

Molecular BioSystems

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



www.rsc.org/molecularbiosystems

Jak2 inhibitor – a jackpot for pharmaceutical industries: a comprehensive computational method in the discovery of new potent Jak2 inhibitors

Kh. Dhanachandra Singh, Queen Naveena and Muthusamy Karthikeyan*

Department of Bioinformatics, Alagappa University, Karaikudi – 630 004, Tamil Nadu, India

****Corresponding author:***

Dr. M. Karthikeyan, Ph.D

Phone (off) +914565-230725,

Fax (off):+914565-225202

email: mkbioinformatics@gmail.com

Abstract

A potent Jak2 inhibitor could solve numerous diseases which includes hypertension and cardiovascular diseases, myeloproliferative neoplasms, polycythemia vera, essential thrombocythemia, primiray myelofibrosis, psoriasis and rheumatoid arthritis. So, identifying a potent Jak2 inhibitors are of great interest to the researchers and pharmaceutical companies. Virtual screening and molecular docking are important tools for structure based drug discovery but selecting an appropriate method to calculate the electrostatic potential is critical. In this study, four semi empirical (AM1, RM1, PM3, and MNDO) and two empirical (DFT, HF) charges were investigated for their performance on the prediction of docking pose using Glide XP. The result shows that AM1 has the best charge model for our study. Further, we performed 3D-QSAR study of 76 decaene derivatives. Since 3D-QSAR methods are known to be highly sensitive to ligand conformation and alignment method, we did comparative 3D-QSAR study of AM1 charge docked pose alignment based QSAR (structure based) and pharmacophore based QSAR. We found a better QSAR model in structure based method. Hence, the result clearly demonstrate that selecting an appropriate method to calculate the electrostatic potential for docking studies and a good alignment of ligand for 3D-QSAR is critical. Finally, extensive pharmacophore and e-pharmacophore based virtual screening followed by subsequent docking studies identifying 27 lead molecules which could be a potent Jak2 inhibitor.

Key words: Jak2, 3D-QSAR, AM1, Docking, Pharmacophore

Introduction

Jak kinases are a family of intracellular nonreceptor tyrosine kinases which includes Jak1, Jak2, Jak3 and Tyk2 and that are transduce cytokine-mediated signals *via* the JAK-STAT pathway.¹ It is playing a significant role in myeloid cell development, proliferation and survival, as well as for the initial stages of the immune response.² Jak2 causes effects on signaling process and it is associated with single chain hormone receptors, the common β chain family and certain members of the class II receptor cytokine family for signaling process.³ Recently, Guilluy and colleagues explained that the Jak2 is involved in the Ang II-mediated activation of the Rho exchange factor, Arhgf1, leads to enhanced vasoconstriction.⁴ Reactive oxygen species (ROS) signaling pathway is implicated in hypertension and vascular pathology^{5,6} and Ang II is the mediator of oxidative stress and oxidant signaling.⁷⁻¹⁰ The Jak2 is essential for mediating ROS dependent vascular smooth muscle cells (VSMC) proliferation¹¹ and their activation results in higher levels of ROS and Jak2 inhibition leads to a dramatic reduction in oxidative stress.¹² The Jak2-V617F causes constitutive activation of Jak2 and thus increases the levels of ROS within cells and also the inhibition of Jak2 leads to reduction of ROS in these same cells.^{12,13}

Jak2 inhibitors could be an important drug for various diseases like myeloproliferative neoplasms, polycythemia vera, essential thrombocythemia, primiray myelofibrosis¹⁴, psoriasis¹⁵⁻¹⁷, rheumatoid arthritis¹⁸, hypertension and cardiovascular diseases.¹⁹ So, finding a potential inhibitor for Jak2 could be a Jackpot for pharmaceutical industries since it can be used in the treatment of several diseases. Numerous Jak2 inhibitors are in phase I/II clinical studies at present.¹⁴ Recent reports were also observed that structure activity relationship (SAR) study using crystal structure of protein-ligand complex could improve the activity and ADME (Absorption, Distribution, Metabolism, and Excretion) of the compounds.^{20,21} The present study

is aimed to explore the SAR of decaene derivatives through ligand based and structured based approach (protein-ligand complex using molecular docking).

