8	17	2	12.95	152.053	8.16
9	18	2	12.95	174.905	9.62
10	18	2	12.95	174.905	7.08
11	18	2	12.95	172.402	8.68
12	18	2	12.95	172.402	8.1
13	18	2	12.95	181.103	6.05
14	18	1	12.95	165.294	6.67
15	18	1	12.95	155.588	8.46
16	18	0	0	142.021	5.52
17	18	2	12.95	181.103	5.99
18	18	1	12.95	165.294	5.54
19	18	1	12.95	155.588	7.81
20	18	0	0	142.021	5.52
21	18	2	12.95	183.607	6
22	18	1	12.95	167.798	5.56
23	18	1	12.95	158.092	7.32
24	18	0	0	144.525	5.52
25	18	1	21.995	176.075	5.52
26	18	1	21.995	176.075	5.52
27	17	1	12.95	135.239	8.14
28	17	0	0	121.672	5.52
29	17	2	12.95	158.25	5.46
30	17	2	12.95	158.25	5.41

MEP analysis, the molecule was optimized using DFT BLYP3 6-31 G method.

3. Results

3.1. QSAR model and statical analysis

After numerous rounds of iteration, a final model consisting of four descriptors, namely (Table 2), was produced: Aro (A_Ar), Double (B_Dou), PEOE VSA PPOS (P.V_P), and Q VSA NEG (Q.V_N). The aromatic ring presence in a molecule is quantified by the descriptor A_Ar. The description B_Dou denotes the number of double bonds in a molecule. Using the PEOE approach, P.V_P describes a molecule's positive partial surface area. Q.V_N also uses the VSA technique to calculate a molecule's negative partial surface area.

3.1.1. Multiple linear regression (MLR)

The MLR models, as shown in Table 3, exhibit good prediction quality for the training set and acceptable performance for the test set. The training set has a high adjusted coefficient of determination R^2, indicating that the model is generalizable and avoids overfitting. The mean squared error (MSE) for the training set is 0.31, indicating that the model's predictions differ by an average of 0.55 units from the real value of the dependent variable (Figure 5). The RMSE, determined as the square root of the MSE, is 0.31, suggesting good accuracy for

the training set. Based on the findings of the multiple linear regression model, it can be concluded that the model is well-fitted and has high prediction quality for the training set as well as acceptable prediction quality for the test set (Table 3).

$$pIC_{50} = -51.65 + 4.86 * A\_Ar + 3.43 * B\_Dou + 0.15 * P.V\_P - 0.21 * Q.V.M, \quad (12)$$

Obtained from Equation (13), it is clear that a linear regression model was used to predict the pIC_{50} value, using four independent variables, namely A_Ar, B_Dou, P.V_P, and Q.VN. According to the regression coefficients associated with these independent variables, the presence of aromatic groups and double bonds, indicated by A_Ar and B_Dou, respectively, favorably increases the pIC50 value, signifying increased inhibitory action. Greater values of P.V_P and Q.V_N, on the other hand, have a negative influence on the pIC50 value, suggesting that compounds with higher values of these variables have lesser inhibitory action. As a result of Equation (1), it is hypothesized that compounds with greater P.V_P and Q.V_N values are more likely to have lower inhibitory activity. Molecule 27 has the following formulas: A_Ar = 17, B_Dou = 1, P.V_p = 12.95, Q.V_N = 135.239, and pIC_{50} = 8.14. A_Ar = 17, B_Dou = 2, P.V_p = 23.274, Q.V_N = 157.317, pIC_{50} = 8.14 for Molecule 5. Molecule 6: A_Ar = 17, B_Dou = 2, P.V_p = 23.274, Q.V_N = 157.317, pIC_{50} = 8.03, and for certain

Table 3. MLR statistical coefficients.

MLR statistical	Observations	Sum of weights	DF	R^2	Adjusted R^2	MSE	RMSE
Training	21	21	16	0.85	0.81	0.31	0.56
Test	9	9	3	0.82	0.748	0.99	0.99
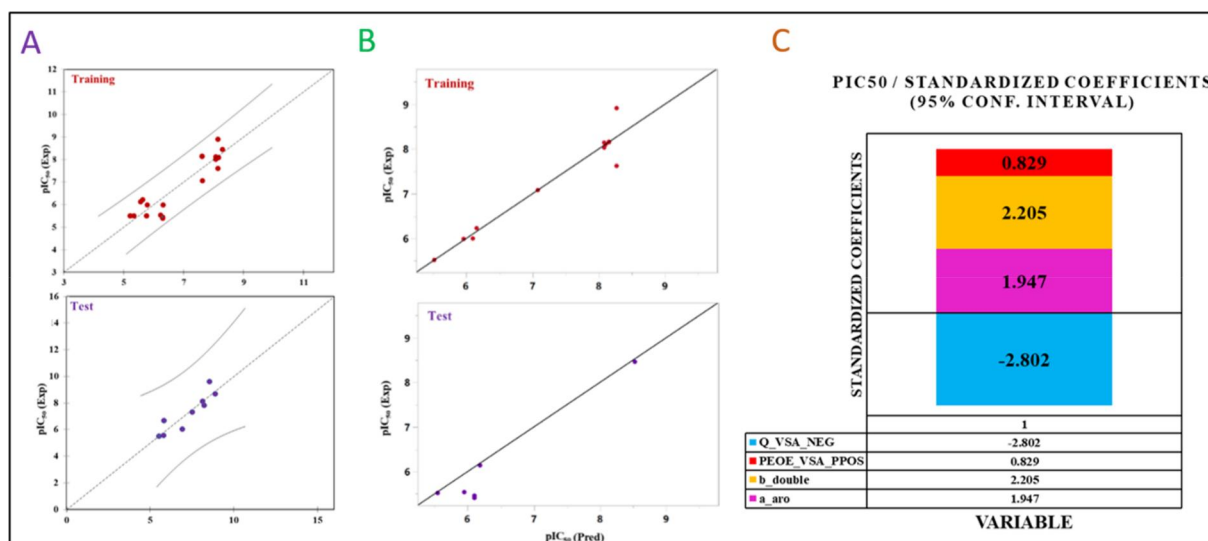
Figure 5. (A) Diagram of linear regression (MLR) in grey for the training set and violet for the test set. (B) Diagram of linear regression (ANN) in red for the training set and violet for the test set. (c) Standardized coefficients diagram for MLR.

Table 4. MLR standardized coefficients (pIC_{50}).

Source	Value	Standard error	t-value	Pr > |t|	Lower bound (95%)	Upper bound (95%)
A_Ar	1.947	0.315	6.191	<0.0001	1.280	2.614
B_Dou	2.205	0.302	7.298	<0.0001	1.565	2.846
P.V_P	0.829	0.170	4.890	0	0.470	1.188
Q.V_N	−2.802	0.378	−7.403	<0.0001	−3.604	−1.999

molecules that may undergo an increase in activity: A_Ar = 17, B_Dou = 2, P.V_p = 23.274, Q.V_N = 157.317, p Molecule 9: A = 18, B = 2, p = 12.95, Q = 152.053, pIC$_{50}$ = 8.16. Molecule 10 has the following formulas: A_Ar = 18, B_Dou = 2, P.V_p = 12.95, Q.V_N = 174.905, and pIC50 = 7.08. A_Ar = 18, B_Dou = 2, P.V_p = 12.95, Q.V_N = 184.31, pIC50 = 6.23 for Molecule 2. Molecule 1: A = 18, B = 2, p = 12.95, Q = 184.657, pIC$_{50}$ = 6.14.

The intensity of each independent variable's influence on the dependent variable is calculated using standardized coefficients that consider the units of measurement and the scale of the variables. These coefficients offer a measure of the relative size of the independent variables' effect on the dependent variable while accounting for any discrepancies in their units or scales. The noteworthy positive standardized coefficients observed in Table 4 and Figure 5 provide compelling evidence that variables A_Ar and B_Dou have a substantial impact on inhibitory activity, underscoring the importance of aromatic groups and double bonds in enhancing the activity. Additionally, the standardized coefficient for P.V_P is also significant, although relatively lower compared to the coefficients for A_Ar and B_Dou. This suggests that the indicator of molecular surface area, P.V_P, also contributes to increasing inhibitory activity, albeit to a lesser degree than aromatic groups and double bonds. The t-value and $p$value are utilized to test the null hypothesis that the true coefficient for the independent variable is equal to zero. A low p-value ($<$0.05) signifies statistical significance, indicating that the coefficient is unlikely to be a result of chance alone. In this instance, all the independent variables have p-value below 0.05, demonstrating their statistical significance as predictors of the dependent variable (Table 4).

Figure 5 presents the standardized coefficient, which reveals a significant negative value for Q.V_N. This suggests that higher values of Q.V_N., an indicator of molecular surface area, are associated with a decrease in inhibitory activity. In summary, the standardized coefficients indicate that the presence of aromatic groups and double bonds in a molecule plays a crucial role in increasing inhibitory activity. Furthermore, higher values of P.V_P have a positive effect, although comparatively less significant. On the other hand, elevated values of Q.V_N tend to result in a decrease in inhibitory activity. These findings provide valuable insights that can assist in the design of new molecules with enhanced inhibitory activity.

3.1.2. Artificial neural network (ANN)

Table 5 presents the outcomes of ANN analysis performed using JMPpro (Ramsey, 2022), incorporating four descriptors: A_Ar, B_Dou, P.V_P, and Q.V_N. These descriptors encompass various molecular structural characteristics, such as the presence of aromatic rings or double bonds, as well as indicators of biological activity, such as molecular polarity. The ANN model was constructed with two hidden layers to capture non-linear relationships between the descriptors and the output, which represents the pIC$_{50}$ value—a quantitative measure of pharmacological activity. By utilizing molecular descriptors, the ANN analysis enables the prediction of pIC$_{50}$ values for novel compounds. This approach proves highly advantageous in rapidly and efficiently predicting the biological activity of a significant number of compounds, thus assisting in the design of innovative pharmaceuticals.

The results obtained from the analysis demonstrate the strong predictive performance of ANN model for both the training and test datasets, as depicted in Table 5 and Figure 5. The coefficient of determination (R^2) values of 0.95 for the training set and 0.85 for the test set indicate that the model effectively accounts for a significant portion of the variation in the dependent variable within the training set and to a satisfactory extent within the test set. Examining the average error of the model's predictions against the actual values, the Root Average Squared Error (RASE) is calculated as 0.25 for the training set and 0.39 for the test set, indicating relatively precise predictions. Similarly, the mean absolute deviation (MAD) values are 0.11 for the training set and 0.27 for the test set, suggesting reasonably accurate model predictions. Additionally, the sum of squared errors (SSE) is 0.85 for the training set and 1.067905 for the test set, indicating favorable prediction quality for both datasets. Finally, the Sum Frequency (Sum Freq) helps determine the average absolute errors and further supports the notion that the ANN model exhibits good predictive accuracy for both sets. The use of the artificial neural network in predicting the dependent variable yields promising outcomes, underscoring its capacity to achieve high predictive performance for both the training and test datasets.

3.1.3. Leave-one-out cross-validation (LOOCV)

The achieved Q^2 result, which stands at 0.66, demonstrates a noteworthy level of predictive performance. This finding suggests that the model exhibits a reasonable degree of accuracy in predicting the values of the dependent variable for new data. The leave-one-out cross-validation method was employed in this study, wherein each data point is sequentially removed from the training set. This approach is recognized for its robustness as it utilizes the complete dataset to evaluate the predictive capability of the model (Faris, Hadni, Saleh, et al., 2023; Hadni & Elhallaoui, 2020).

3.1.4. External validation (Golbraikh & Tropsha)

The Golbraikh and Tropsha test results are utilized to evaluate the predictive accuracy of a regression model and determine its reliability in predicting new data (refer to Table 6) (Faris, Hadni, Ibrahim, et al., 2023; Faris, Hadni, Saleh, et al., 2023).

The outcomes presented in Table 6 exhibit that the model has successfully passed all Golbraikh tests, indicating its trustworthiness in predicting new data. With a Q^2 value of

Table 5. Statistics on ANN coefficients.

Training	Value	Test	Value
R^2	0.95	R^2	0.85
RASE	0.25	RASE	0.39
Mean Abs Dev	0.11	Mean Abs Dev	0.27
SSE	0.85	SSE	1.07
Sum Freq	14	Sum Freq	7

Table 6. Golbraikh and Tropsha test.

Parameter	Formula	Threshold value	Model score
Q^2	$Q^2 = 1 - \dfrac{\sum (Ypred(test) - Y(test))^2}{\sum (Y(test) - Y_{tr})^2}$	>0.5	0.65
R^2	The coefficient of determination for the graph of expected versus seen for the test set	>0.6	0.89
$\left| R_0^2 - R_0'^2 \right|$		<0.3	0.012
r_0^2	r^2 at zero intercept		
$R_0'^2$	r_0^2 for the plot of observed against projected activity for the test set at zero intercept		
$\frac{R^2 - R_0^2}{R^2}$		<0.1	0.01412
$\frac{R^2 - R_0'^2}{R^2}$		<0.1	0.00003
k	The slope of the plot of observed versus projected activity at the intercept	0.85 < k < 1.15	1.014
k′		0.85 < k′ < 1.15	0.98

0.65, surpassing the threshold of 0.5, the model demonstrates a reasonably effective predictive ability for the dependent variable in new data. Moreover, the R^2 value of 0.89 exceeds the threshold of 0.6, indicating that the model can explain a significant portion of the variability in the dependent variable. The difference between the coefficient of determination R^2 and the adjusted coefficient of determination R'^2, represented as $|R\_0^2 - 〚R'〛\_0^2|$, is 0.01243, which falls below the threshold of 0.3. This suggests that the model avoids overfitting and can generalize appropriately. The Golbraikh coefficients K and K' are 1.014 and 0.98, respectively, which fall within the acceptable range of 0.85–1.15, further confirming the model's reliability in predicting new data. The percentage change in R^2 from R'^2, calculated as $[(R^2 - R'^2)/R^2]$, is 1.014, indicating effective generalization by the model. Similarly, the percentage change in R^2 from $R'0^2$, computed as $[(R^2 - R'0^2)/R^2]$, is 0.00003, highlighting the model's correct generalization.

