



OPEN

Computer aided designing of novel pyrrolopyridine derivatives as JAK1 inhibitors

Seketoulie Keretsu¹, Suparna Ghosh¹ & Seung Joo Cho^{1,2}✉

Janus kinases (JAKs) are a family of non-receptor kinases that play a key role in cytokine signaling and their aberrant activities are associated with the pathogenesis of various immune diseases. The JAK1 isoform plays an essential role in the types 1 and II interferon signaling and elicits signals from the interleukin-2, interleukin-4, gp130, and class 2 receptor families. It is ubiquitously expressed in humans and its overexpression has been linked with autoimmune diseases such as myeloproliferative neoplasm. Although JAK1 inhibitors such as Tofacitinib have been approved for medical use, the low potency and off-target effects of these inhibitors have limited their use and calls for the development of novel JAK1 inhibitors. In this study, we used computational methods on a series of pyrrolopyridine derivatives to design new JAK1 inhibitors. Molecular docking and molecular dynamics simulation methods were used to study the protein-inhibitor interactions. 3D-quantitative structure–activity relationship models were developed and were used to predict the activity of newly designed compounds. Free energy calculation methods were used to study the binding affinity of the inhibitors with JAK1. Of the designed compounds, seventeen of the compounds showed a higher binding energy value than the most active compound in the dataset and at least six of the compounds showed higher binding energy value than the pan JAK inhibitor Tofacitinib. The findings made in this study could be utilized for the further development of JAK1 inhibitors.

Janus kinases (JAK) are non-receptor tyrosine kinases that mediate cellular signaling from the extracellular matrix via the cytokine receptors¹. They play a crucial role in type 1 and type 2 cytokines mediated signaling². The binding of endogenous ligands such as cytokines and interferons at the transmembrane receptors leads to the activation of JAKs, which in turn phosphorylate the transcription factor STAT kinases³. The activated STAT kinases modulate the expression of genes that are associated with cell proliferation, apoptosis, and angiogenesis. The dysregulation of the JAK/STAT signaling has been linked with various autoimmune disorders. The JAK family consists of the isoforms JAK1, JAK2, JAK3, and the tyrosine kinase 2 (TYK2)⁴. While the JAK1 and JAK2 are ubiquitously expressed, the TYK2 is predominantly expressed in the endothelial. The JAK1 isoform regulates cytokine signaling from the interleukin-2 (IL2) receptor, interleukin-4 (IL4) receptor, and glycoprotein 130 receptor families. They also transduce signaling from type 1 and type 2 interferons⁵. The aberrant activity of JAK1 is linked with the pathogenesis of inflammatory disorders and immune diseases like rheumatoid arthritis (RA), ulcerative colitis, and Crohn's disease^{6,7}.

The JAK1 shares a high sequence identity with the other members of the JAK family. Structurally, JAK1 consists of seven distinct domains termed JAK homology (JH1-JH7) domains. The JH1 domain consists of the protein tyrosine kinase domain and plays a significant role in physiological function. The JH2 domain consists of a kinase-like domain called the pseudokinase domain that regulates the activity of the protein tyrosine kinases (PTK) domain. The JH3 and JH4 share some homology with the Src homology 2 (SH-2) domains and stabilize the conformation of JAK1. The JH5 and JH7 domains make up the N-terminus of JAK1 and play an essential role in the regulation of the JAK1. The ATP binding takes place at the cavity formed by the residues from the glycine-rich loop, activation loop, hinge region, and α C-helix at the JH2 domain.

Currently, JAK1 inhibitors such as Filgotinib, Tofacitinib, Peficitinib, and several other second-generation inhibitors are in consideration for the treatment of inflammatory disorders and autoimmune disorders^{8–10}. Tofacitinib was approved for oral use against RA. However, it is a multi-JAK inhibitor with inhibitory activity against JAK2, JAK3, and TYK2¹¹. Filgotinib is a second-generation, orally available drug in clinical trials for RA. However, the drug is a potent inhibitor of JAK1, JAK2, and TYK2¹². Peficitinib is another pan JAK inhibitor

¹Department of Biomedical Sciences, College of Medicine, Chosun University, 375 Seosuk-dong, Dong-gu, Gwangju 501-759, Republic of Korea. ²Department of Cellular Molecular Medicine, College of Medicine, Chosun University, Gwangju 501-759, Republic of Korea. ✉email: chosj@chosun.ac.kr

that has been approved for clinical use in Japan. Clinical trials have shown that Peficitinib is effective against RA with an acceptable safety profile^{10,13}. Several other JAK1 inhibitors are also in consideration for the treatment of autoimmune and inflammatory diseases¹⁴. However, due to the low potency, non-selectivity, and off-target effects of existing drugs, developing novel JAK1 inhibitors with high potency is in crucial need for the treatment of RA.

We used computational modeling techniques such as 3D-QSAR, molecular docking, molecular dynamics simulation, and free energy calculation methods to study a series of pyrrolopyrimidine based JAK1 inhibitors. The study was carried out to explore the protein–ligand interactions and to characterize structural features important for the inhibition of JAK1. 3D-QSAR models were developed for structure–activity relationship analysis and to predict the activity value of novel JAK1 inhibitors. The relative binding affinity of the inhibitors and the newly designed compounds were calculated using free energy calculation methods.

Methodology

Data collection and protein preparation. The dataset of 50 pyrrolopyrimidine based compounds that showed inhibitory activity for JAK1 was collected from the literature¹⁵. The half-maximal inhibitory concentration (IC₅₀) values were converted to negative log values (pIC₅₀). The inhibitory activity (pIC₅₀) values of the compounds ranged from 5 to 9.7. The structure of the dataset compounds and the activity values are given in Table 1. The X-ray crystallographic structure of JAK1 in complex with the Tofacitinib (PDB ID 3EYG) at 1.9 Å resolutions was collected from the protein repository¹⁶. The protein structure was prepared by removing solvents molecules and adding the missing atoms.

Molecular docking. The protein–ligand binding was carried out using Autodock 4.2 (The Scripps Research Institute, La Jolla, CA, USA)¹⁷. Prior to the docking, the protein was prepared by adding polar hydrogen atoms, followed by the addition of Kollman charges. The inhibitor was sketched and minimized in Sybyl X 2.1 (Tripos Inc., St. Louis, MO, USA). Gasteiger charges were added as partial charges. The conformational search area within the receptor was defined by a 60 × 60 × 60 grid box. The grid space was set to 0.37 Å. The Lamarckian Genetic algorithm within the Autodock program was used to perform the conformational search and generate 100 conformations. The autodock tools package was used to analyze the outcome of the docking¹⁷. The protocol described above was used during the docking of the dataset compounds as well as the docking of the newly designed compounds.