In the present study, we used decaene derivatives as multikinase inhibitor for 3D-QSAR studies. In this study, we focus on the effect of six different electrostatic potential on Jak2 inhibitors. Several charge calculation methods are available but their fundamental differences in algorithms can result in significant differences in the electrostatic assignments for various atoms. It should be noted that the charge models could have effect on not only the docking energy/scores, but also the docking conformations, and thus could interfere with the accuracy in docking.²² Therefore, we compare six different charge calculation methods to identify the accuracy of charge calculation methods among 15 protein-ligand complexes. The ligands were extracted from the complexes and then docked with their host protein using Glide, Schrodinger, LLC. Then, docking accuracy was evaluated mainly considering the ability to find the correct conformations as well as to accurately estimate binding energy. Further, charge were calculated for all the 76 decaene derivatives^{23, 24} and docked with the Jak2 protein. The conformations from the docked pose were used for structure based 3D-QSAR generation. The un-charged compounds were used for pharmacophore based 3D-QSAR generation. Finally, extensive virtual screening was performed against zinc database.

Materials and methods

Charge calculation methods

Jaguar version 7.8 was used to assign the charges for Jak2 inhibitors.^{25, 26} Electrostatic point charges on the atoms of Jak2 inhibitors can be calculated by multiple semiempirical and empirical methods. In this study, the semiempirical and empirical charge calculation methods were used to calculate the point charges of the compounds. We used 15 ligand bound proteins crystal structures for charge calculation and docking studies. These 15 ligand bound proteins

crystal structures and their active site residues for Jak2 inhibition were obtained from PDB (Protein Data Bank). Ligands were extracted from the complex structures and then charges of the ligands have been assigned using semiempirical (AM1, RM1, PM3, and MNDO) and empirical (DFT, HF) charge calculation methods. These charged ligands were taken to perform docking with Jak2 protein to find correct conformation as well as to estimate accurate binding energy and accuracy of charge calculation methods.

Dataset

The multi-kinase inhibitors, decaene derivatives were used to understand the mechanistic insight into the inhibitory activity of the compounds. The 76 compounds were retrieved from the literature.^{23,24} These 76 compounds were sketched in maestro, Schrodinger, LLC²⁷ and inhibitory activity (IC₅₀) for all the compounds were converted into pIC₅₀ (-logIC₅₀) values (Table1-7).

Structure based 3D-QSAR

In QSAR studies, a bioactive conformation of the compound is required for accurate calculation of 3D descriptors. Hence, the 76 compounds were subjected to energy minimization in ligprep module using the OPLS 2005 force field.²⁸ The prepared compounds have been assigned their charges using AM1 charge calculation method and these charged compounds were allowed to perform docking with Jak2 protein (PDB ID 3E64). The best docked conformation of the ligands were used for direct structure based 3D-QSAR using Phase version 3.4.²⁹⁻³¹

Pharmacophore based 3D-QSAR

In pharmacophore based 3D-QSAR, the QSAR models were generated based on pharmacophoric features such as hydrogen-bond acceptor (A), hydrogen-bond donor (D), aromatic ring (R), positively charged groups (P) and negatively charged groups (N) to predict common pharmacophore hypothesis. The pharmacophoric sites were created by using following

steps. The ligands were divided as active and inactive by giving proper activity threshold value based on dataset activity threshold value (5.0-8.15). The threshold value was 7 for active and 6.5 for inactive ligand. Ligands were chosen to derive a set of suitable common pharmacophore and QSAR model building. Details of the methodology are already been explained in our previous papers.^{32, 33}

Building 3D-QSAR models

The 3D-QSAR models were generated using the selected hypothesis by dividing the dataset into training set (70%) and test set (30%) in random manner. Here, 51 compounds were selected as training set and 25 were selected as test set. The training set and test are same for structure based QSAR and Pharmacophore based QSAR. The regression is done by constructing a series of models with 7 PLS factors.

Validation of QSAR model

External validation has been carried out to check the robustness of the model and also to evaluate the true predictive abilities of the established model. This includes the statistical measure of significance including (R^2), leave-n-out (LNO), cross-validation coefficient (Q^2), least squares error (LSE), and lack of-fit measure (LOF) which was developed by Friedman. According to literatures,³⁴⁻³⁶ 3D-QSAR models are accepted if they satisfy all of the following conditions:

$$r_{cv}^2 > 0.5, r^2 > 0.6, [(r^2 - r_o^2)/r^2] < 0.1, 0.85 \leq k \leq 1.15 \text{ and } r_m^2 > 0.5 \quad (1)$$

The r^2 value calculated by the following formula,

$$R = \frac{\sum (y_i - \bar{y}_o)(\tilde{y}_i - \bar{y}_p)}{\sqrt{\sum (y_i - \bar{y}_o)^2 \sum (\tilde{y}_i - \bar{y}_p)^2}} \quad (2)$$