All results from the Golbraikh test demonstrate successful outcomes, providing support for the reliability of the regression model in predicting new data.

3.1.5. Y-randomization

The findings presented in Table 7, derived from the analysis of dependent variable randomization, are utilized to evaluate the significance of the regression model and determine if the observed associations between variables are statistically meaningful or merely a result of chance correlation.

The original model demonstrates a strong correlation coefficient (R) of 0.92 and a coefficient of determination (R^2) of 0.89, suggesting a robust relationship between the independent variables and the dependent variable. Additionally, the Q^2 coefficient of 0.65 indicates a reasonable predictive capability of the model for new data. The results obtained from randomizing the dependent variable reveal that the R, R^2, and Q^2 values of the random models are notably lower compared to those of the original model. These outcomes derived from the dependent variable randomization analysis provide evidence supporting the validity of the original regression model. The original model accurately predicts the values of the dependent variable for new data, and the strong correlation observed between the independent variables and the dependent variable is not merely due to chance correlation. Lastly, the fact that cRp^2 0.76 > 0.6 assures that the observed $Q^2_{(LOO)}$ is statistically significant and indicates that the model is a good fit for the data.

Table 7. Y-randomization test.

Parameters	R	R^2	$Q^2_{(LOO)}$	cRp^2
Random models	0.39	0.17	−0.4	0.76 > 0.6
Model original	0.94	0.89	0.65	

3.2. Pharmacophore prediction

Based on the findings presented in Table 8 regarding the pharmacophore hypothesis, the initial model chosen is DDRRR_1.

To understand the rationale behind this selection, it is essential to conduct a comprehensive analysis of the distinct scores and compare them across models. Among the different models, the DDRRR_1 model demonstrates the following scores: Survival Score: 5.456, Site Score: 0.860, Vector Score: 0.918, Volume Score: 0.630, Selectivity Score: 1.901, Inactive Score: 2.242, and Adjusted Score: 3.213. Notably, the Survival Score achieves the highest value among all models, indicating that the DDRRR_1 model exhibits superior resilience during a predetermined set of tests. Furthermore, the DDRRR_1 model also achieves a higher adjusted score (3.213) compared to the other models. The adjusted score considers all other scores (site, vector, volume, selectivity, and inactive) to provide a holistic evaluation of the model's overall quality. Although the DDRRR_1 model may not possess the highest scores for each criterion, it strikes a balance across the different scores. For instance, the Site Score (0.860) and Vector Score (0.918) both exhibit significant magnitudes, indicating a favorable alignment between the ligand anchor points and the features of the pharmacophore model. In summary, the DDRRR_1 model is chosen due to its equilibrium between the various scores and its superior overall performance compared to the other models.

3.3. Data screening

Utilizing the pharmacophore model DDRRR_1, which is depicted in Figure 6, a virtual screening was performed using ZINCPharmer. This screening process resulted in the identification of a set of 9 molecules, as presented in Tables 8 and 14. These molecules exhibit similarity to the reference compounds known for their high activity. In this context, the symbol D represents a hydrogen donor (H-donor), while R indicates a three-ring group.

3.4. Strucural design

In the search for more potent ligands, we reach the stage of utilizing QSAR models. Initially, we consider the pharmacophore model (DDRRR) obtained. To enhance our determination, we proceed with the calculation of the molecular electronic surface (MEP), based on the structure of the most active reference molecule 9 (pIC50 = 9.62).

The negative electrostatic potential is represented by the red color, indicating that these areas are conducive to

Table 8. Pharmacophore models.

| Model	Survival.S*	Site.S*	Vector.S*	Volume.S*	Selectivity.S*	Inactive.S*	Adjusted.S*
DDRRR_1	5.456	0.860	0.918	0.630	1.901	2.242	3.213
DDRRR_2	5.435	0.848	0.920	0.606	1.915	2.369	3.066
DDRRR_3	5.434	0.881	0.903	0.601	1.903	2.357	3.077
DDRRR_4	5.397	0.795	0.919	0.637	1.900	2.236	3.161
DDRRR_5	5.385	0.797	0.908	0.643	1.891	2.134	3.252
DDRRR_6	5.380	0.788	0.921	0.617	1.908	2.243	3.138
DDRRR_7	5.379	0.802	0.927	0.598	1.906	2.352	3.028
DDRRR_8	5.367	0.834	0.892	0.592	1.903	2.435	2.931
DDRRR_9	5.353	0.760	0.921	0.610	1.915	2.357	2.996
DDRR_1	5.039	0.963	0.968	0.626	1.336	2.548	2.491
DDRR_2	5.028	0.960	0.970	0.605	1.348	2.531	2.497
DDRR_3	5.025	0.960	0.980	0.591	1.348	2.536	2.489
DDRR_4	5.024	0.942	0.971	0.620	1.345	2.360	2.664
DDRR_5	4.990	0.962	0.934	0.584	1.363	2.364	2.625
DRRR_1	4.980	0.880	0.914	0.634	1.406	2.278	2.702
DRRR_2	4.974	0.864	0.915	0.640	1.409	2.293	2.681
DRRR_3	4.972	0.847	0.922	0.651	1.406	2.337	2.634
DDRR_6	4.971	0.916	0.950	0.601	1.358	2.361	2.610
DDRR_7	4.963	0.938	0.959	0.575	1.344	2.379	2.584
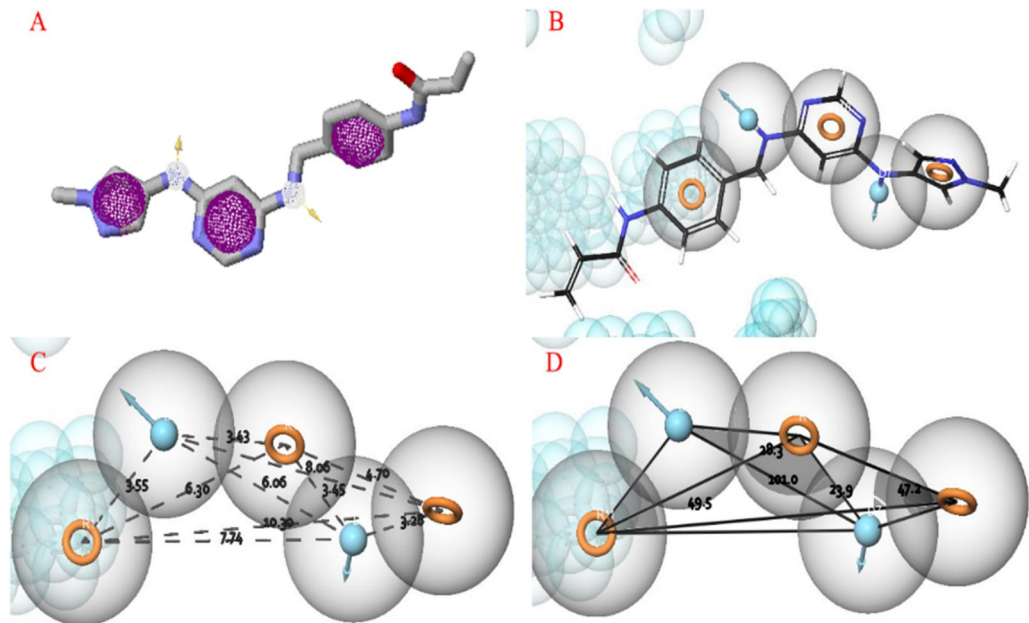
Figure 6. A: Useful features for ZINCPharmer. B: DDRRR pharmacophore model characteristics. C: the distances between the characteristics. D: the angles between the characteristics.

Table 9. Electrostatic potential mapping of the compound 9 (with pIC$_{50}$ = 9.62) after geometric optimization.

E_{LUMO}: 0.093 E_{HOMO}: −0.305 E_{TOTAL}: = −1399

Visualizations 3D of MEP	**Modified Structure**
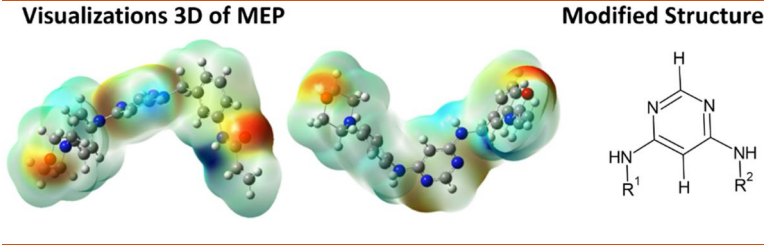
electrophilic reactions. Conversely, the positive electrostatic potential is denoted by the blue color, signifying that these regions are conducive to nucleophilic reactions. The green contours, on the other hand, do not exhibit any potential reactive sites. Based on the results provided by the MEP (Table 9) and the various previous approaches, we were able to encode the structures of diamines with the R1 and R2 groups. This guided us in generating several suggested compounds, from which we selected two compounds, D1 and D2 (Table 14), based on ADMET considerations, specifically their non-toxic and non-hepatotoxic properties, and according to the prediction of biological activity using QSAR models.

3.5. Pharmacophore validation

H-validation is a widely employed technique in molecular modeling to assess the predictive performance of models (Faris, Ibrahim, Al kamaly, et al., 2023). This method involves dividing the dataset into multiple subsets or groups. Each of these subsets is then utilized to train and test the model in a cross-validated manner.

The results presented in Table 10 are derived from the H-validation of the pharmacophore model DDRRR1. Various metrics, including ROC, BEDROC, EF%, AUAC, and FOD, have been employed to evaluate the model's performance. The ROC score of 0.86 indicates that the model exhibits high accuracy in distinguishing between active and inactive compounds. The BEDROC score, with a value of 1.0 for alpha = 160.9, suggests that the model is exceptionally effective in identifying active compounds within the dataset. The EF% values further support the model's efficacy in enriching the dataset with active compounds. For instance, the model

demonstrates an EF of 2.7 for the top 1% of the dataset, indicating a retrieval rate of 2.7 times more actives than expected by chance. The AUAC score of 0.73 indicates that the model's predictions are consistent with the actual dataset. Additionally, the FOD values emphasize the model's capability to filter out false positives, with an FOD of 0.01 for the top 1% of the dataset. Collectively, these results suggest that the pharmacophore model DDRRR1 is a highly effective tool for identifying active compounds within the dataset.

In Figure 7, the receiver operating characteristic (ROC) curve illustrates the performance of a binary classification model, depicting the relationship between the false positive rate (FPR) and the true positive rate (TPR). The false positive rate represents the proportion of incorrectly classified positive instances compared to the total number of true negatives. The ROC curve's positioning closer to the upper left corner indicates a better model performance. In this specific example, the model demonstrates good performance as its ROC curve approaches the upper left corner. This suggests that the model exhibits high sensitivity, meaning it has a strong ability to correctly identify true positives, while maintaining a low false positive rate or specificity.

3.6. ADMET properties

In the context of drug discovery, it is of utmost importance to anticipate the properties associated with the absorption, distribution, metabolism, excretion, and toxicity of novel drugs to prevent their failure during clinical phases (Dearden, 2007; Faris, Ibrahim, Hadni, et al., 2023). Accurate prediction of pharmacokinetic and bioavailability properties stands as a fundamental stage in drug development. The outcomes of these predictions, utilizing the pkCSM predictive model, are presented in Tables 11 and 12. A notable caco-2 permeability value reflects favorable absorption, with predicted values surpassing 50%, while values below 30% imply inadequate intestinal absorbance. A low total clearance value (logCLtot) signifies a prolonged drug half-life. Furthermore, a logBB value below −1 indicates limited distribution to the brain. Additionally, a positive outcome in the AMES test indicates the potential mutagenic activity of the compound. AMES toxicity was observed in all compounds, as indicated by the results presented in Tables 11 and 12. As per the established ADMET guidelines, the evaluation of a molecule's

Table 10. Evaluation metrics for DDRRR_1 hypothesis.