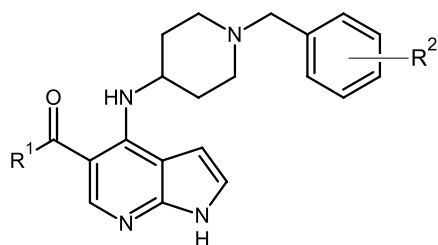
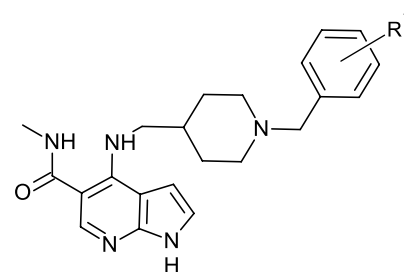
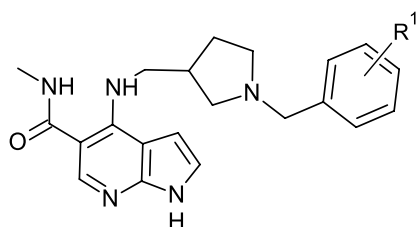
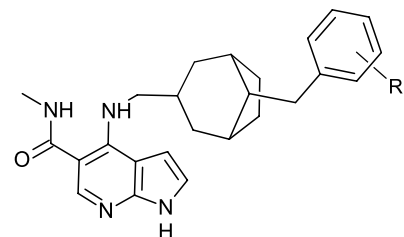
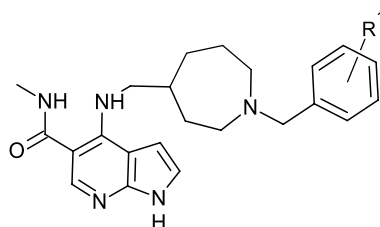
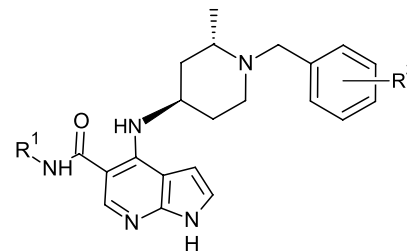
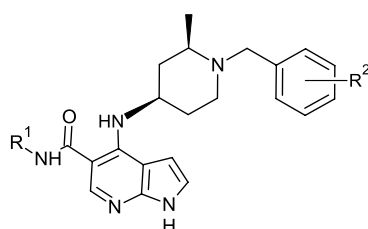
Molecular dynamics. Molecular dynamics simulation was performed using the Gromacs 2020¹⁸. The docked protein–ligand structure from the docking study was used as the initial structure for simulation. The ACPYPE (or AnteChamber PYthon Parser interface) script was used to develop the ligand parameters^{19,20}. The amber force field Amber99SB was used to parameterize the protein²⁰. The protein–ligand complex was placed inside a cubic box and was solvated with a TIP3P water model. Na⁺ ions and Cl[−] ions were added to the system to neutralize the system. The system was energy minimized with the maximum force F_{\max} set to 1000 kJ mol^{−1} nm^{−1} to remove steric contacts. Position restrained NVT ensemble and NPT ensemble equilibrations were carried out to gradually heat the system to a pressure and temperature of 1 bar and 300 K, respectively. During the equilibration, V-rescale was used for temperature coupling and Parrinello–Rahman was used for pressure coupling. The long-range electrostatic interactions were estimated using the Particle Mesh Ewald method²⁰. During the simulation, hydrogen bonds were constrained using LINCS methods and the minimum time step was kept at 2 fs. The unrestrained production was performed using the leap-frog integrator. A more detailed explanation of the simulation methodology can be found in previous protein–ligand interaction studies by our group^{21–24}. Protein–ligand snapshots were collected from the equilibrated region of the MD simulation and the binding energy values were estimated using the Molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) method implemented by Kumari et al.²⁵.

The *g-mmpbsa* method implemented by Kumari et al.²⁵ does not consider the entropy term hence, the methods can only give the relative BE. Though there are several methods to calculate the entropy term^{26,27}, these methods are time-consuming and the magnitude of the standard error is high. Hence, the binding energy results presented in the study are considered as relative binding energy values. More detail regarding the MM/PBSA methods can be found in the supplementary material.

Three dimensional structure–activity relationships (3D-QSAR). The compounds were sketched based on the binding conformation of the most active compound 42 inside JAK1 as observed in the MD simulation. The compounds were prepared by adding Gasteiger charges as partial charges and then minimizing the substituents in Sybyl X 2.1 (Tripos Inc., St. Louis, MO, USA). Using the distill-rigid alignment method, the compounds were superimposed on the common substructure, using the most active compound as the template.

3D-QSAR models were developed to analyze the relationship between the physicochemical descriptor of the compounds and their biological activity values (pIC₅₀). The dataset was split into test sets and training set for training and validating the models using a stratified random sampling approach. As a first step, the compounds were split into various mutually exclusive and non-overlapping groups (or strata) based on the pIC₅₀ values. Based on a random draw, compounds belonging to each group were split into a training set and test set in a 2:1 ratio. The final training set consists of the training sets compounds from all the groups. Similarly, the final test set consists of the test set compounds from all the groups. This process of drawing test set and training set was repeated five times resulting in five sets of the test set and training set pairs (Sets 1–5).

Taking the first test set and training set pair (Set 1), Comparative Molecular Field Analysis (CoMFA) models were developed for various charge schemes namely Gasteiger, Gasteiger–Hückel, Delre, Pullman, and MMFF94^{28,29}. In CoMFA, the electrostatic and steric descriptors were used as the independent variables and the

**Structure A****Structure B****Structure C****Structure D****Structure E****Structure F****Structure G**

	Structure	R1	R2	pIC₅₀
1	A	-NH ₂	H	7.06
2	A	-NH ₂	3-OMe	6.71
3	A	-NHMe	H	7.19
4	A	-NHMe	2-OMe	6.39
5*	A	-NHMe	3-OMe	7.74
6*	A	-NHMe	4-OMe	6.69
7	A	-NHMe	-3, 4-diOMe	5.08
8*	A	-NHMe	3-CN	7.09

Table 1. (continued)