In these equations, y_i and \tilde{y}_i are the observed and predicted activity, \bar{y}_o and \bar{y}_p are the average values of the observed and predicted pIC₅₀ values of the test set molecules. For the ideal QSAR model, the r^2 value should be close to 1. Meanwhile the regression of y against \bar{y} through origin: $y_i^{ro} = k\tilde{y}_i$ should be characterized by k close to 1. Slope k is calculated as follow:

$$k = \frac{\sum y_i \tilde{y}_i}{\sum \tilde{y}_i^2} \quad (3)$$

Another essential parameter c was defined as follow:

$$r_m^2 = r^2 (1 - \sqrt{|r^2 - r_o^2|}) \quad (4)$$

Where, the r^2 was the non-cross-validated correlation coefficient obtained from the PLS process, and the r_o^2 was calculated as follows:

$$r_o^2 = 1 - \frac{\sum (\tilde{y}_i - y_i^{ro})^2}{\sum (\tilde{y}_i - \bar{y}_p)^2} \quad (5)$$

The y_i^{ro} was obtained by this formula,

$$y_i^{ro} = k\tilde{y}_i \quad (6)$$

The cross validated co-efficient, r_{cv}^2 , was calculated using the following equation:

$$r_{cv}^2 = 1 - \frac{\sum (Y_{predicted} - Y_{observed})^2}{\sum (Y_{observed} - Y_{mean})^2} \quad (8)$$

Where, $Y_{predicted}$, $Y_{observed}$, and Y_{mean} are the predicted, observed, and mean values of the target property (pIC₅₀), respectively.

The best validated pharmacophore model was used to screen against zinc database. The crystal structure of Jak2 (PDB id: 3E64) was selected for docking studies based on binding

affinity. The Jak2 protein was prepared in protein preparation wizard and docking was performed against the receptor using the same set of ligands in Schrodinger, LLC, New York, USA. We have performed three different level of docking precision including High Throughput Virtual Screening (HTVS), Standard Precision (SP) and Extra Precision (XP).³⁷⁻⁴⁰ First, we carried out our calculations in HTVS, then in SP and XP mode for further refinement of good ligand pose.

Prime/MM-GBSA is used to predict the free energy of binding for set of ligands to receptor.⁴¹⁻⁴³ The binding free energy (ΔG_{bind}) is then estimated using equation:

$$\Delta G_{\text{bind}} = E_{\text{R:L}} - (E_{\text{R}} + E_{\text{L}}) + \Delta G_{\text{solv}} + \Delta G_{\text{SA}} \quad (7)$$

where $E_{\text{R:L}}$ is energy of the complex, $E_{\text{R}} + E_{\text{L}}$ is sum of the energies of the ligand and apo protein, using the OPLS-2005 force field, ΔG_{solv} (ΔG_{SA}) is the difference between GBSA solvation energy (surface area energy) of complex and sum of the corresponding energies for the ligand and apo-protein.

Pharmacophore based virtual screening

Pharmacophore models containing five sites were generated using 76 decaene derivatives, which were reported as potent Jak2 inhibitors. The common hypothetical pharmacophores with three and four features were rejected for the study based on the molecule occupancy of the pharmacophoric features. The best two hypothetical pharmacophores, ADHRR and AAHRR, were selected based on the scoring function for further pharmacophore-based screening (**Supplementary Figure 1**). The selected hypothetical pharmacophore consists of one hydrogen bond acceptors (A), one donor (D), one hydrophobic group and two aromatic rings (R) for the first hypothesis, and the second hypothetical pharmacophore consists of two hydrogen bond acceptors (A), one hydrophobic group (H) and two aromatic rings (R).

The validated pharmacophore model was used to search against a 3D database for structures that match the pharmacophoric features of the model. Virtual screening was carried out using Phase in Schrodinger software that uses the pharmacophore to efficiently search the ZINC database of fixed conformers for pharmacophore matches. A molecule which fits well with the pharmacophoric features of the ADHRR and AAHRR hypothesis was retrieved as a hit.

E-Pharmacophores were generated by mapping the energetic terms from Glide XP scoring function onto atom centres. The docking pose of protein-ligand complexes were given as input for generating pharmacophore sites. The constructed e-pharmacophore was used as a query for virtual screening.^{44, 45} Phase module of Schrodinger, LLC²⁹⁻³¹ was used for generation of pharmacophore sites such as hydrogen bond acceptor, hydrogen bond donor, hydrophobic region, positive ionizable region, negative ionizable region and aromatic ring. The screening molecules should be match a minimum of 3 sites for a hypothesis with 3 or 4 sites and a minimum of 4 sites for a hypothesis with 5 or more sites. For the e-Pharmacophore approach, explicit matching was required for the most energetically favorable site, provided that it scored better than $-1.0 \text{ kcal mol}^{-1}$. Multiple sites were included in cases where more than one site had the top score. Screening of compounds was performed against Zinc database.⁴⁶ Database hits were ranked in order of fitness score, a measure of how well the aligned ligand conformer matches the hypothesis based on RMSD site matching, vector alignments, and volume terms.