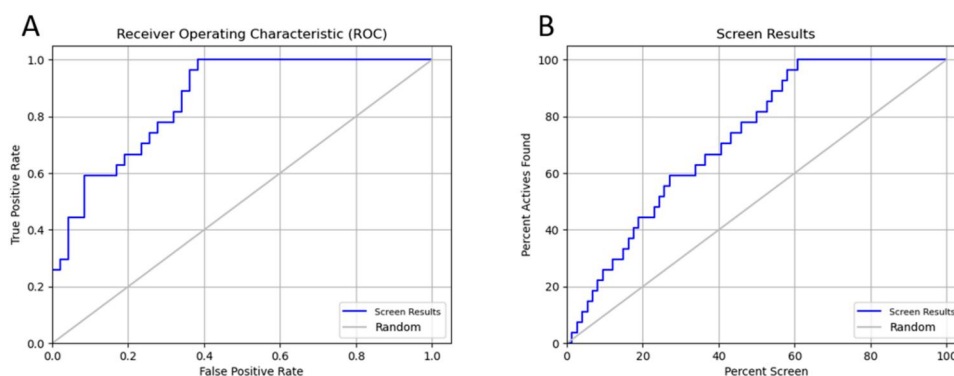| Hypothesis	DDRRR_1
Phase hypo score	1.33
EF1%	2.74
BEDROC160.9	1
ROC	0.86
AUAC	0.73
Ave outranking decoys	6.56
Total actives	27
Ranked actives	27
Matches	4 of 5
Excluded volumes	Yes

Figure 7. ROC analysis and scree results for screening performance assessment.

Table 11. Properties of ADMET for new ligands with a positive AMES test.

ID		ZINC72622261	ZINC14568680	ZINC52306282	ZINC13631917	ZINC52190419	ZINC4810303
Absorption	*Water solubility*	−3.858	−3.816	−3.652	−3.231	−4.447	−3.693
	Caco2 permeability	0.349	−0.788	0.519	−0.161	1.061	0.494
	Intestinal absorption (human)	94.517	91.1	88.268	91.93	89.689	91.113
Permeability	*Skin*	−2.749	−2.741	−2.735	−2.735	−2.906	−2.735
	BBB	−0.135	−0.544	−0.296	−0.063	−0.901	0.033
Metabolism	*CNS*	−1.89	−2.181	−1.9	−1.994	−2.914	−1.922
substrate	*CYP2D6*	No	No	No	No	No	No
	CYP3A4	Yes	Yes	Yes	Yes	Yes	Yes
inhibitor	*CYP1A2*	Yes	Yes	Yes	Yes	Yes	Yes
	CYP2C19	Yes	Yes	Yes	Yes	Yes	Yes
	CYP2C9	Yes	Yes	Yes	No	No	Yes
	CYP2D6	No	No	No	No	No	No
	CYP3A4	Yes	Yes	Yes	Yes	Yes	Yes
Toxicity	*AMES toxicity*	Yes	Yes	Yes	Yes	Yes	Yes
	Skin Sensitisation	No	No	No	No	No	No

Table 12. Properties of ADMET for new ligands with a negative AMES test.

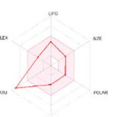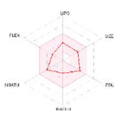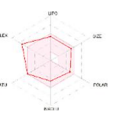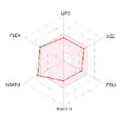| ID		ZINC52190408	ZINC74072092	ZINC78345744	Tofacitinib
Absorption	*Water solubility*	−3.741	−4.273	−4.025	−3.526
	Caco2 permeability	0.919	0.868	0.817	1.36
	Intestinal absorption (human)	94.638	90.579	92.776	93.481
Permeability	*Skin*	−0.293	−0.312	0.538	−3.154
	BBB	−2.257	−1.839	0.538	−0.752
Metabolism	*CNS*	−2.257	−1.839	No	−3.064
Substrate	*CYP2D6*	Yes	Yes	Yes	No
	CYP3A4	Yes	Yes	Yes	No
Inhibitor	*CYP1A2*	Yes	Yes	Yes	Yes
	CYP2C19	No	No	No	No
	CYP2C9	No	No	No	No
	CYP2D6	Yes	Yes	Yes	No
	CYP3A4	Yes	Yes	Yes	No
Toxicity	*AMES toxicity*	No	No	No	No
	Skin Sensitisation	No	No	No	No
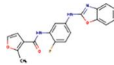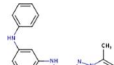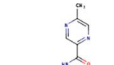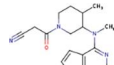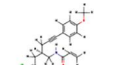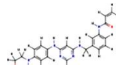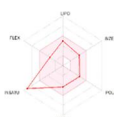
ability to penetrate the blood-brain barrier (BBB) and central nervous system (CNS) permeability is crucial for assessing its potential to elicit effects on the CNS. Acceptable BBB permeability is determined by LogBB values within the range of $-1 < \text{LogBB} < 0.3$, while CNS permeability is considered acceptable for LogPS values falling within $-3 < \text{LogPS} < -2$. Values below −3 or above −2 are considered to be outside the acceptable range.

The results of the compounds ZINC72622261, ZINC14568680, ZINC52306282, ZINC13631917, ZINC52190419, and ZINC4810303 can be interpreted as follows: Absorption: The water solubility values of the compounds range from −3.858 to −4.447, suggesting low solubility in water. Caco-2 permeability, which is used to assess intestinal absorption, varies from −0.788 to 1.061. Positive values indicate better intestinal absorption. Permeability: The values of skin permeability range from −2.735 to −2.906, suggesting relatively low permeability through the skin. The values of blood-brain barrier (BBB) permeability range from −0.544 to 0.033, indicating variable penetration into the central nervous system (CNS). Negative values indicate low BBB penetration. Metabolism: Inhibition of CYP3A4 is important in drug metabolism, as it can lead to an increase in the plasma concentration of drugs and thus to an increase in their efficacy. CYP3A4 inhibition may be beneficial for JAK3 inhibition, as many JAK3 inhibitors are metabolized by CYP3A4. By inhibiting CYP3A4, the drug's half-life can be prolonged and its

bioavailability increased, which may improve its efficacy and reduce the doses needed to achieve sufficient JAK3 inhibition (Faris, Hadni, Ibrahim, et al., 2023; Jain et al., 2019; Li et al., 2017). All compounds are substrates of CYP3A4, meaning they can be metabolized by this hepatic enzyme. The compounds ZINC72622261, ZINC14568680, ZINC52306282, ZINC13631917, ZINC52190419, and ZINC4810303 also inhibit the enzymes CYP1A2, CYP2C19, and CYP3A4. Toxicity: All compounds exhibit positive toxicity according to the AMES test, indicating a potential for toxicity. None of the compounds show skin sensitization according to the skin sensitization test. Moreover, these molecules are eliminated because they represent AMES toxicity.

The findings for the compounds ZINC52190408, ZINC74072092, and ZINC78345744 can be concisely summarized as absorption: The water solubility values of the compounds range from −3.741 to −4.273, indicating low solubility in water. Caco-2 permeability, used to evaluate intestinal absorption, varies from 0.817 to 0.919. Positive values indicate better intestinal absorption. Permeability: The values of skin permeability range from −0.312 to 0.538, suggesting moderate permeability through the skin. The values of the blood-brain barrier (BBB) permeability are not available in the provided data. Metabolism: In this study, CYP3A4 was identified as the primary human enzyme responsible for the metabolism of rheumatoid (Sun et al., 2019; Thatcher et al., 2010) and colon cancer drugs (*Thyroid Hormones*

Table 13. The design and predicted compounds.

ID	ZINC52190408	ZINC74072092	ZINC78345744	Tofacitinib	D1	D2
Compound 2d						
Plot Physicochemical Properties						
MW	351.33	344.37	304.35	312.37	439.94	430.5
Log P	3.71	2.61	2.72	2.09	2.28	1.96
Rotatable bonds	5	5	5	4	11	9
H-bond acceptors	5	4	3	4	4	4
H-bond donors	2	2	2	1	3	3
MR	94.96	99.51	90.76	93.481	124.61	129.74
TPSA	80.3	84.21	66.91	88.91	102.16	91.41
Log S	−5.43	−4.65	−3.96	−2.75	−5.69	−5.01
log Kp (cm/s)	−5.57	−6.12	−6.1	−7.14	−6.25	−6.25

Induce Doxorubicin Chemosensitivity through Enzymes Involved in Chemotherapy Metabolism in Lymphoma T Cells - PMC, n.d.; Zhou et al., 2019). The results revealed that all newly designed compounds appear to function as substrates and inhibitors of CYP3A4. The compounds ZINC52190408, ZINC74072092, and ZINC78345744 are also substrates of CYP1A2 and CYP2D6. However, they are not substrates of CYP2C19 and CYP2C9 enzymes. Toxicity: None of the compounds exhibit toxicity according to the AMES test, suggesting a low potential for toxicity. None of the compounds show skin sensitization according to the skin sensitization test. In comparison to the drug tofacitinib the compounds ZINC52190408, ZINC74072092, and ZINC78345744 demonstrate good intestinal absorption, moderate permeability through the skin, and potential metabolism by the enzymes CYP3A4, CYP1A2, and CYP2D6. These results highlight their potential for advancing drug development. The newly discovered compounds ZINC52190408, ZINC74072092, and ZINC78345744 exhibit promising, they display favorable characteristics ADMET.

3.7. Lipinski's rule

The rules that have been established to assess the physicochemical properties of molecules play a crucial role in determining their potential for biological activity and toxicity. According to these rules, a molecule needs to exhibit an LIPO value within the range of −0.7 to 5 to be classified as lipophilic. Additionally, the molecular weight is evaluated, and it must fall between 150 and 500 g/mol. Remarkably, all the molecules satisfy this criterion. Furthermore, the polarity assessment is based on the determination of the topological polar area (TPSA), which should lie within the range of 20–130 Å².

The provided results, as shown in Table 13, illustrate that all the molecules under examination meet the required molecular weight (MW) criterion, which stipulates a range between 150 and 500 g/mol. Regarding lipophilicity, it is
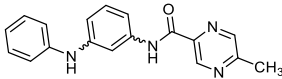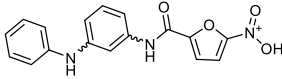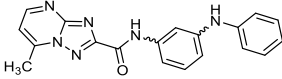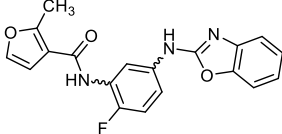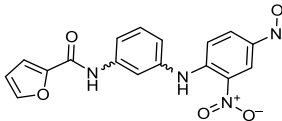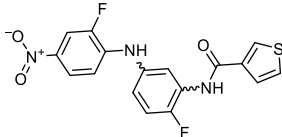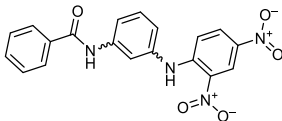
noteworthy that all the molecules fall within the acceptable range of −0.7 to 5 for Log pvalue, indicating their classification as lipophilic compounds by established guidelines for assessing molecular physicochemical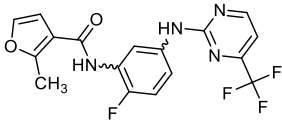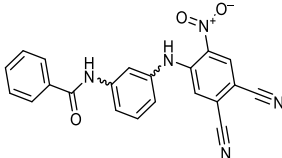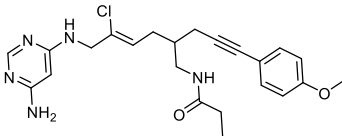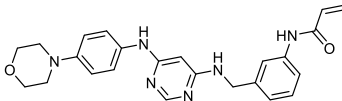 properties. In terms of other characteristics, all the molecules possess five rotatable bonds, except for molecule ZINC35324135, which has four. Furthermore, the number of hydrogen acceptors (H-bond acceptors) among the molecules ranges from two to five, while the number of hydrogen donors (H-bond donors) varies between two and three, with molecules D1 and D2 having 11 and 9 rotatable bonds, respectively. Regarding polarity, molecules ZINC52190408, ZINC74072092, ZINC35324135, D1, and D2 meet the specified condition for topological polar surface area (TPSA), which requires a range of 20–130 Å2. Consequently, in terms of Log S, all four molecules exhibit good solubility in water. The investigated molecules exhibit physicochemical properties, suggesting their potential as drugs for future investigations.

3.8. Prediction of biological activity

Following the acquisition of a pharmacophore model capable of facilitating the prediction of structurally viable compounds with potential biological activity against JAK3, the utilization of ZINCPharmer furnished us with 9 molecules recommended by the DDRRR pharmacophore model. Subsequently, we ensured the model's reliability and robustness by employing five descriptors, coupled with internal and external validation methodologies, which yielded favorable outcomes. Using this model, we utilized it to make forecasts about the biological activity of novel identifiers and create ligands, and the outcomes are displayed in Table 14.

Said table encompasses the descriptors corresponding to each predicted structure, accompanied by their respective predicted biological activity. The biological activity of the novel molecules was predicted utilizing the MLR and ANN-based model derived from Equation (1), which demonstrated significant activity against JAK3 for all compounds.

Table 14. Predicted biological activity pIC50 using QSAR models MLR and ANN.