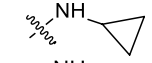
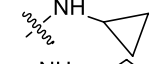
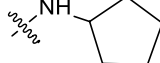
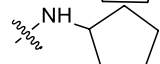
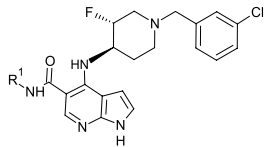
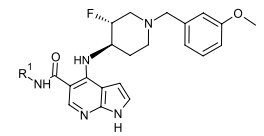
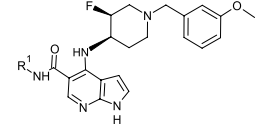
9	A		H	6.94
10	A		3-OMe	5.53
11	A		H	5.00
12	A		3-OMe	5.48
13	A	H	H	6.80
14	B	H	–	6.67
15*	B	3-Cl	–	6.61
16*	B	4-Cl	–	7.14
17	B	3-OMe	–	6.88
18	B	4-OMe	–	7.03
19	C	H	–	6.40
20*	C	3-Cl	–	7.22
21	C	4-Cl	–	7.10
22	C	3-OMe	–	6.08
23*	C	4-OMe	–	6.83
24	D	H	–	7.13
25	D	3-Cl	–	7.39
26	D	4-Cl	–	7.76
27*	D	3-OMe	–	7.34
28	D	4-OMe	–	7.47
29	D	4-CN	–	7.82
30	E	H	–	7.42
31*	E	3-Cl	–	8.03
32	E	4-Cl	–	7.92
33*	E	3-OMe	–	8.07
34	E	4-OMe	–	7.52
35*	E	4-CN	–	7.49
36	F	Me	H	7.97
37*	F	Me	3-Cl	8.95
38	F	Me	4-Cl	8.25
39	F	Me	3-OMe	8.47
40*	F	Me	4-OMe	7.56
41	G	Me	3-Cl	8.50
42	G	Me	4-Cl	9.72
43	G	Me	3-OMe	7.89
44*	G	Me	4-OMe	7.62
45	G	Me	4-CN	8.71
46	G	H	3-Cl	8.16
47	G	H	3-OMe	7.76
48*				7.77
49				8.1
50				9.12

Table 1. Dataset of pyrrolopyridine based JAK1 inhibitors. *Test set compounds.

activity values of the compounds were used as the dependent variables. The models were validated using various internal and external validation techniques. The non-validated Leave-One-Out (LOO) method was used to estimate the predictive ability of the models.

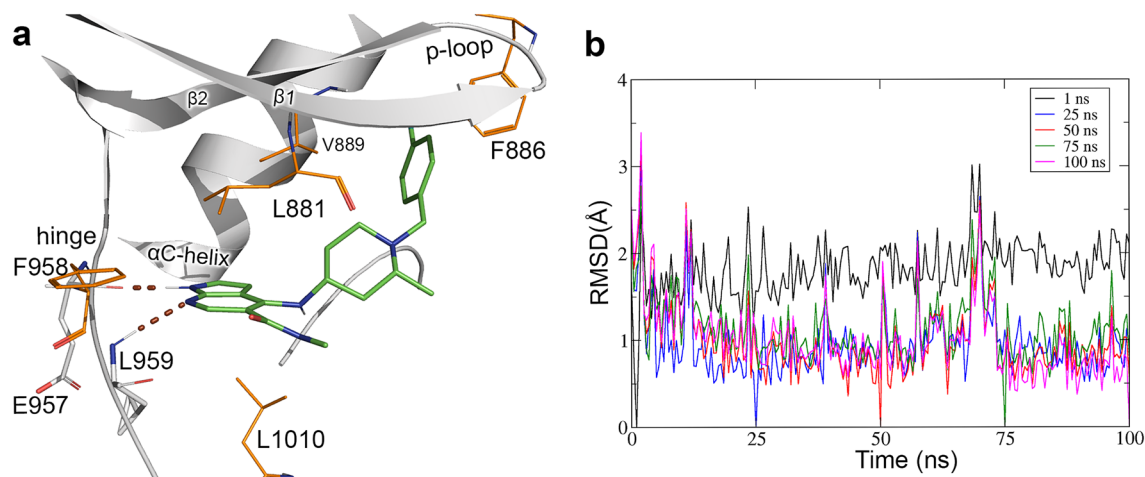


Figure 1. H-bond interactions of the most active compound (42) and JAK1 from the MD simulation. **(a)** Binding interactions between compound 42 (green) and JAK1 (grey). H-bond interactions are represented by brown dotted lines and residues forming H-bond interactions are shown in grey stick representations. Residues that formed hydrophobic interactions are shown in orange color lines. **(b)** Root mean square deviation (RMSD) of the ligand (compound 42) with respect to various snapshots (1 ns, 25 ns, 50 ns, 75 ns, and 100 ns) of the trajectory as references.

The bootstrapping method was used to evaluate the robustness of the model. The sensitivity of the model to chance correlation was tested using the progressive scrambling method (progressive scrambling Q^2). To evaluate the predictive ability of the models against the external dataset, the r_{pred}^2 , Q_{F1}^2 , Q_{F2}^2 , Q_{F3}^2 , Concordance Correlation Coefficient (CCC), mean absolute error (MAE), and r_m^2 validations were performed³⁰. A more detailed description of the validation techniques is provided in the Supplementary Material. Comparative Molecular Similarity Index Analysis (CoMSIA) models were also developed based on various combinations of electrostatic, steric, H-bond donor, and acceptor and hydrophobic descriptors, and the models were tested using the various validation techniques. Finally, based on statistical results, the models with the highest q^2 , r^2 , and r_{pred}^2 values were selected for further contour map study. We have also analyzed the applicability domain (AD) of the 3D-QSAR models to assess their applicability and reliability in predicting the activity values of the test set compounds. The Williams plots showing the standard residuals and the leverage are provided below. In Williams' plot, the standard residuals value within the range of +3 and -3 are considered acceptable and the leverage value less than $h^* = 3(p + 1)/n$ where, p is the number of descriptors and n is the number of compounds.

Results

Molecular docking. The binding of the most active compound 42 with JAK1 was carried out to study the protein–ligand interaction. The Lamarckian Genetic algorithm in Autodock 4.2 was used to generate a maximum of 100 ligand conformations. Analysis of the result was carried out by clustering the conformations of the ligand-based on root mean square deviation (RMSD) and a representative structure was selected based on nonbonded interactions and binding affinity score. The docking procedure was validated by docking the ligand (Tofacitinib) which was reported in a complex with JAK1 (PDB 3EYG) from an X-ray crystallography experiment¹⁶. The docked pose of Tofacitinib overlapped closely with that of the crystalized structure. The overlap between the docked pose and crystal pose of Tofacitinib is shown in Fig. S1a (Supplementary Material). The compound 42 was bound to JAK1 with a binding affinity score of -10.2 kcal/mol. The interactions between compound 42 and JAK1 are shown in Fig. S1b (Supplementary Materials). The pyrrolopyridine moiety of the compound formed H-bond interactions with F958 and L959 at the hinge between the N-lobe and the C-lobe domains. The methyl group of the methyl piperidine moiety extended out of the binding pocket and was exposed to the bulk solvent. The chlorobenzyl moiety folded inward into the hydrophobic pocket formed by the residues from the activation loop, α C-helix, and the P-loop. This folding of the chlorobenzyl moiety towards the hydrophobic pocket was reminiscent of the conformation observed in the crystallographic structure of the selective JAK1 inhibitor LKT (PDB ID 6SM8) reported by Su et al.³¹. The binding conformation of compound 42 selected from the docking analysis was prepared for further molecular dynamics simulation study.