ID	Smile	pIC$_{50}$ (MLR)	pIC$_{50}$ (ANN)
ZINC78345744		8.57	7.84
ZINC72622261		6.88	7.02
ZINC74072092		8.67	8.45
ZINC52190408		8.99	6.78
ZINC14568680		6.17	7.32
ZINC52306282		7.18	7.60
ZINC13631917		7.69	8.41
ZINC52190419		8.07	7.47
ZINC4810303		6.49	7.60
D1		8.12	8.43
D2		7.68	6.53

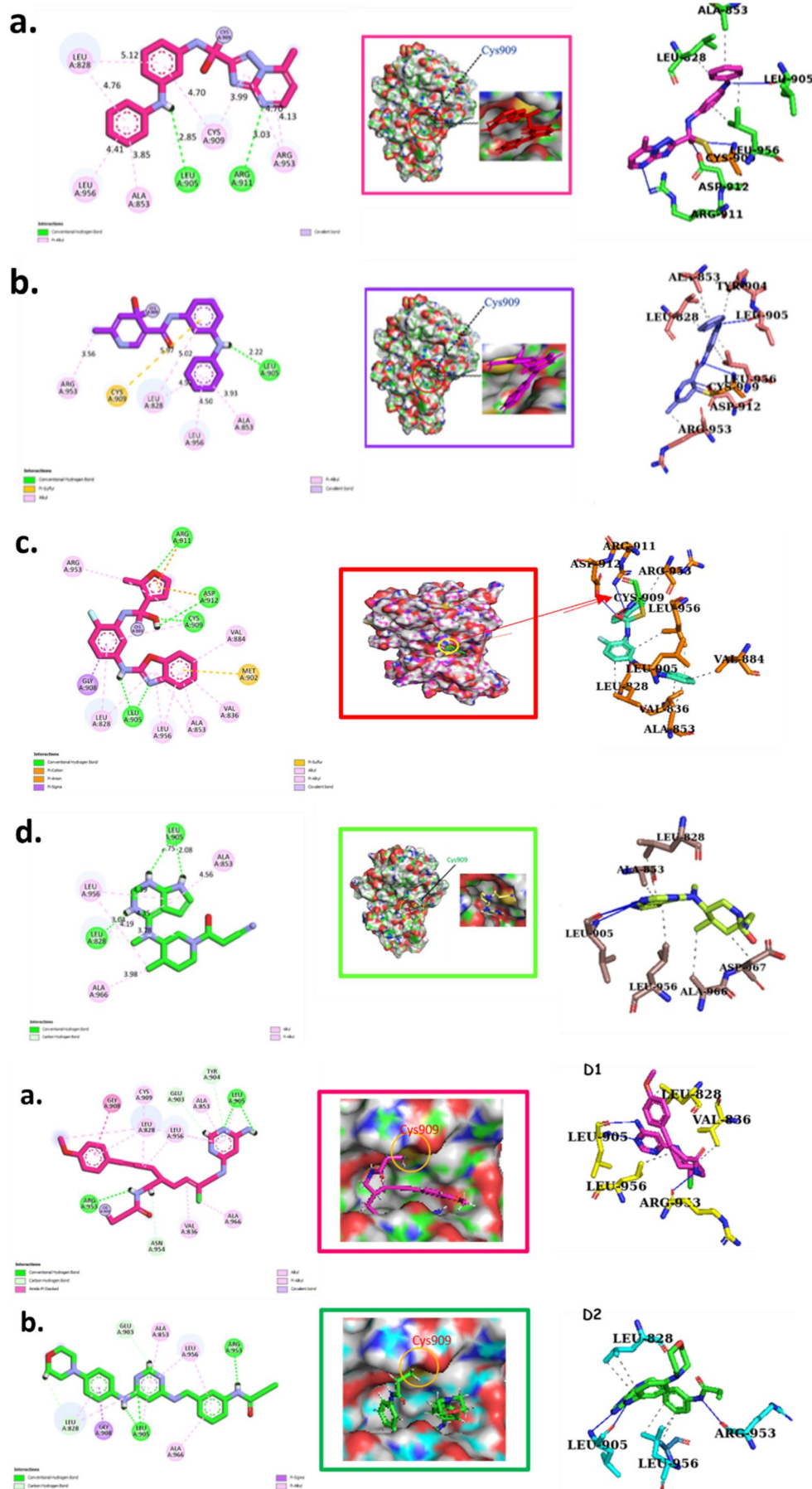## 3.9. Covalent docking docking

3.9.1. Analysis of new compound identifiers

A fundamental method for researching covalent bonds between a target protein and possible ligands is known as covalent docking. It is essential for creating novel pharmacological inhibitors, especially for the JAK3 protein and Cys909 as its active site. JAK3 (4Z16) complex, which exhibits exceptional resolution, may be used to evaluate the effectiveness of covalent docking using the crystallographic structures that are currently available (Bank, n.d.; Tan et al., 2015). This high-resolution data allows for more precise *In-silico* studies of ligand-protein interactions, confirming the relevance of covalent docking in exploring new potential ligands. The protein JAK3 (ID: 4Z16) is

Table 15. Affinity and Non-Covalent interactions of novel compound identifiers (a. ZINC74072092: −6.95 kcal/mol b. ZINC78345744: −5.54 kcal/mol c. ZINC52190408: −9.47 kcal/mol d. Tofacitinib-Drug: −7.5 kcal/mol) and Designs (a1. D1: −9.57 kcal/mol b. D2: −7.45 kcal/mol.) with JAK3.

Novel identifier and design compounds

an exciting discovery in the field of biochemistry. This protein exhibits an intriguing characteristic: it can form a covalent bond with Cys909 when it interacts with the ligand 4LH.

The covalent docking outcomes depicted in Table 15 reveal the formation of distinct bond types between the JAK3 protein and the three ligands. The distances associated with each bond, measured in Angstroms, assume significance in evaluating the strength and stability of the respective interactions. The JAK3-ZINC74072092 complex displayed binding interactions with several key residues, including Leu905, Arg911, Arg953, Cys909, Leu828, and Ala853 (Table 15). The nature and distances of these binding interactions were as follows: a hydrogen bond at 2.85 Å for LEU905 and at 3.03 Å for Arg911, Pi-Alkyl interactions at 4.13 and 4.70 Å for Arg953, at 3.99 and 4.70 Å for Cys909, at 5.12 and 4.76 Å for Leu828, and a Pi-Alkyl interaction at 3.85 Å for Ala853. These findings highlight the specific amino acid residues involved in the binding and provide insights into the molecular interactions between JAK3 and ZINC74072092. JAK3-ZINC78345744 complex exhibited binding interactions with specific amino acid residues, namely Arg953, Leu828, Cys909, Leu956, and Ala853 (Table 15). The nature of these interactions and their corresponding distances are as follows: a hydrogen bond at 2.04 Å for ARG953, hydrogen bonds at 2.91 and 2.93 Å (as well as a carbon hydrogen bond) for Leu828, a hydrogen bond at 2.59 Å and a Pi-Alkyl interaction at 4.70 Å for Cys909, a Pi-Alkyl interaction at 4.52 Å for Leu956, and a Pi-Alkyl interaction at 5.24 Å for Ala853. These findings provide detailed insights into the specific residues involved in the binding process, shedding light on the molecular interactions between JAK3 and ZINC78345744. The JAK3- ZINC52190408 complex exhibited binding interactions with multiple amino acid residues, including Asp912, Leu905, Cys909, Arg911, Gly908, Met902, Val836, Val884, Leu828, and Ala853 (Table 15). The binding distances and types of these interactions are as follows: a hydrogen bond at 2.24 Å for Asp912, hydrogen bonds at 1.92 and 2.36 Å for Leu905, a hydrogen bond at 1.86 Å and an alkyl interaction at 3.37 Å for Cys909, a hydrogen bond at 2.50 Å and a Pi-Cation interaction at 3.64 Å for Arg911, a Pi-Sigma interaction at 2.45 Å for Gly908, a Pi-Sulfur interaction at 5.15 Å for Met902, and a Pi-Alkyl interaction at 4.91 Å for Leu828. Furthermore, additional Pi-Alkyl interactions were observed for Val836 (at 5.02 Å), Ala853 (at 3.64 Å), Val884 (at 5.46 Å), Leu956 (at 4.74 Å), Leu828 (at 4.60 Å), Ala853 (at 4.18 Å), and Leu956 (at 4.30 Å).

3.9.2. CovDock of new compounds design
For D1 (Table 15): Direct hydrogen bonds are observed between the ligand and residues Arg953, and Leu905, with distances ranging from 1.78 to 1.88 Å, indicating strong and specific interactions. Carbon hydrogen bond interactions are present with Asn954, Glu903, and Tyr904 at longer distances, between 2.33 and 2.72 Å. Π-Π interactions of the amide-aromatic stacking or alkyl type are detected with Gly908, Ala966, and various aliphatic residues (Val836, Leu, Cys909) between 3.65 and 5.39 Å. For D2 (Table 15): The residues Arg953 and Leu905 form direct hydrogen bonds with the

ligand again, at distances similar to D1. A carbon-hydrogen bond interaction is observed for Glu903. The presence of longer distance π-alkyl interactions is noted, up to 5.46 Å, notably involving Leu956, Ala966, and Leu828/905.

3.9.3. Molecular docking analysis of Tofacitinib-Drug
Tofacitinib is a small-molecule drug that belongs to the class of Janus kinase (JAK) inhibitors. It is primarily used for the treatment of rheumatoid arthritis, psoriatic arthritis, and ulcerative colitis. Tofacitinib works by inhibiting the activity of JAK enzymes, which play a crucial role in controlling immune cell function and inflammation. By blocking the activity of JAK enzymes, tofacitinib helps to reduce inflammation and alleviate the symptoms of these autoimmune diseases. At distances of 3.04, 2.75, 2.08, and 3.78, tofacitinib established numerous hydrogen bonds with residues Leu828 and Leu906. Leu956, Ala966, Leu828 (4.35), Ala853 (4.56), and Leu956 (4.39) were among the residues with which tofacitinib interacted hydrophobically (Table 15).

Following an extensive examination of the ADMET characteristics of the anticipated and engineered substances, we moved forward by conducting covalent docking of the most highly ranked compounds, namely ZINC52190408, ZINC74072092, ZINC78345744, D1, and D2, in comparison to the drug Tofacitinib. The objective of this docking investigation was to assess the interaction between the JAK3 receptor and the ligands, as well as to evaluate their mutual affinity. Furthermore, MD simulations were carried out to analyze the dynamic behavior and stability of the target protein.

3.10. Molecular dynamics

Molecular dynamics analysis is a vital approach for comprehending the motions and interactions of molecules under physiological or pathological conditions. To interpret the outcomes of this analysis, various parameters, such as Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Solvent Accessible Surface Area (SASA), Radius of Gyration (RoG), and hydrogen bonds (H-Bonds), can be utilized. RMSD is employed to gauge the average deviation between the molecular structures during the simulation, enabling the assessment of structural stability and comparisons at different time points. RMSF quantifies the magnitude of atom fluctuations, shedding light on molecular flexibility. High RMSF values in certain regions may indicate interaction areas with other molecules or crucial regions for the molecule's biological function. SASA measures the accessible surface area of the molecule to the solvent, aiding in the understanding of interactions with the surrounding environment. Notable changes in SASA signify alterations in the molecule's interaction with its environment, potentially impacting its biological function. RoG quantifies the compactness of the molecule and allows monitoring of conformational changes throughout the simulation. Hydrogen bonds play a crucial role in molecular dynamics, and their analysis provides insights into molecule-to-molecule interactions. Integrating these diverse parameters offers a comprehensive

Table 16. RMSD, RMSF, SASA, and RoG ZINC74072092, ZINC78345744, ZINC52190408, Tofacitinib drug, D1 and D2.
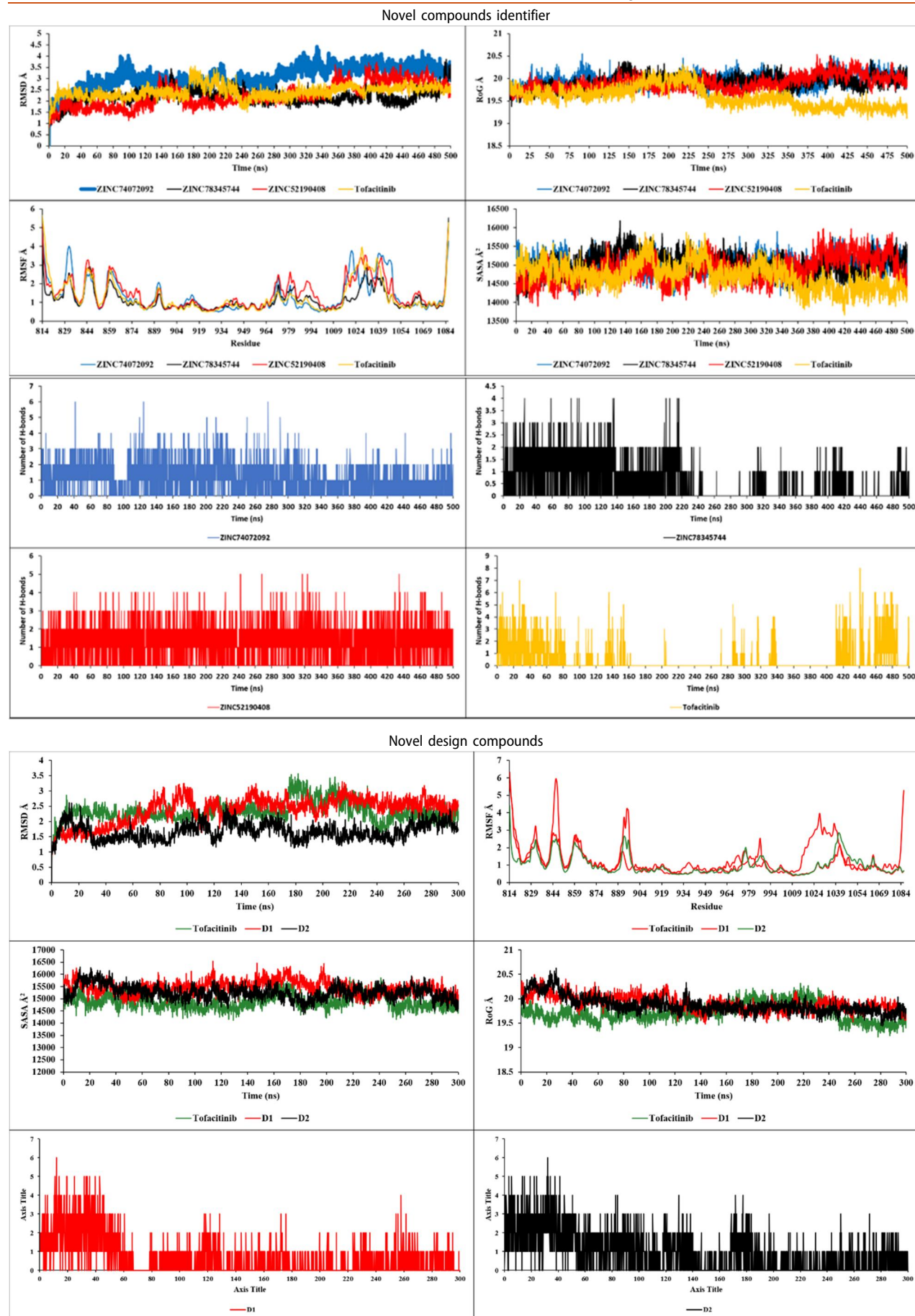
understanding of molecular dynamics and its influence on the molecule's biological function.

In this study, the stability of three compounds, namely ZINC78345744, ZINC52190408, and ZINC74072092, was analyzed during a 500 ns simulation. The remarkable stability of these compounds was assessed based on the results obtained from Table 16, which revealed average RMSD values of 3.05, 2.18, and 2.24 Å, respectively. To evaluate the stability of residues within the complexes, RMSF calculations were performed. Table 16 presents the results, indicating remarkable stability of the residues in the studied ligands, with mean values of 1.36, 1.17, and 1.47 Å for ZINC78345744, ZINC52190408, and ZINC74072092, respectively. Additionally, SASA and RoG calculations were conducted to assess the compactness of the complexes throughout the 500 ns

simulation. As depicted in Table 16 the complexes maintained their compact structures, with mean SASA and RoG values of 14958 Å2 and 20 Å, respectively.

The analysis of hydrogen bond donors and acceptors (H-bonds) in Table 16 for the new ligands ZINC74072092, ZINC78345744, and ZINC52190408 reveals the maximum number of H-bonds formed during the 500 ns simulation to be 6, 4, and 5, respectively, with a minimum of 1 H-bond formed for all three ligands. During the 500 ns simulation, it was observed that ZINC74072092 exhibited stability with the JAK3 protein, forming 1–3 bonds. For ZINC78345744, it was observed to maintain its stability for 250 ns with 1–2 bonds, but no further remarks were made for the remaining simulation, with occasional formation of 1 and 3 bonds. The compound ZINC52190408 demonstrated stabilization with 1–3

Table 17. DSSP analysis for the newly identified, designed compounds and Tofacitinib drug.

bonds throughout the entire 500 ns simulation. The analysis of tofacitinib compared to a new compound showed discrepancies during the 500 simulations, suggesting that the results of the covalent docking analysis indicate a strong affinity between the studied ligands and the JAK3 protein.

The results of MD analysis for the newly designed compounds (Table 16) show an acceptable and stable RMSD compared to tofacitinib. D1 and D2 in interaction with JAK3 exhibited an RMSD ranging from 1.5 to 2.5 Å throughout the trajectory. The RMSF remained stable with a few exceptions for residues greater than 3, which were located outside the active site during the interaction of JAK3 protein with D1 and D2. These residues include a, b, and so on. The SASA and RoG showed stability during the trajectory. SASA ranged between 15,500 and 160,000 Å^2, while RoG ranged between 19.5 and 20.9 Å, indicating favorable compactness of the complexes. The analysis of Hbond demonstrated stable bond formations during the simulation, with fewer disruptions. D1 and D2 in interaction with JAK3 exhibited maximum stability with several hydrogen bonds reaching up to 2, and minimum stability with 1 hydrogen bond. It is worth noting that D1 and D2 maintained stability throughout the trajectory with hydrogen bonds.

3.10.1. DSSP (define secondary structure of proteins)

DSSP algorithm is a commonly used method for assigning the secondary structure of proteins. It identifies alpha helices, beta sheets, kinks, and loops, as well as disulfide bridges and 3/10 helices. The method uses the protein's atomic coordinates to determine the local conformation of each residue.

DSSP analysis assigns a code to each residue based on its secondary structure, enabling quantitative and qualitative analysis of the protein's structure. The results of DSSP analysis can be used to study the stability of protein structure, and protein-protein interactions, as well as for the prediction of protein function and dynamics. In short, DSSP analysis is a powerful tool for characterizing protein structure and studying protein function and dynamics.

DSSP analysis for the predicted new ligands shows significant results for the secondary structure, these results may be useful for understanding the interactions between ligands and the target protein, and for designing new ligands with similar properties. Table 17 shows that ligands ZINC74072092, ZINC78345744, ZINC52190408, D1, and D2 exhibit pseudo transitions similar to those of the known apoprotein, passing in particular through structural elements H (alpha helix), B (isolated strand), E (extended strand in parallel and/or antiparallel beta-sheet conformation), G (3-10 helix), I (pi helix), T (turn of 3 to 4 residues), S (bend) or C (turn). In a parallel manner, all three compounds exhibit alterations akin to those observed in the apoprotein. The DSSP analysis, when contrasted with tofacitinib, implies that the anticipated novel ligands manifest a secondary structural profile closely resembling that of the established apoprotein, characterized by quasi-analogous transitions among distinct structural elements. Additionally, the congruence in secondary structural attributes between the ligands and the apoprotein may imply a commensurate stability for these compounds, a facet of significance in the context of designing ligands that would offer an optimal level of biological activity.
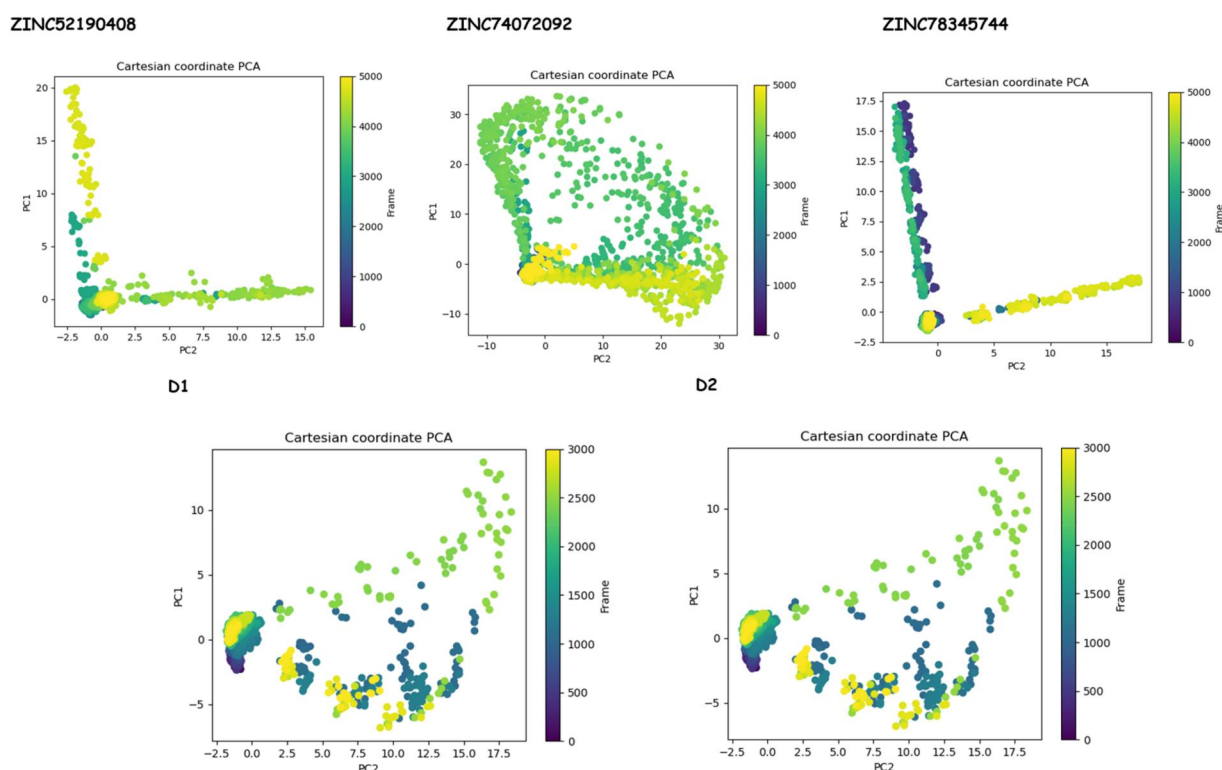

Figure 8. Principal component analysis (PCA) for the newly identified and designed compounds.

3.10.2. Free-energy-landscapes (FEL)

FEL is a valuable tool for interpreting and analyzing biomolecular processes such as molecular folding, aggregation, and recognition. The free energy landscape of a molecule can be computed using the formula:

$$F(CV1, \ CV2) \ = \ -k_b TlnP(CV1, \ CV2), \qquad (13)$$

where kb and T represent the Boltzmann constant and absolute temperature, respectively, and P (CV1, CV2) is the probability distribution of the molecular system along the reaction coordinates or collective variables (CV1 and CV2). A range of reaction coordinates can be selected, including contact distances between two atoms, the root means square deviation (RMSD), the radius of gyration, angles, dihedral angles, principal component analysis (PCA), and others.

The PCA analysis results provided the following (Figure 8): in the complex with ZINC52190406, JAK3 showed cluster motion covering a range on PC1 and PC2 between (0, 20, and −2.5). In the complex with ZINC74072092, JAK3 showed cluster motion covering a range on PC1 and PC2 between (−10, 30 and −10, 30). In the complex with ZINC783445744, JAK3 showed cluster motion covering a range on PC1 and PC2 between (0, 15 and −2.5, 17.5). In the complex with D1, JAK3 showed cluster motion covering a range on PC1 and PC2 between (−25, 10 and −2.5, 17.5). In the complex with D2, JAK3 showed cluster motion covering a range on PC1 and PC2 between (−2.5, 10 and −2.5, 17.5). In the case of a
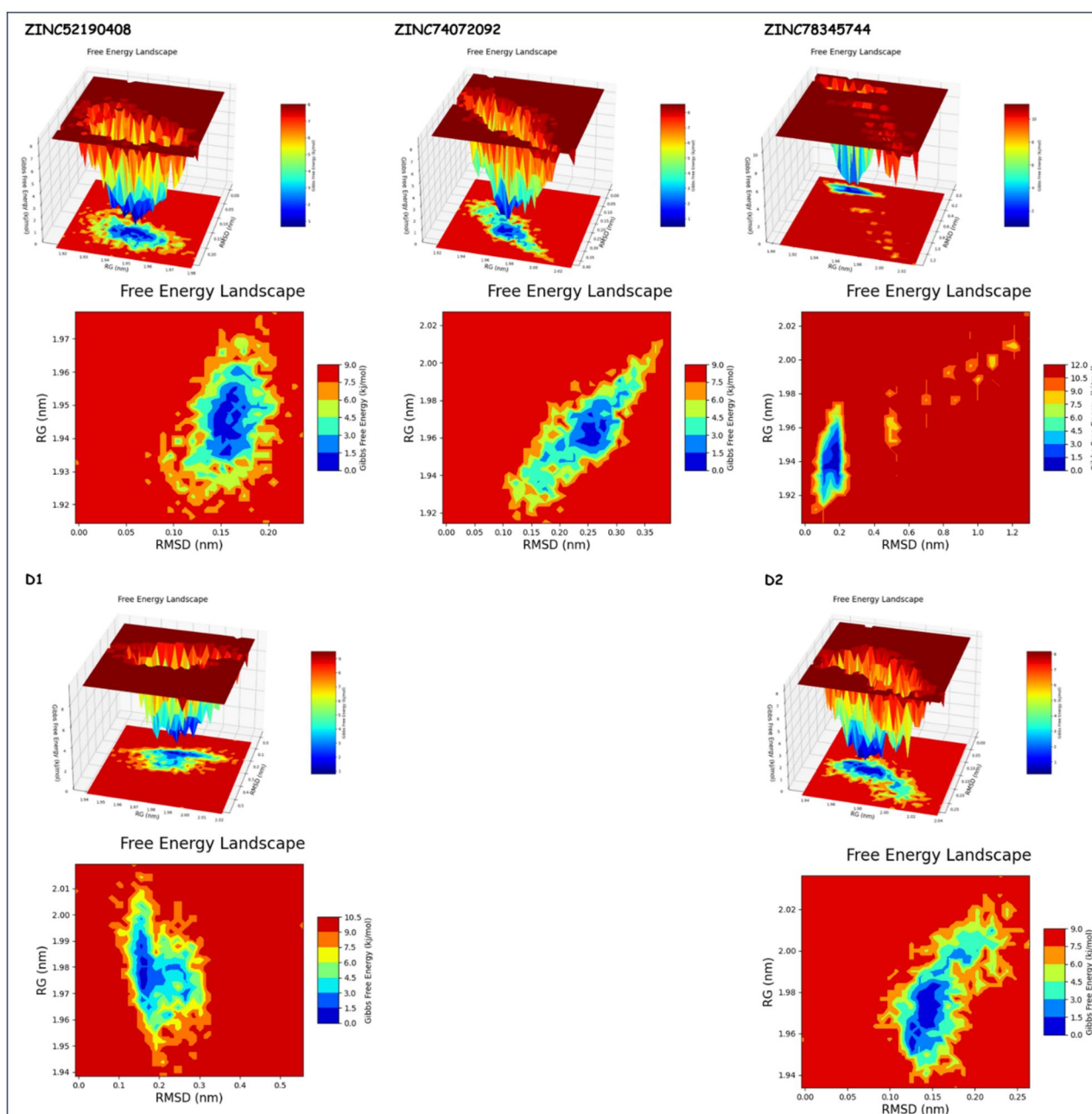