MD simulation and binding energy (BE) calculation. Molecular dynamics simulation was performed to explore the binding conformations and to analyze the stability of the inhibitors in binding with JAK1. The MD simulation of the most active compound 42 was carried out for 100 ns in explicit solvent. To analyze the binding interactions, an average protein–ligand complex structure was extracted from the last 2 ns of the simulation trajectory. The binding interactions between compound 42 and JAK1 are shown in Fig. 1a. Root mean square deviation (RMSD) analysis suggested that the system converged at around 25 ns into the simulation. The RMSD of the ligand with respect to the conformations at the initial 1st ns as well as the 25th ns, 50th ns, 75th ns, and 100th ns of the trajectory are shown in Fig. 1b. During the simulation, the pyrrolopyridine of compound 42 formed H-bond interactions with F958 and L959 at the hinge region of JAK1. The chlorobenzyl end of the

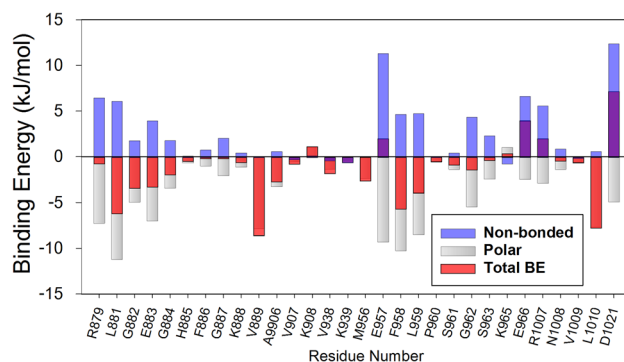


Figure 2. Residues that showed a high contribution to the total binding energy during the MD simulation of compound 42-JAK1 interaction. The energy values of non-bonded, polar, and total binding energy are shown in blue, grey, and red color respectively. Energy values are given in kJ/mol.

compound was enclosed within the hydrophobic pocket under the P-loop and formed interactions with hydrophobic residues such as F886 and V889. This conformation of compound 42 was similar to the conformation of the selective JAK1 inhibitor (LKT) reported from the X-ray crystallography experiment by Su et al.³¹. The carboxamide moiety from pyrrolopyridine extended towards the bulk solved. Hydrophobic interactions were also observed between the inhibitor and the hydrophobic residues L881, F958, and L1010.

The BE of compound 42 was evaluated to estimate its binding affinity with JAK1 using the MM-PBSA method. Fifty snapshots of the protein–ligand complex were collected from the last 5 ns of the MD simulation trajectory and were used for the BE calculation. The most active compound was bound to JAK1 with a BE value of -113.9 kJ/mol. The result of the BE calculation is shown in Table 4. The contribution of the Van der Waals, electrostatic, polar, and non-polar energy terms were -192.9 kJ/mol, -42.7 kJ/mol, -141 kJ/mol, and -19.3 kJ/mol, respectively. Energy contribution analysis for individual residues showed that the residues V889, L1010, L881, F958, L959, G882, E883, A906, M956, K979 were among the highest contributors to the total BE with each residue contributing more than -2 kJ/mol. The non-bonded energy term, polar energy term, and total binding energy of the binding site residues that made vital interactions with the most active compound 42 are shown in Fig. 2. The non-bonded energy term which includes both hydrophobic and electrostatic interactions was represented in blue color. Polar solvation energy was represented in grey color and the total binding energy was shown in red color. The Van der Waal energy term made the biggest contribution to the binding of the inhibitor. This observation from the BE calculation corroborates the interaction study which showed that compound 42 formed only two polar interactions while forming multiple hydrophobic interactions with JAK1. These results suggested that hydrophobic interaction could play an important role in the binding of the JAK1 inhibitor. For reference, we also evaluated the BE of Tofacitinib. Tofacitinib is an FDA-approved, pan JAK inhibitor that has been comprehensively studied and with structural data available in protein repository (PDB ID 3EYG). In the binding energy evaluation, Tofacitinib showed a high BE value of -126 kJ/mol suggesting a tight binding with JAK1. To verify the simulation results of the compound 42-JAK1 interaction, the simulation of compound 42 was repeated and the results were compared with the results from the first simulation. In the second simulation of compound 42 (simulation 2), the total BE was -110 kJ/mol. The analysis showed that the binding conformation of compound 42 was similar to that of the conformation observed in simulation 1. The comparison between the binding poses of compound 42 from the first and the second simulation is shown in Fig. S2a (Supplementary Material). The binding interaction of Tofacitinib with JAK1 from the MD simulation is shown in Fig. S2b (supplementary material).

3D-QSAR. For the 3D-QSAR study, the dataset compounds were sketched and minimized in Sybyl X 2.1 (Tripos, St Louis, MO, USA). The alignment of the compounds was performed by superimposing the common substructure of the compounds with that of the structure of compound 42 from the MD study. The alignment of the compounds is shown in Fig. 3a. To develop a predictive model and validate its predictive ability, we used a stratified random sampling approach to split the dataset into a test set (15 compounds) and training set (35 compounds) following a 1:2 ratio. By repeating the random sampling, we generated five different sets of the training set and test set pairs (Sets 1–5). For each of the five sets, five CoMFA models were developed using the five charge schemes namely Gasteiger, Gasteiger-Hückel, Delre, Pullman, and MMFF94. The statistical results of the CoMFA models for the different charge schemes are shown in Table S1 (Supplementary material). Finally, a final model was selected based on high internal and external predictive ability and a low optimal number of components (ONC). The CoMFA model based on the Gasteiger charge scheme showed relatively better statistical values than the models based on other charge schemes. The selected CoMFA model showed internal predictive q^2 values of 0.62 at an optimal number of component values of 3 and a non-validated r^2 value of 0.82. When evaluated against the test set compounds, the model showed an r^2_{pred} value of 0.86, suggesting that the model can predict the activity values reasonably. The statistical results of the CoMFA models are shown in Table 2.

We also developed the CoMSIA models based on the training set and charge scheme used for the CoMFA model development. Several models were derived for each combination of the CoMSIA descriptors and the statistical results were evaluated (Table S2). The CoMSIA model based on hydrophobic descriptor with q^2 , r^2 , and ONC values of 0.52, 0.82, and 3 was selected since this model gave the highest external predictive value of 0.69. The statistical results of the CoMSIA models are shown in Table 2. The scatter plot between the actual and the

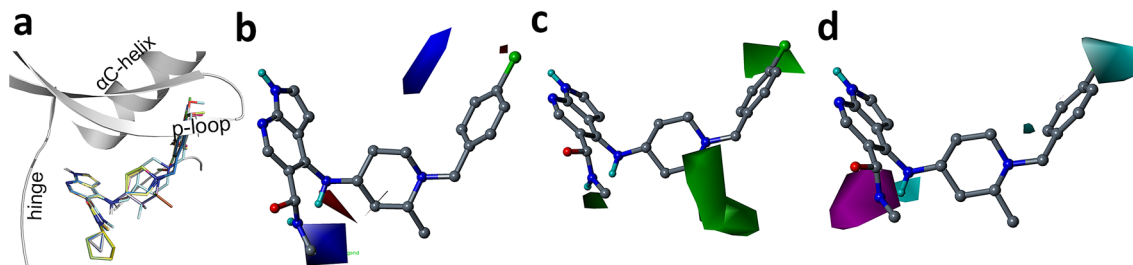


Figure 3. Alignment and contour maps from the CoMFA and CoMSIA models. (a) Alignment of the compounds inside the receptor. (b) Electrostatic contour map. Blue contour represents region favorable for electropositive substituents. (c) Steric contour map. Green contour represents a region favorable for bulky substituents. (d) Hydrophobic contour. Cyan contour represents regions favorable for hydrophobic substituents whereas, magenta color represent region favorable for non-hydrophobic substituents.