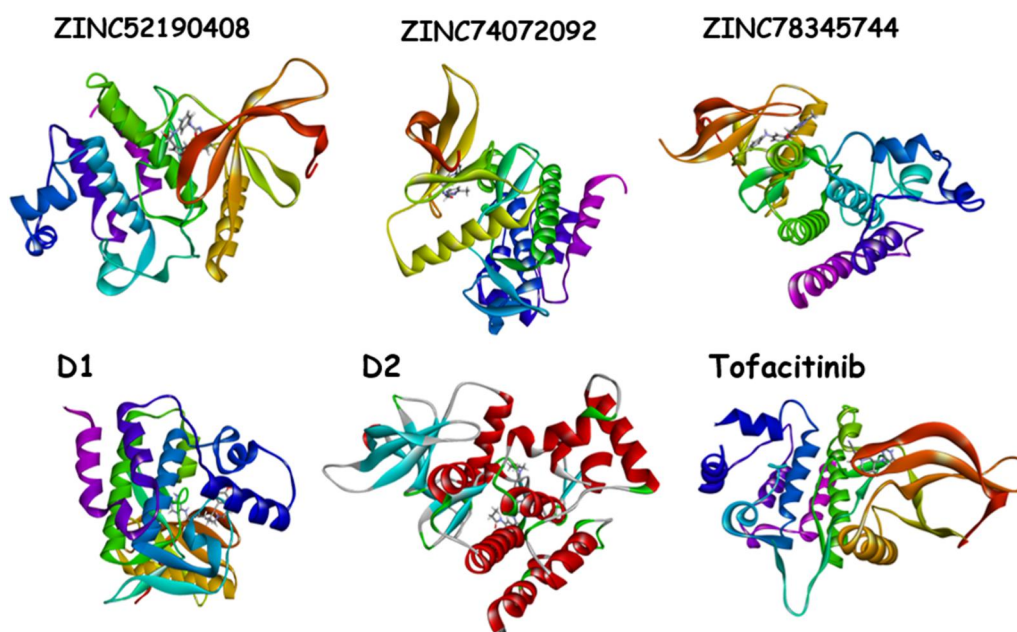**Figure 9.** Free energy landscape For the newly identified and designed compounds.

Figure 10. Stable conformations for each complex.

Table 18. Delta Energy *via* MM/GSBA for ligand complexes (ZINC74072092, ZINC78345744, and ZINC52190408) interacting with JAK3.

Delta energy (Kcal/mol)	ZINC74072092	ZINC78345744	ZINC52190408	D1	D2	Tofacitinib
$\Delta_{VDWAALS}$	−26.18	−34.40	−32.16	−40.32	−47.19	−22.82
ΔE_{EL}	−124.81	−20.82	−10.69	−39.78	−28.82	−32.93
ΔE_{GB}	129.68	37.81	25.76	53.40	50.22	55.89
ΔE_{SURF}	−4.12	−5.25	−4.25	−5.98	−6.33	−3.34
ΔG_{GAS}	−151.00	−55.22	−42.85	−80.10	−76.01	−55.75
ΔG_{SOLV}	125.56	32.56	21.51	47.43	43.88	52.55
Δ_{TOTAL}	−25.44	−22.66	−21.34	−32.67	−32.13	−3.20

complex with Tofacitinib, JAK3 showed cluster motion covering a range on PC1 and PC2 between (−4, 8) and (−2.5, 17.5).

The results of the FEL analysis in Figures 9 and 10 provided the following: The compound ZINC52190408, JAK3, exhibited confirmation stability with an RMSD and RoG of 0.16 and 1.94 nm, respectively. The compound ZINC74072092, JAK3, exhibited confirmation stability with an RMSD and RoG of 0.25 and 1.96 nm, respectively. The compound ZINC52190408, JAK3, exhibited confirmation stability with an RMSD and RoG of 0.22 and 1.94 nm, respectively. The compound ZINC78345744, JAK3, exhibited confirmation stability with an RMSD and RoG of 0.13 and 1.94 nm, respectively. The compound D1, JAK3, exhibited confirmation stability with an RMSD and RoG of 0.23 and 1.98 nm, respectively. The compound D2, JAK3, exhibited confirmation stability with an RMSD and RoG of 0.15 and 1.98 nm, respectively. The drug Tofacitinib, JAK3 (Figure 8), exhibited confirmation stability with an RMSD and RoG of 0.2 and 1.95 nm, respectively. Figure 9 shows the stable conformations for each complex.
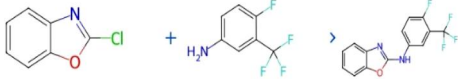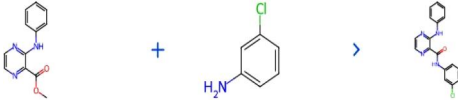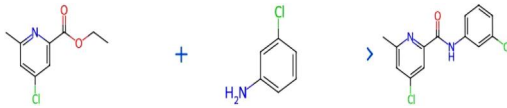
3.10.3. Free binding energy (MM/GBSA)

In the assessment of covalent bond stability between ligands and proteins, MM/GBSA is a commonly employed method following covalent docking simulations (Jin et al., 2020;

Kumar et al., 2022; Wang et al., 2019). While covalent docking simulations aid in predicting potential binding sites, they do not consider the stability of the covalent bond itself. Thus, MM/GBSA serves as a valuable tool for evaluating covalent bond stability and selecting the most stable complexes. Moreover, MM/GBSA plays a crucial role in assessing the stability of ligand-protein complexes post-simulated covalent docking, thereby facilitating a comprehensive understanding of ligand-protein interactions under various physiological or pathological conditions.

The binding energies between JAK3 and the ligands (ZINC74072092, ZINC78345744, ZINC52190408, D1, and D2) were estimated using the MM/GBSA method, as presented in Table 18. The results indicate that three discovery ligands and newly designed exhibit negative binding energies, indicating spontaneous binding to the target protein. Among the ligands, ZINC74072092 displays the lowest binding energy ($\Delta_{TOTAL} = -25.44$ kcal/mol), followed closely by ZINC78345744 ($\Delta_{TOTAL} = -22.66$ kcal/mol) and ZINC52190408 ($\Delta_{TOTAL} = -21.34$ kcal/mol), D1($\Delta_{TOTAL} = -32.67$ kcal/mol) and), D2($\Delta_{TOTAL} = -32.13$ kcal/mol). The contributions of $\Delta_{VDWAALS}$ and ΔE_{EL} are also negative, signifying favorable interactions between the ligands and the target protein. The solvent environment significantly impacts the stability of the complexes. Positive ΔE_{GB} values suggest that the ligands interacting with JAK3 may exhibit reduced affinity for the surrounding solvent or environment. Conversely, negative ΔE_{GB} values indicate a

Table 19. Retrosynthesis of predicted ligands.

ID	Method	Reaction
ZINC52190408	Chloro N-arylation Reference: **US04088770A** Substituted 2-anilinobenzoxazoles used as immunosuppressive agents Similarity: 0.47 Time: 16 hours For a solution of 17.9 g. (0.1 mole) of 4-fluoro-3-(trifluoromethyl)aniline in 200 ml. of tetrahydrofuran was added dropwise 15.4 g. (0.1 mole) of 2-chlorobenzoxazole in 150 ml. of tetrahydrofuran with vigorous agitation. Following the addition, the reaction mixture was heated in a steam bath under reflux for 16 hours. Then the tetrahydrofuran solvent was removed by distillation under vacuum, and 100 ml. of water was added to the residue. The residue was an oily liquid that solidified with the addition of water. The residue was removed by filtration and dried in vacuo. The dried residue was dissolved in a minimum amount of warm acetone, and about 200 ml. of water was added to the solution, resulting in the crystallization of the reaction product. The precipitates were filtered and dried and 24.5 g. of 2-[4-fluoro-3-(trifluoromethyl)anilino]benzoxazole having a melting point of 190°-191° C. was recovered.	
ZINC74072092	Carboxylic ester + amine reaction Reference: **US07947685B2_0145** Pyrazine-2-carboxamide derivatives Similarity: 0.28 The title compound, MS: m/e = 325.1 (M + H+), was prepared by the general method of example 1, step 2 from 3-phenylamino-pyrazine-2-carboxylic acid methyl ester and 3-chloroaniline.	
ZINC78345744	Carboxylic ester + amine reaction Reference: **US20100227887A1_0198** Pyridine-2-yl Carboxylic Acid Amides Similarity: 0.31 The title compound, white solid, MS (ISP) m/e = 281.0 [(M + H)+], mp 174° C., was prepared from 4-chloro-6-methyl-pyridine-2-carboxylic acid ethyl ester [CAS-No. 315494-03-2] and commercially available 3-chloroaniline according to the general method of Example 3.	

higher affinity of the ligands for the interacting surface or interface. Nevertheless, the ΔG_{SOLV} values are lower than the ΔG_{GAS} values, resulting in negative binding energies for all three complexes.

The novel compounds ZINC74072092, ZINC78345744, ZINC52190408, D1, and D2 had more favorable interactions with the target protein than Tofacitinib, MMGBSA data imply that the new compounds can form stable complexes with the JAK3 protein and may have promise as JAK3 therapeutic candidates. The retrosynthesis of D1 and D1 is not available.

3.10.4. Retrosynthesis

Retrosynthesis of identified compounds using ZINCPharmer. The automated retro-synthesis tool, the SPAYA web server, serves as a valuable resource for identifying potential

synthetic pathways for a given target molecule. By leveraging a database of pre-registered reactions, the server generates and suggests feasible syntheses by considering the target molecule and the availability of reagents. The outcomes obtained for the three predicted ligands have been comprehensively compiled and presented in Table 19.

3. Conclusion

In summary, this study employs a multifaceted molecular modeling approach to design and predict novel JAK3/STAT inhibitors based on the pyrimidine-4,6-diamine scaffold. The robust QSAR model, validated through various methods, reliably predicts inhibitory potency with pIC50 values ranging from 6 to 8. A robust pharmacophore model identifies three potential JAK3 inhibitors with favorable ADMET properties

and strong binding interactions. Molecular dynamics simulations confirm inhibitor stability and MM/GBSA calculations validate binding affinity. These inhibitors exhibit promising ADMET profiles and lack AMES toxicity, suggesting their potential for further development. The retrosynthetic analysis aids in synthesis planning. This study advances JAK3 inhibition through comprehensive computational techniques, offering potent therapeutic candidates with favorable profiles and practical insights for future research, contributing to our understanding of JAK3/STAT inhibitors and drug development for related diseases.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Abdelmoujoud Faris http://orcid.org/0000-0003-0781-6523
Ibrahim M. Ibrahim http://orcid.org/0000-0001-7875-7538
Radwan Alnajjar http://orcid.org/0000-0003-1869-9596
Hanine Hadni http://orcid.org/0000-0002-4952-9533
Mashooq Ahmad Bhat http://orcid.org/0000-0001-8426-4692
Muhammad Yaseen http://orcid.org/0000-0002-2259-4589
Souvik Chakraborty http://orcid.org/0000-0003-2628-3988
Nada Alsakhen http://orcid.org/0000-0003-4543-4243
Israa M. Shamkh http://orcid.org/0000-0002-3260-4162
Fazal Mabood http://orcid.org/0000-0002-7282-774X
Ahmed M. Naglah http://orcid.org/0000-0003-4377-5239
Noha Ziedan http://orcid.org/0000-0002-9522-2188

Author contributions

A.F.: Conceptualization, methodology, calculation, investigation, formal analysis; and wrote the original manuscript. I.M.I.: Evaluated the results, Investigation, and formal analysis. R.A: Calculation, formal analysis, evaluated, and edited manuscript. H.H. and M.A.B.: Evaluated the results and methodology. M.Y.: Investigation, and reviewed manuscript. S.V., N.A., I.M.S, F.M., A.M.N., I.U., N.Z.: reviewed and edited manuscript. M.E.: Supervision; formal analysis, reviewed and evaluated the results.

All authors have read and agreed to the published version of the manuscript.

References

Abel, R., Mondal, S., Masse, C., Greenwood, J., Harriman, G., Ashwell, M. A., Bhat, S., Wester, R., Frye, L., Kapeller, R., & Friesner, R. A. (2017). Accelerating drug discovery through tight integration of expert molecular design and predictive scoring. *Current Opinion in Structural Biology*, 21(11), 38–44. https://doi.org/10.1016/j.sbi.2016.10.007

Abraham, M., Alekseenko, A., Bergh, C., Blau, C., Briand, E., Doijade, M., Fleischmann, S., Gapsys, V., Garg, G., Gorelov, S., Gouaillardet, G., Gray, A., Irrgang, M. E., Jalalypour, F., Jordan, J., Junghans, C., Kanduri, P., Keller, S., Kutzner, C., … Lindahl, E. (2023). GROMACS 2023.1 Manual. https://doi.org/10.5281/zenodo.7852189

Alunno, A., Padjen, I., Fanouriakis, A., & Boumpas, D. T. (2019). Pathogenic and therapeutic relevance of JAK/STAT signaling in systemic lupus erythematosus: Integration of distinct inflammatory pathways and the prospect of their inhibition with an oral agent. *Cells*, 8(8), 898. https://doi.org/10.3390/cells8080898

Aqvist, J. (1990). Ion-water interaction potentials derived from free energy perturbation simulations. *The Journal of Physical Chemistry*, 94(21), 8021–8024.

Atkins, P., Atkins, P. W., & de Paula, J. (2014). *Atkins' physical chemistry*. Oxford University Press.

Banerjee, S., Biehl, A., Gadina, M., Hasni, S., & Schwartz, D. M. (2017). JAK–STAT signaling as a target for inflammatory and autoimmune diseases: Current and future prospects. *Drugs*, 77(5), 521–546. https://doi.org/10.1007/s40265-017-0701-9

Bank, R. P. D. (n.d.). *RCSB PDB - 4Z16: Crystal structure of the Jak3 kinase domain covalently bound to N-(3-(((5-chloro-2-((2-methoxy-4-(4-methylpiperazin-1-yl)phenyl)amino)pyrimidin-4-yl)amino)methyl)phenyl)acrylamide*. Retrieved February 5, 2023, from https://www.rcsb.org/structure/4z16

Barf, T., & Kaptein, A. (2012). Irreversible protein kinase inhibitors: Balancing the benefits and risks. *Journal of Medicinal Chemistry*, 55(14), 6243–6262. https://doi.org/10.1021/jm3003203

Boyenle, I. D., Adelusi, T. I., Ogunlana, A. T., Oluwabusola, R. A., Ibrahim, N. O., Tolulope, A., Okikiola, O. S., Adetunji, B. L., Abioye, I. O., & Oyedele, A.-Q K. (2022). Consensus scoring-based virtual screening and molecular dynamics simulation of some TNF-alpha inhibitors. *Informatics in Medicine Unlocked*, 28, 100833. https://doi.org/10.1016/j.imu.2021.100833

Bryan, M. C., & Rajapaksa, N. S. (2018). Kinase inhibitors for the treatment of immunological disorders: Recent advances. *Journal of Medicinal Chemistry*, 61(20), 9030–9058. https://doi.org/10.1021/acs.jmedchem.8b00667

Calculations were performed using Gaussian, D. F. T. (2009). Program: Gaussian 09 (Revision A. 02), MJ Frisch, et al. *Gaussian, Inc.: Wallingford, CT*.

CHARMM-GUI. (n.d.). Retrieved February 5, 2023, from https://www.charmm-gui.org/

Chen, C., Lu, D., Sun, T., & Zhang, T. (2022). JAK3 inhibitors for the treatment of inflammatory and autoimmune diseases: A patent review (2016–present). *Expert Opinion on Therapeutic Patents*, 32(3), 225–242. https://doi.org/10.1080/13543776.2022.2023129

Chen, C., Yin, Y., Shi, G., Zhou, Y., Shao, S., Wei, Y., Wu, L., Zhang, D., Sun, L., & Zhang, T. (2022). A highly selective JAK3 inhibitor is developed for treating rheumatoid arthritis by suppressing γc cytokine–related JAK-STAT signal. *Science Advances*, 8(33), eabo4363. https://doi.org/10.1126/sciadv.abo4363

Chowdhury, M. Z. I., & Turin, T. C. (2020). Variable selection strategies and its importance in clinical prediction modelling. *Family Medicine and Community Health*, 8(1), e000262. https://doi.org/10.1136/fmch-2019-000262

Cox, I., & Gaudard, M. (2013). *Discovering partial least squares with JMP*. SAS Institute.

Daoui, O., Elkhattabi, S., Bakhouch, M., Belaidi, S., Bhandare, R. R., Shaik, A. B., Mali, S. N., & Chtita, S. (2023). Cyclohexane-1,3-dione derivatives as future therapeutic agents for NSCLC: QSAR modeling, in silico ADME-Tox properties, and structure-based drug designing approach. *ACS Omega*, 8(4), 4294–4319. https://doi.org/10.1021/acsomega.2c07585

Dearden, J. C. (2007). In silico prediction of ADMET properties: How far have we come? *Expert Opinion on Drug Metabolism & Toxicology*, 3(5), 635–639. https://doi.org/10.1517/17425255.3.5.635

Dhiman, S., Saha, M., Ali, A., Ali, A., Gupta, G. D., & Asati, V. (2023). Structural aspects of triazole derivatives as HSP90 inhibitors for the treatment of cancer: In silico studies. *Journal of Biomolecular Structure*

& *Dynamics*, *41*(10), 4756–4769. https://doi.org/10.1080/07391102. 2022.2083686

Ding, Q., Hu, W., Wang, R., Yang, Q., Zhu, M., Li, M., Cai, J., Rose, P., Mao, J., & Zhu, Y. Z. (2023). Signaling pathways in rheumatoid arthritis: Implications for targeted therapy. *Signal Transduction and Targeted Therapy*, *8*(1). https://doi.org/10.1038/s41392-023-01331-9

Dong, Z., Liu, C.-H., Wang, Y., Lin, M., & Yu, Z.-X. (2013). Gold (I)-catalyzed endo-selective intramolecular a-alkenylation of b-Yne-furans: Synthesis of seven-membered-ring-fused furans and DFT calculations. *Angewandte Chemie International Edition*, *52*(52), 14157–14161. https://doi.org/10.1002/anie.201306965

Emery, P., Pope, J. E., Kruger, K., Lippe, R., DeMasi, R., Lula, S., & Kola, B. (2018). Efficacy of monotherapy with biologics and JAK inhibitors for the treatment of rheumatoid arthritis: A systematic review. *Advances in Therapy*, *35*(10), 1535–1563. https://doi.org/10.1007/s12325-018-0757-2

En-Nahli, F., Baammi, S., Hajji, H., Alaqarbeh, M., Lakhlifi, T., & Bouachrine, M. (2022). High-throughput virtual screening approach of natural compounds as target inhibitors of plasmepsin-II. *Journal of Biomolecular Structure & Dynamics*, *41*(19), 10070–10080. https://doi. org/10.1080/07391102.2022.2152871

Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T., McDowell, R. M., & Gramatica, P. (2003). Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs. *Environmental Health Perspectives*, *111*(10), 1361–1375. https://doi.org/10.1289/ehp.5758

Esposito, E. X., Hopfinger, A. J., & Madura, J. D. (2004). Methods for applying the quantitative structure-activity relationship paradigm. In *Chemoinformatics: Concepts, methods, and tools for drug discovery* (pp. 131–213).

Faris, A., Cacciatore, I., Ibrahim, I. M., Al Mughram, M. H., Hadni, H., Tabti, K., & Elhallaoui, M. (2023). In silico computational drug discovery: A Monte Carlo approach for developing a novel JAK3 inhibitors. *Journal of Biomolecular Structure & Dynamics*, 1–23. https://doi.org/10.1080/ 07391102.2023.2270709

Faris, A., Hadni, H., Ibrahim, I. M., & Elhallaoui, M. (2023). In silico discovery of potent and selective Janus kinase 3 (JAK3) inhibitors through 3D-QSAR, covalent docking, ADMET analysis, molecular dynamics simulations, and binding free energy of pyrazolopyrimidine derivatives. *Journal of Biomolecular Structure & Dynamics*, 1–17. https://doi.org/10. 1080/07391102.2023.2222839

Faris, A., Hadni, H., Saleh, B. A., Khelfaoui, H., Harkati, D., Ait Ahsaine, H., Elhallaoui, M., & El-Hiti, G. A. (2023). In silico screening of a series of 1,6-disubstituted 1H-pyrazolo[3,4-d]pyrimidines as potential selective inhibitors of the Janus kinase 3. *Journal of Biomolecular Structure & Dynamics*, 1–19. https://doi.org/10.1080/07391102.2023.2220829

Faris, A., Ibrahim, I. M., Al Kamaly, O., Saleh, A., & Elhallaoui, M. (2023). Computer-aided drug design of novel derivatives of 2-amino-7,9-dihydro-8H-purin-8-one as potent Pan-Janus JAK3 inhibitors. *Molecules (Basel, Switzerland)*, *28*(15), 5914. https://doi.org/10.3390/molecules28155914

Faris, A., Ibrahim, I. M., Hadni, H., & Elhallaoui, M. (2023). High-throughput virtual screening of phenylpyrimidine derivatives as selective JAK3 antagonists using computational methods. *Journal of Biomolecular Structure & Dynamics*, 1–26. https://doi.org/10.1080/07391102.2023. 2240413

Ferreira, L. L. G., & Andricopulo, A. D. (2019). ADMET modeling approaches in drug discovery. *Drug Discovery Today*, *24*(5), 1157–1165. https://doi.org/10.1016/j.drudis.2019.03.015

Fetter, T., Smith, P., Guel, T., Braegelmann, C., Bieber, T., & Wenzel, J. (2020). Selective janus kinase 1 inhibition is a promising therapeutic approach for lupus erythematosus skin lesions. *Frontiers in Immunology*, *11*, 344. https://doi.org/10.3389/fimmu.2020.00344

Free Chemical Drawing Software for Students | ChemSketch | ACD/Labs. (n.d). Retrieved March 18, (2023). from https://www.acdlabs.com/ resources/free-chemistry-software-apps/chemsketch-freeware/

Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., & Mainz, D. T. (2006). Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein- ligand complexes. *Journal of Medicinal Chemistry*, *49*(21), 6177–6196. https://doi.org/10.1021/jm051256o

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J. A., Vreven, T., Kudin, K. N., & Burant, J. C. (2009). *Gaussian, Inc: Gaussian 03, Revision B. 03 AND Gaussian 09, Revision A. 02*. Gaussian Inc.

Gadina, M., Johnson, C., Schwartz, D., Bonelli, M., Hasni, S., Kanno, Y., Changelian, P., Laurence, A., & O'Shea, J. J. (2018). Translational and clinical advances in JAK-STAT biology: The present and future of jakinibs. *Journal of Leukocyte Biology*, *104*(3), 499–514. https://doi.org/10. 1002/JLB.5RI0218-084R

Gholamhoseinnia, M., & Asadollahi-Baboli, M. (2022). Ranked binding energies of residues and data fusion to identify the active and selective pyrimidine-based Janus kinases 3 (JAK3) inhibitors. *SAR and QSAR in Environmental Research*, *33*(1), 23–34. https://doi.org/10.1080/ 1062936X.2021.2013318

Ghoreschi, K., Laurence, A., & O'Shea, J. J. (2009). Janus kinases in immune cell signaling. *Immunological Reviews*, *228*(1), 273–287. https://doi.org/10.1111/j.1600-065X.2008.00754.x

Golbraikh, A., & Tropsha, A. (2002). Beware of q2!. *Journal of Molecular Graphics & Modelling*, *20*(4), 269–276. https://doi.org/10.1016/S1093-3263(01)00123-1

Hadni, H., & Elhallaoui, M. (2019). Molecular docking and QSAR studies for modeling the antimalarial activity of hybrids 4-anilinoquinoline-triazines derivatives with the wild-type and mutant receptor pf-DHFR. *Heliyon*, *5*(8), e02357. https://doi.org/10.1016/j.heliyon.2019.e02357

Hadni, H., & Elhallaoui, M. (2020). 2D and 3D-QSAR, molecular docking and ADMET properties: In silico studies of azaaurones as antimalarial agents. *New Journal of Chemistry*, *44*(16), 6553–6565. https://doi.org/ 10.1039/C9NJ05767F

Hadni, H., Mazigh, M., Charif, E., Bouayad, A., & Elhallaoui, M. (2018). Molecular modeling of antimalarial agents by 3D-QSAR study and molecular docking of two hybrids 4-aminoquinoline-1, 3, 5-triazine and 4-aminoquinoline-oxalamide derivatives with the receptor protein in its both wild and mutant types. *Biochemistry Research International*, *2018*, 8639173–8639115. https://doi.org/10.1155/2018/8639173

Hosking, A.-M., Juhasz, M., & Mesinkovska, N. A. (2018). Topical Janus kinase inhibitors: A review of applications in dermatology. *Journal of the American Academy of Dermatology*, *79*(3), 535–544. https://doi.org/10. 1016/j.jaad.2018.04.018

Hosseini, A., Gharibi, T., Marofi, F., Javadian, M., Babaloo, Z., & Baradaran, B. (2020). Janus kinase inhibitors: A therapeutic strategy for cancer and autoimmune diseases. *Journal of Cellular Physiology*, *235*(9), 5903–5924. https://doi.org/10.1002/jcp.29593

Hu, L., Liu, R., & Zhang, L. (2022). Advance in bone destruction participated by JAK/STAT in rheumatoid arthritis and therapeutic effect of JAK/STAT inhibitors. *International Immunopharmacology*, *111*, 109095. https://doi.org/10.1016/j.intimp.2022.109095

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

Izadyar, M., Housaindokht, M. R., Zavvar, N., Khavani, M., & Reisi-Vanani, A. (2015). Secondary structure effects on the acidity of histidine and lysine-based peptides model; A theoretical study. *Physical Chemistry Research*, *3*(1), 67–77.