Parameters	CoMFA	CoMSIA
q^2	0.62	0.52
ONC	3	3
SEP	0.45	0.52
r^2	0.82	0.74
SEE	0.66	0.53
F value	48.39	30.08
BS r^2	0.84	0.86
BS SD	0.22	0.30
Progressive scrambling (Q^2)	0.54	0.52
r^2_{pred}	0.70	0.69
Q^2_{F1}	0.69	0.69
Q^2_{F2}	0.67	0.68
Q^2_{F3}	0.68	0.69
CCC	0.94	0.59
MAE	0.25	0.25
r^2_m	0.79	0.83
Influence of different fields (%)		
S	43	–
E	57	–
H	–	100

Table 2. Statistical results of the CoMFA and CoMSIA models. q^2 : cross-validated correlation coefficient; ONC: Optimal number of components; SEP: Standard Error of Prediction; r^2 : non-cross-validated correlation coefficient; SEE: Standard Error of Estimation; F value: F-test value; r^2 ; BS- r^2 : Bootstrapping r^2 mean; BS-SD: Bootstrapping Standard deviation; r^2_{pred} : predictive correlation coefficient; S: Steric; E: Electrostatic; CCC: Concordance correlation coefficient; More details on Q^2_{F1} , Q^2_{F2} , Q^2_{F3} , MAE and r^2_m are given in Supplementary Material.

predicted activity values from the CoMFA and CoMSIA models is shown in Fig. S3 (Supplementary Material). The analysis showed that the CoMFA and CoMSIA models were able to predict the activities of inhibitors with reasonable accuracy.

The CoMFA model showed a Q^2 value of 0.54 while the CoMSIA model showed a slightly lower value of 0.52 in progressive scrambling. The predictive ability of the selected CoMFA and CoMSIA models were tested using various external validation techniques¹ and the results are provided in Table 2. The CoMFA model showed Q^2_{F1} , Q^2_{F2} , Q^2_{F3} , (CCC), MAE, and r^2_m of 0.69, 0.67, 0.68, 0.94, 0.25 and 0.79 respectively. Similarly, The CoMSIA model showed Q^2_{F1} , Q^2_{F2} , Q^2_{F3} , (CCC), MAE, and r^2_m of 0.69, 0.68, 0.69, 0.59, 0.25 and 0.83 respectively. These observations suggested that the CoMFA and CoMSIA models have reasonable predictive ability and reliability.

The williams plots for the CoMFA and CoMSIA models are provided in Fig. S6 (Supplementary materials). The predictions in both the models showed standard residual values between +3 and –3 and leverage values less

than the threshold value of 0.25. These results indicated that the test set predictions are within the AD of the CoMFA and CoMSIA models. Hence, these predictions may be considered reliable.

Contour map analysis. The contour map analysis of the CoMFA and CoMSIA maps was carried out to study the favorable and unfavorable sites for various chemical substitutions. The electrostatic and steric contour maps from the CoMFA model and the hydrophobic contour map from the CoMSIA model are shown in Fig. 3. Blue contours in the electrostatic map represent regions favorable to positive substitutions. The green contours in the steric contour map indicate regions favorable for bulky substitutions. Cyan and magenta color contours in the hydrophobic map represent hydrophobic favorable and hydrophilic favorable regions.

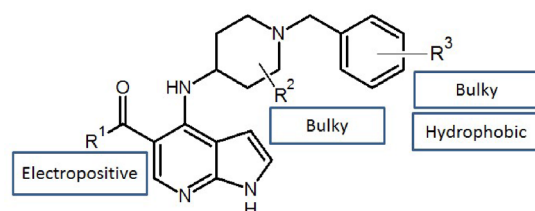
Blue contours near the carboxamide and also near the meta-position of the chlorobenzyl suggested that electropositive substituents in those positions could enhance the inhibitory activity against JAK1 (Fig. 3b). From the analysis, we observed that bulky substitutions at the chlorobenzyl moiety and near the methyl group of the methyl piperidine could improve the inhibitory activity against JAK1 (Fig. 3c). Hydrophobic substitutions near the chlorobenzyl could enhance the activity against JAK1 whereas non-hydrophobic substituents are favored near the carboxamide (Fig. 3d). Taken together, these contour maps provide an overview of the regions favorable for various chemical substitutions.

Designing new compounds. Following the contour map analysis, an inhibitor design scheme was developed as shown in Table 3. Up to 150 compounds were designed by adding substituents based on the design scheme. The inhibitory activities of the designed compounds were predicted using the selected CoMFA and CoMSIA models and seventeen compounds that showed a higher predicted activity value than the most active compound (42) in the dataset were further analyzed. The structure and the predicted activity values of the seventeen designed compounds are given in Table 3.

The interaction study from MD simulation and the contour maps were co-analyzed to explore the structural features necessary to improve the activity values against JAK1. Some of the compound fragments which were used as substituents were collected from published literature while some were taken from the in-house compound library. We performed the designing of new compounds by building upon the most active compound 42 as shown in Table 3. As a first step, we substituted the R¹ and R² with electropositive and bulky substituents while keeping the Cl substituent at the R³ position unchanged. The 3D-QSAR prediction showed that the electropositive and bulky substituents at R¹ and R² improved the activity values as observed in compounds D01 and D05. Next, we also examined the R³ position by adding bulky and hydrophobic substituents with three or four carbon atoms. The prediction showed that the substitution of bulky and hydrophobic substituents increased the activity values as seen in D129, D135, D62, etc. When we substituted the R³ position with 3-NH₂ or 4-NH₂, we found the activity values of the compounds increased considerably higher than the compounds with methyl or ethyl groups as seen in compounds D108, D128, and D129. In D70 and D72, we have also substituted the Cl atom of the compound D01 and D05 with other heavy atoms such as N, and F atoms which increased the activity values. We also substituted the R³ position with aromatic rings such as five and six-membered rings, which showed a decrease in the activity value. Similarly, aromatic substituents in R¹ and R² positions also lead to a decrease in activity values. Hence we deduce that, though bulky substituents are favorable at R¹, R², and R³, very large substituents are not favorable. The SMILES structures of the 150 compounds were provided in Table S5 (supplementary material).

Molecular dynamics simulation and BE calculation of the seventeen designed compounds were carried out to study the binding interactions and the binding affinity with JAK1. The binding energy values are shown in Table 4. The binding interactions of compounds D01, D07, D64, D108, D127, and D135 showed higher binding affinity than the reference compound Tofacitinib (−126 kJ/mol) and are further analyzed and shown in Fig. 4. Of the selected design compounds, compound D127 showed the highest binding affinity with a binding energy value of −137 kJ/mol (Fig. 4e).

The compound D01 (Fig. 4a), like compound 42, has a chlorine atom at the benzyl moiety and formed a conformation similar to that of compound 42. The pyrrolopyridine moiety formed H-bond interactions with E957 and L959. The tert-butyl moiety from the piperidine extended out into the bulk solvent while the chlorobenzyl moiety folded into the hydrophobic pocket. The compound D07 (Fig. 4b), formed H-bond with the E957 and L959 at the hinge region while, the benzyl moiety extended into the hydrophobic pocket, similar to compound 42. Compound D07 which has a methyl group at the benzyl moiety also exhibited a similar binding pattern with 42 and folded in-ward into the hydrophobic pocket. Compound D64 also formed H-bond interactions with the hinge at E957 and L959. However, the dimethyl benzyl moiety extended out of the hydrophobic pocket and formed non-bonded interactions with K956, R1007, and N1008 (Fig. 4c). The compound D108, which possessed an amine group at the benzyl moiety, also folded toward the hydrophobic pocket. D108 formed three H-bond interactions at the hinge with E957 and L959. The amine group at the benzyl formed H-bond interaction with N1008 at the activation loop and D1021 at the DFG motif. The methyl group at the piperidine extended up and out towards the glycine-rich loop (Fig. 4d). Interaction studies of the D127-JAK1 interaction in Fig. 4e showed that the pyrrolopyridine of D127, like compound 42, was able to form H-bond interaction at the hinge with E957 and L959. However, the diethyl benzyl moiety of compound D127 extended outward of the hydrophobic pocket and formed non-bonded interactions with E966 and K965 of the α D-helix and R007 of the activation loop. The H-bond interactions with the hinge residues were also observed in all the designed compounds. The compound D135 which showed the third-highest binding affinity with a BE value of −126 kJ/mol showed an additional H-bond interaction with Leu959 at the hinge (Fig. 4f). In addition to the non-bonded interactions at the α D-helix, the compound D135 also interacted with the R1007 via a cation- π interaction. The energy values of the individual residues that make significant contributions are given in Table S4 (Supplementary Materials). The interactions of the compounds D05, D16, D62, D63, D66, D70, D72, D118, D126, D128, and D129 showed



Compounds	R1	R2	R3	Predicted activity (pIC ₅₀)	
				CoMFA	CoMSIA
Compound D01			4-Cl	8.7	9.7
Compound D05			4-Cl	8.7	9.8
Compound D07			4-Me	8.4	9.3
Compound D16			3-Me -CH ₂ CH ₂ CH ₃	8.1	8.9
Compound D62			3-Me 5-Me	8.1	9
Compound D63			2-Me 5-Me	8.3	9.1
Compound D64			2-Me 5-Me	8.3	9.1
Compound D66			3- 5-OH	8.4	9.2
Compound D70		-Me	4-N	8.1	9.4
Compound D72			4-F	8.9	9.7
Compound D108		-Me	4-NH ₂	8.9	10.5
Compound D118	-NH ₂		3-	8.5	9.4
Compound D126			3-Me 4-Me	8.3	9.2
Compound D127			3-Eth 4-Eth	8.2	9.2
Compound D128			4-	8.3	9.9
Compound D129			3-NH ₂ 3-NH ₂	8.3	9.9
Compound D135	-NH ₂		3- 4-Me	8.2	9.3

Table 3. Structure of the designed compounds along with the predicted pIC₅₀ values.

Compounds	Van der Waals (kJ/mol)	Electrostatics (kJ/mol)	Polar solvation (kJ/mol)	SASA (kJ/mol)	Total binding energy (kJ/mol)
Compound 42 (most active compound)	-192.9	-42.7	141.0	-19.3	-113.9
Tofacitinib	-205	-74.2	170.5	-17.3	-126
Designed compounds					
Compound D01	-230.4	-51.1	177.4	-23.6	-127.2
Compound D05	-210.8	-40.9	150.8	-21.3	-122.2
Compound D07	-225.2	-40.9	162.2	-22.0	-125.9
Compound D16	-217.3	-47.8	162.5	-21.0	-124.2
Compound D62	-184.5	-46.3	135.1	-20.0	-115.7
Compound D63	-203.5	-45.7	151.3	-20.4	-118.2
Compound D64	-202.2	-50.3	145.4	-19.5	-126.6
Compound D66	-209.2	-78.0	193.9	-22.3	-115.6
Compound D70	-232.7	-55.9	188.0	-23.3	-123.8
Compound D72	-198.7	-53.5	148.9	-19.2	-122.5
Compound D108	-214.0	-69.8	175.7	-21.3	-129.4
Compound D118	-201.6	-38.5	147.2	-22.1	-115.0
Compound D126	-205.8	-54.8	160.4	-20.8	-120.9
Compound D127	-200.2	-43.1	127.0	-20.9	-137.2
Compound D128	-224.0	-55.4	185.7	-23.8	-117.5
Compound D129	-206.1	-51.3	166.3	-21.9	-113.1
Compound D135	-218.1	-50.9	164.2	-21.5	-126.3

Table 4. The energy contribution of the various energetic terms (van der Waals energy, electrostatic energy, polar solvation energy, and non-polar solvation energy/SASA) to the total binding energy. The energy values are given in kJ/mol.