Jaffar, S. (2015). *Optimizing selectivity in heterocycle CH functionalization through computational design* [PhD Thesis]. University of Oxford.

Jain, D., Udhwani, T., Sharma, S., Gandhe, A., Reddy, P. B., Nayarisseri, A., & Singh, S. K. (2019). Design of novel JAK3 inhibitors towards rheumatoid arthritis using molecular docking analysis. *Bioinformation*, *15*(2), 68–78. https://doi.org/10.6026/97320630015068

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. *Computational Biology and Chemistry*, *85*, 107241. https://doi.org/10. 1016/j.compbiolchem.2020.107241

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

Kapetanovic, I. M. (2008). Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. *Chemico-*

Biological Interactions, *171*(2), 165–176. https://doi.org/10.1016/j.cbi.2006.12.006

Kim, J., Kim, S., & Schaumann, G. E. (2013). Reliable predictive computational toxicology methods for mixture toxicity: Toward the development of innovative integrated models for environmental risk assessment. *Reviews in Environmental Science and Bio/Technology*, *12*(3), 235–256. https://doi.org/10.1007/s11157-012-9286-7

Kukol, A. (2014). *Molecular modeling of proteins*. 2nd ed. (Vol. 1215, p. 474). https://doi.org/10.1007/978-1-4939-1465-4

Kumar, B. K., Faheem, n, Sekhar, K. V. G. C., Ojha, R., Prajapati, V. K., Pai, A., & Murugesan, S. (2022). Pharmacophore based virtual screening, molecular docking, molecular dynamics and MM-GBSA approach for identification of prospective SARS-CoV-2 inhibitor from natural product databases.*Journal of Biomolecular Structure & Dynamics*, *40*(3), 1363–1386. https://doi.org/10.1080/07391102.2020.1824814

Kumari, R., Kumar, R., Consortium, O. S. D. D., & Lynn, A. (2014). G_mmpbsa— A GROMACS tool for high-throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling*, *54*(7), 1951–1962. https://doi.org/10.1021/ci500020m

Kurdi, M., & Booz, G. W. (2009). JAK redux: A second look at the regulation and role of JAKs in the heart. *American Journal of Physiology. Heart and Circulatory Physiology*, *297*(5), H1545–H1556. https://doi.org/10.1152/ajpheart.00032.2009

Lai, Y., & Dong, C. (2016). Therapeutic antibodies that target inflammatory cytokines in autoimmune diseases. *International Immunology*, *28*(4), 181–188. https://doi.org/10.1093/intimm/dxv063

Li, G., Waite, E., & Wolfson, J. (2017). T-cell prolymphocytic leukemia in an adolescent with ataxia-telangiectasia: Novel approach with a JAK3 inhibitor (tofacitinib). *Blood Advances*, *1*(27), 2724–2728. https://doi.org/10.1182/bloodadvances.2017010470

Malemud, C. J., & Pearlman, E. (2009). Targeting JAK/STAT signaling pathway in inflammatory diseases. *Current Signal Transduction Therapy*, *4*(3), 201–221. https://doi.org/10.2174/157436209789057467

McKeage, K. (2015). Ruxolitinib: A review in polycythaemia vera. *Drugs*, *75*(15), 1773–1781. https://doi.org/10.1007/s40265-015-0470-2

Moe, C. C. G. (2014). Molecular operating environment. *H3A 2R7*.

Moriwaki, H., Tian, Y.-S., Kawashita, N., & Takagi, T. (2018). Mordred: A molecular descriptor calculator. *Journal of Cheminformatics*, *10*(1), 4. https://doi.org/10.1186/s13321-018-0258-y

Nascimento, I. J. D. S., de Aquino, T. M., & da Silva-Júnior, E. F. (2021). Drug repurposing: A strategy for discovering inhibitors against emerging viral infections. *Current Medicinal Chemistry*, *28*(15), 2887–2942.

Nascimento, I. J. D. S., de Aquino, T. M., & da Silva-Júnior, E. F. (2022). The New era of drug discovery: The power of computer-aided drug design (CADD). *Letters in Drug Design & Discovery*, *19*(11), 951–955. https://doi.org/10.2174/1570180819666220405225817

O'Shea, J. J., Kontzias, A., Yamaoka, K., Tanaka, Y., & Laurence, A. (2013). Janus kinase inhibitors in autoimmune diseases. *Annals of the Rheumatic Diseases*, *72* (2), ii111–ii115. https://doi.org/10.1136/annrheumdis-2012-202576

Ouattara, O., Affi, T. S., Koné, M. G.-R., Bamba, K., & Ziao, N. (2017). Can empirical descriptors reliably predict molecular lipophilicity? A QSPR study investigation. *International Journal of Engineering Research and Applications*, *7*(5), 50–56. https://doi.org/10.9790/9622-0705015056

Palacios-Prado, N., Soto, P. A., López, X., Choi, E. J., Marquez-Miranda, V., Rojas, M., Duarte, Y., Lee, J., González-Nilo, F. D., & Sáez, J. C. (2022). Endogenous pannexin1 channels form functional intercellular cell–cell channels with characteristic voltage-dependent properties. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(18), e2202104119. https://doi.org/10.1073/pnas.2202104119

Pires, D. E. V., Blundell, T. L., & Ascher, D. B. (2015). pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *Journal of Medicinal Chemistry*, *58*(9), 4066–4072. https://doi.org/10.1021/acs.jmedchem.5b00104

Ramsey, P. J. (2022). Analysis strategies for constrained mixture and mixture process experiments using JMP Pro 14. JMP Discovery Summit, 7.

Salas, A., Hernandez-Rocha, C., Duijvestein, M., Faubion, W., McGovern, D., Vermeire, S., Vetrano, S., & Vande Casteele, N. (2020). JAK–STAT pathway targeting for the treatment of inflammatory bowel disease.

Nature Reviews. Gastroenterology & Hepatology, *17*(6), 323–337. https://doi.org/10.1038/s41575-020-0273-0

Schrödinger Release 2021-1; Maestro. (2021). *Schrödinger, LLC: New York, NY, USA*.

Schrödinger Release. (2021). *2021-1 (Version 2021-1) [Computer software]*.

Snedecor, G. W., & Cochran, W. G. (1967). *Methods, statistical*. Oxford and IBH, p. 381.

Spaya.ai. (n.d). Retrieved March 15, 2023 from https://spaya.ai/app/search

Su, W., Chen, Z., Liu, M., He, R., Liu, C., Li, R., Gao, M., Zheng, M., Tu, Z., Zhang, Z., & Xu, T. (2022). Design, synthesis and structure-activity relationship studies of pyrido[2,3-d]pyrimidin-7-ones as potent Janus Kinase 3 (JAK3) covalent inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *64*, 128680. https://doi.org/10.1016/j.bmcl.2022.128680

Suma, K. B., Kumari, A., Shetty, D., Fernandes, E., Chethan, D. V., Jays, J., & Murahari, M. (2020). Structure based pharmacophore modelling approach for the design of azaindole derivatives as DprE1 inhibitors for tuberculosis. *Journal of Molecular Graphics & Modelling*, *101*, 107718. https://doi.org/10.1016/j.jmgm.2020.107718

Sun, R., Chen, M., Hu, Y., Lan, Y., Gan, L., You, G., Yue, M., Wang, H., Xia, B., Zhao, J., Tang, L., Cai, Z., Liu, Z., & Ye, L. (2019). CYP3A4/5 mediates the metabolic detoxification of humantenmine, a highly toxic alkaloid from Gelsemium elegans Benth. *Journal of Applied Toxicology: JAT*, *39*(9), 1283–1292. https://doi.org/10.1002/jat.3813

Systèmes, D. (2020, March 20). *Free Download: BIOVIA Discovery Studio Visualizer*. Dassault Systèmes. https://discover.3ds.com/discovery-studio-visualizer-download

Tan, L., Akahane, K., McNally, R., Reyskens, K. M. S. E., Ficarro, S. B., Liu, S., Herter-Sprie, G. S., Koyama, S., Pattison, M. J., Labella, K., Johannessen, L., Akbay, E. A., Wong, K.-K., Frank, D. A., Marto, J. A., Look, T. A., Arthur, J. S. C., Eck, M. J., & Gray, N. S. (2015). Development of selective covalent Janus kinase 3 inhibitors. *Journal of Medicinal Chemistry*, *58*(16), 6589–6606. https://doi.org/10.1021/acs.jmedchem.5b00710

Thatcher, J. E., Zelter, A., & Isoherranen, N. (2010). The relative importance of CYP26A1 in hepatic clearance of all-trans retinoic acid. *Biochemical Pharmacology*, *80*(6), 903–912. https://doi.org/10.1016/j.bcp.2010.05.023

Thyroid hormones induce doxorubicin chemosensitivity through enzymes involved in chemotherapy metabolism in lymphoma T cells—PMC. (n.d). Retrieved May 31, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6508960/

Tropsha, A., Gramatica, P., & Gombar, V. K. (2003). The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science*, *22*(1), 69–77. https://doi.org/10.1002/qsar.200390007

Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), 455–461. https://doi.org/10.1002/jcc.21334

Valdés-Tresanco, M. S., Valdés-Tresanco, M. E., Valiente, P. A., & Moreno, E. (2021). gmx_MMPBSA: A new tool to perform end-state free energy calculations with GROMACS. *Journal of Chemical Theory and Computation*, *17*(10), 6281–6291. https://doi.org/10.1021/acs.jctc.1c00645

Villoutreix, B. O., & Taboureau, O. (2015). Computational investigations of hERG channel blockers: New insights and current predictive models. *Advanced Drug Delivery Reviews*, *86*, 72–82. https://doi.org/10.1016/j.addr.2015.03.003

Wang, D.-P., Wu, L.-H., Li, R., He, N., Zhang, Q.-Y., Zhao, C.-Y., & Jiang, T. (2023). A Novel Aldisine Derivative Exhibits Potential Antitumor Effects by Targeting JAK/STAT3 Signaling. *Marine Drugs*, *21*(4), 218. https://doi.org/10.3390/md21040218

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

White, J. R., Phillips, F., Monaghan, T., Fateen, W., Samuel, S., Ghosh, S., & Moran, G. W. (2018). Review article: Novel oral-targeted therapies in

inflammatory bowel disease. *Alimentary Pharmacology & Therapeutics*, *47*(12), 1610–1622. https://doi.org/10.1111/apt.14669

Xin, P., Xu, X., Deng, C., Liu, S., Wang, Y., Zhou, X., Ma, H., Wei, D., & Sun, S. (2020). The role of JAK/STAT signaling pathway and its inhibitors in diseases. *International Immunopharmacology*, *80*, 106210. https://doi.org/10.1016/j.intimp.2020.106210

Yap, C. W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry*, *32*(7), 1466–1474. https://doi.org/10.1002/jcc.21707

Ylilauri, M., & Pentikäinen, O. T. (2013). MMGBSA As a Tool To Understand the Binding Affinities of Filamin–Peptide Interactions. *Journal of Chemical Information and Modeling*, *53*(10), 2626–2633. https://doi.org/10.1021/ci4002475

Yu, R.-N., Chen, C.-J., Shu, L., Yin, Y., Wang, Z.-J., Zhang, T.-T., & Zhang, D.-Y. (2019). Structure-based design and synthesis of pyrimidine-4,6-diamine derivatives as Janus kinase 3 inhibitors. *Bioorganic & Medicinal Chemistry*, *27*(8), 1646–1657. https://doi.org/10.1016/j.bmc.2019.03.009

Zhong, H. A., & Almahmoud, S. (2023). Docking and selectivity studies of covalently bound janus kinase 3 inhibitors. *International Journal of Molecular Sciences*, *24*(7), 6023. https://doi.org/10.3390/ijms24076023

Zhou, X.-Y., Hu, X.-X., Wang, C.-C., Lu, X.-R., Chen, Z., Liu, Q., Hu, G.-X., & Cai, J.-P. (2019). Enzymatic activities of CYP3A4 allelic variants on quinine 3-hydroxylation in vitro. *Frontiers in Pharmacology*, *10*, 591. https://doi.org/10.3389/fphar.2019.00591

Zhu, J., Sun, D., Li, X., Jia, L., Cai, Y., Chen, Y., Jin, J., & Yu, L. (2023). Developing new PI3Kγ inhibitors by combining pharmacophore modeling, molecular dynamic simulation, molecular docking, fragment-based drug design, and virtual screening. *Computational Biology and Chemistry*, *104*, 107879. https://doi.org/10.1016/j.compbiolchem.2023.107879

Zhu, Y., Zheng, X., Wang, C., Sun, X., Sun, H., Ma, T., Li, Y., Liu, K., Chen, L., & Ma, X. (2020). Synthesis and biological activity of thieno[3,2-d]pyrimidines as potent JAK3 inhibitors for the treatment of idiopathic pulmonary fibrosis. *Bioorganic & Medicinal Chemistry*, *28*(2), 115254. https://doi.org/10.1016/j.bmc.2019.115254