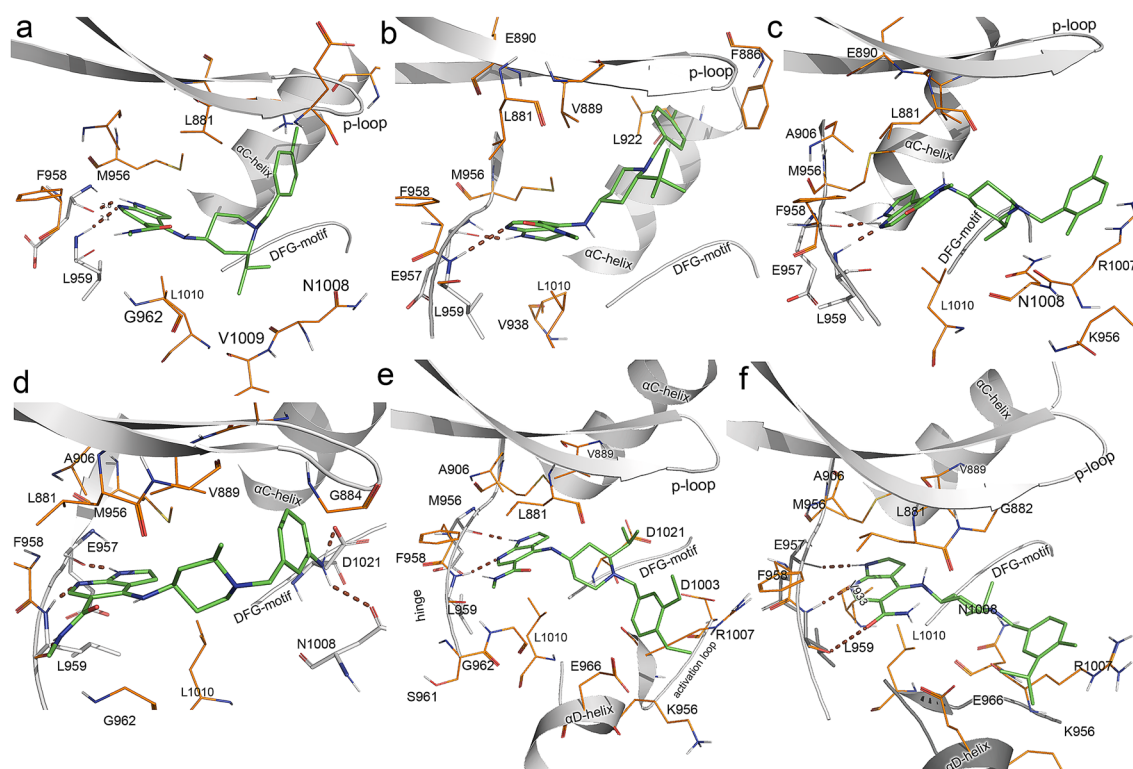


Figure 4. H-bond interactions of the designed compounds with JAK1 from the MD simulation. (a) D01-JAK1 interactions, (b) D07-JAK1 interactions, (c) D64-JAK1 interactions, (d) D108-JAK1 interactions, (e) D127-JAK1 interactions, (f) D135-JAK1 interactions. The inhibitors are shown in green color stick representations. H-bond interactions are represented by brown dotted lines and residues forming H-bonds are shown in grey stick representations. Residues that formed hydrophobic interactions are shown in orange color lines.

binding energy values lower than -113 kJ/mol are given in Fig. S4 (Supplementary Materials). The ligand and protein RMSD for all the compounds are given in Fig. S5 (Supplementary Materials).

Discussion

The dataset compounds shared a common substructure consisting of a fused pyrrolopyridine and a benzyl moiety which are linked together by either piperidine, a pyrrolidine, an azabicyclo[3.2.1]octane, or an azepane ring. The pyrrolopyridine moiety, as seen in the most active compound 42, was anchored at the hinge while the piperidine was bound between the activation loop and glycine-rich loop through hydrophobic interactions. The benzyl end (chlorobenzyl in compound 42) of the compound took an inward turn and occupied the hydrophobic pocket formed by residues from the P-loop, the α C-helix, and the activation loop.

The conformational analysis of compounds D01, D07, D64, D108, D127, and D135 showed that these designed compounds were able to bind to the hinge via H-bond interactions with the E957 and L959. This interaction with the hinge is considered to be important for anchoring the inhibitors inside the kinase^{32,33}. However, the benzyl end of the designed compounds adopted either a fold-in conformation, or a fold-out conformation. The designed compounds D01 and D05 adopted the inward fold-in conformation in which the benzyl moiety occupied the hydrophobic pocket, similar to the observation in compound 42 (most active compound). On the other hand, the compounds D64, D127, and D135 adopted the fold-out conformation and extended the substituted benzyl end toward the bulk solvent. This fold-out conformation is reminiscent of the conformation of the selective inhibitor EYQ–JAK1 interaction (PDB ID 6GGH). This conformation contradicts the observation made in compound 42–JAK1 interaction (Fig. 1a) in which the chlorobenzyl folded into the hydrophobic pocket. The reason behind the difference in the conformation of the designed compounds could not be verified. However, further analysis of compounds D64, D108, and D135 showed that all these compounds possessed bulky substituents at the benzyl ring and adopted the fold-out conformation. The bulky substituents at the benzyl may cause steric clash at the hydrophobic pocket in a fold-in conformation, while the fold-out conformation may allow non-bonded interactions with residues from the α D-helix and the activation loop. These observations suggested that compounds with small substituents tend to occupy the hydrophobic pocket whereas the compounds with bulky substituents tend to extend outward. Earlier computational studies by our group and also by other experimental researchers have shown that, while the interactions with the hinge of kinases are stable, the interactions at the hydrophobic pocket are more flexible^{21,34–36}. This flexibility allows for variation in inhibitor binding and is considered to be a mechanism for selective inhibition. The high binding affinity with JAK1 and their unique binding conformation make these compounds interesting potential candidates for JAK1 inhibition. Further assessment through experimental studies is essential to verify the potential of the designed compounds.

Taken together, the protein–ligand interaction studies suggested that these pyrrolopyridine based JAK1 inhibitors can adopt two distinct conformations. The structure–activity analysis revealed vital structural properties to improve inhibitory activity against JAK1. At least six of the compounds showed a higher binding affinity against JAK1 as compared to the approved JAK1 inhibitor Tofacitinib. The outcome of this study could provide vital information for the development of more potent JAK1 inhibitors.

Data availability

All relevant data are contained within the manuscript and the supplementary material. Additional raw data will be available upon request.

Received: 4 September 2021; Accepted: 10 November 2021

Published online: 29 November 2021

References

- Shuai, K. & Liu, B. Regulation of JAK–STAT signalling in the immune system. *Nat. Rev. Immunol.* **3**(11), 900–911 (2003).
- Gadina, M. *et al.* Signaling by type I and II cytokine receptors: Ten years after. *Curr. Opin. Immunol.* **13**(3), 363–373 (2001).
- Levy, D. E. & Darnell, J. Stats: Transcriptional control and biological impact. *Nat. Rev. Mol. Cell Biol.* **3**(9), 651–662 (2002).
- Schwartz, D. M. *et al.* JAK inhibition as a therapeutic strategy for immune and inflammatory diseases. *Nat. Rev. Drug Discovery* **16**(12), 843–862 (2017).
- O’Shea, J. J. *et al.* The JAK–STAT pathway: Impact on human disease and therapeutic intervention. *Annu. Rev. Med.* **66**, 311–328 (2015).
- Choy, E. H. *et al.* The effect of JAK1/JAK2 inhibition in rheumatoid arthritis: Efficacy and safety of baricitinib. *Clin. Exp. Rheumatol.* **37**(4), 407–694 (2019).
- Wu, G., Yang, Y., Liu, M., Wang, Y. & Guo, Q. Systematic review and network meta-analysis: Comparative efficacy and safety of biosimilars, biologics and JAK1 inhibitors for active Crohn disease. *Front. Pharmacol.* **12**, 604 (2021).
- Singh, J. A. Filgotinib, a JAK1 inhibitor, for treatment-resistant rheumatoid arthritis. *JAMA* **322**(4), 309–311 (2019).
- Boyle, D. *et al.* The JAK inhibitor tofacitinib suppresses synovial JAK1–STAT signalling in rheumatoid arthritis. *Ann. Rheum. Dis.* **74**(6), 1311–1316 (2015).
- Takeuchi, T. *et al.* Efficacy and safety of the oral Janus kinase inhibitor peficitinib (ASP015K) monotherapy in patients with moderate to severe rheumatoid arthritis in Japan: A 12-week, randomised, double-blind, placebo-controlled phase IIb study. *Ann. Rheum. Dis.* **75**(6), 1057–1064 (2016).
- Dhillon, S. Tofacitinib: A review in rheumatoid arthritis. *Drugs* **77**(18), 1987–2001 (2017).
- Kavanaugh, A. *et al.* Filgotinib (GLPG0634/GS-6034), an oral selective JAK1 inhibitor, is effective as monotherapy in patients with active rheumatoid arthritis: Results from a randomised, dose-finding study (DARWIN 2). *Ann. Rheum. Dis.* **76**(6), 1009–1019 (2017).
- Markham, A. & Keam, S. J. Peficitinib: First global approval. *Drugs* **79**(8), 887–891 (2019).
- Roskoski, R. Jr. Janus kinase (JAK) inhibitors in the treatment of inflammatory and neoplastic diseases. *Pharmacol. Res.* **111**, 784–803 (2016).
- Park, E. *et al.* Discovery and biological evaluation of N-methyl-pyrrolo [2, 3-b] pyridine-5-carboxamide derivatives as JAK1-selective inhibitors. *J. Med. Chem.* **64**(2), 958–979 (2021).

16. Williams, N. K. *et al.* Dissecting specificity in the Janus kinases: The structures of JAK-specific inhibitors complexed to the JAK1 and JAK2 protein tyrosine kinase domains. *J. Mol. Biol.* **387**(1), 219–232 (2009).
17. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**(16), 2785–2791 (2009).
18. Van Der Spoel, D. *et al.* GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **26**(16), 1701–1718 (2005).
19. Da Silva, A. W. S. & Vranken, W. F. ACPYPE-Antechamber python parser interface. *BMC. Res. Notes* **5**(1), 1–8 (2012).
20. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**(9), 1157–1174 (2004).
21. Keretsu, S., Bhujbal, S. P. & Cho, S. J. Computational study of paroxetine-like inhibitors reveals new molecular insight to inhibit GRK2 with selectivity over ROCK1. *Sci. Rep.* **9**(1), 1–14 (2019).
22. Keretsu, S., Bhujbal, S. P. & Cho, S. J. Rational approach toward COVID-19 main protease inhibitors via molecular docking, molecular dynamics simulation and free energy calculation. *Sci. Rep.* **10**(1), 1–14 (2020).
23. Keretsu, S., Bhujbal, S. P. & Cho, S. J. Molecular modeling studies of pyrrolo [2, 3-d] pyrimidin-4-amine derivatives as JAK1 inhibitors based on 3D-QSAR, molecular docking, molecular dynamics (MD) and MM-PBSA calculations. *J. Biomol. Struct. Dyn.* **39**(3), 753–765 (2021).
24. Keretsu, S., Ghosh, S. & Cho, S. J. Molecular modeling study of c-KIT/PDGFR α dual inhibitors for the treatment of gastrointestinal stromal tumors. *Int. J. Mol. Sci.* **21**(21), 8232 (2020).
25. Kumari, R., Kumar, R., Open Source Drug Discovery Consortium, Lynn, A. g_mmpbsa—A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* **54**(7), 1951–1962 (2014).
26. Halder, A. K. & Cordeiro, M. AKT inhibitors: The road ahead to computational modeling-guided discovery. *Int. J. Mol. Sci.* **22**(8), 3944 (2021).
27. Miller, B. R. III. *et al.* MMPBSA.py: An efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* **8**(9), 3314–3321 (2012).
28. Cramer, R. D., Patterson, D. E. & Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **110**(18), 5959–5967 (1988).
29. Klebe, G. Comparative molecular similarity indices analysis: CoMSIA. *3D QSAR Drug Des.* **3**, 87–104 (1998).
30. Gramatica, P. & Sangion, A. A historical excursus on the statistical validation parameters for QSAR models: A clarification concerning metrics and terminology. *J. Chem. Inf. Model.* **56**(6), 1127–1131 (2016).
31. Su, Q. *et al.* Discovery of (2R)-N-[3-[2-[(3-Methoxy-1-methyl-pyrazol-4-yl) amino] pyrimidin-4-yl]-1H-indol-7-yl]-2-(4-methylpiperazin-1-yl) propanamide (AZD4205) as a Potent and Selective Janus Kinase 1 Inhibitor. *J. Med. Chem.* **63**(9), 4517–4527 (2020).
32. Xing, L., Klug-Mcleod, J., Rai, B. & Lunney, E. A. Kinase hinge binding scaffolds and their hydrogen bond patterns. *Bioorg. Med. Chem.* **23**(19), 6520–6527 (2015).
33. Del Rio, A., Baldi, B. F. & Rastelli, G. Activity prediction and structural insights of extracellular signal-regulated kinase 2 inhibitors with molecular dynamics simulations. *Chem. Biol. Drug Des.* **74**(6), 630–635 (2009).
34. Keretsu, S., Bhujbal, S. P. & Cho, S. J. Docking and 3D-QSAR studies of hydrazone and triazole derivatives for selective inhibition of GRK2 over ROCK2. *Lett. Drug Des. Discovery* **17**(5), 618–632 (2020).
35. Cheetham, G., Charlton, P., Golec, J. & Pollard, J. Structural basis for potent inhibition of the Aurora kinases and a T315I multi-drug resistant mutant form of Abl kinase by VX-680. *Cancer Lett.* **251**(2), 323–329 (2007).
36. Treiber, D. K. & Shah, N. P. Ins and outs of kinase DFG motifs. *Chem. Biol.* **20**(6), 745–746 (2013).

Acknowledgements

This study was supported by research funds from Chosun University 2021.

Author contributions

S.K. designed the study and completed the molecular modeling under the supervision of S.J.C. The analysis and interpretation of the results were done by S.K. The manuscript was written by S.K and S.G. All authors read and approved the manuscript for publication.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-02364-2>.

Correspondence and requests for materials should be addressed to S.J.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021