Finally, the database hits were used for docking calculations and to score the lead like compounds. In the first step, Glide was run in high-throughput virtual screen mode. The top scoring ligands from High Throughput Virtual Screening (HTVS) were taken for second step, Glide Standard Precision (SP) stage. The top-scoring leads from Glide SP were retained and docked with Glide extra precision (XP). All the Glide protocols were run using default parameters. An extensive search was carried out for generating all possible conformations. Then,

we carried out docking using these conformations using default parameters. The schematic representation of the workflow is presented in **Figure 1**.

Results and Discussion

Analysis of essential amino acids for Jak2 inhibition

In this study, 15 ligand bound proteins crystal structures were obtained from PDB (www.rcsb.org) and their active site residues were identified by analyzing the protein-ligand interactions. Majority of the ligands in the complex structure are interacted with Leu932 and Glu930 (**Table 8**).

The effect of electrostatic potentials on docking accuracy using six types of charges and 15 ligand bound proteins crystal structures

In computer aided drug designing, especially in structure based drug design, virtual screening and molecular docking can search the preferred orientation and conformation of a ligand for its optimal binding to a receptor or enzyme active site but selecting an appropriate method to calculate the electrostatic potential is critical.^{22, 47} So, in this study we have selected four different semiempirical and two empirical methods to calculate the electrostatic potential of ligands, and their formal charges were investigated for their performance on the prediction of docking poses using Glide XP. The complex crystal structures used for this study were extracted from PDB (**Table 1**). RMSD was calculated between top-scoring docked and the complex structure ligand pose was evaluated. Usually, an RMSD of 2Å is considered as the cutoff of correct docking, probably because the resolution in an X-ray crystal structure analysis is not meaningful.^{22, 48, 49} The docking was performed for fifteen complex proteins using six different charges *viz.* AM1, RM1, PM3, MNDO, HF, and DFT. The RMSD < 1Å is considered as the best prediction, moderate prediction as 1Å < RMSD < 2Å, and the worst prediction as RMSD > 2Å.

These results show that the AM1 charge model gave the best RMSD value less than 2Å and their average RMSD values was comparatively low than others (**Figure 2**). In order to validate the accuracy of those charges, their energy scores after molecular docking were correlated with binding free energy. Known dissociation constants (K_d) were retrieved from PDBbind database.^{50, 51} Their binding free energies were calculated using Prime, Schrodinger, LLC. The correlation coefficient was calculated between binding free energy and K_d values. Among all six charge methods, AM1 has the highest correlation constant (R) in docking (**Figure 2**). So, the result reveals that AM1 charge has lowest average value and highest correlation constant. Therefore, our study suggests that AM1 charge could be the best charge model for ligand docking with Jak2 protein. The results are consistent with previous reports.^{22, 46}

Dataset

The Jak2 protein inhibitors, decaene derivatives were collected from the literatures^{23, 24}. From the above result, AM1 was considered as the best charge calculation method. Hence, we performed AM1 charge calculation for all 76 compounds and these charged compounds were docked using Glide XP in the active site of Jak2 protein.

Molecular Docking

The docking was performed with crystal structure of Jak2 (PDB id: 3E64) and a set of 76 Jak2 inhibitors. The flexible receptor docking was carried out against the receptor with 76 set of ligands using Glide XP. The calculation of ligand binding energies and ligand strain energies for a set of ligands and a single receptor, we used MM-GBSA method. Here, the inhibitors taken for QSAR studies were docked into the active site of Jak2 protein. It shows all the molecules were binding in the catalytic amino acid residues.

Structure based 3D-QSAR

The best conformation from the Glide XP docked pose were used to perform direct structure based 3D-QSAR. Additional insight into the inhibitory activity can be gained by visualizing the QSAR model in the context of most active and least active compounds. Structure based QSAR was performed to know how the 3D-QSAR methods can identify pharmacophoric features important for the interaction between ligands and their target protein. The structure based QSAR model was generated using 7 PLS factors.

Based on the overall performance of various models with respect to different statistical parameters such as SD, R^2 , F-value, RMSE, Q^2 and Pearson-R and external validation, we have selected the best QSAR model. The correlation between observed and predicted activities is critical in identifying QSAR models.

The QSAR model has R^2 value of 0.957 and Q^2 of 0.841 indicating that the model has good predictive model. In addition, this model has a low RMSE value of 0.190 and the highest Pearson-r value of 0.931 which also supports this model.

External validation

Each of the selected QSAR models was validated internally using the leave-n-out technique and externally using the corresponding test set compounds. All the models were also validated by the process randomization technique. From the internal validation technique, the value of Q^2 was determined and from the external validation technique the value of R^2 was calculated which were then used as the parameters for determining the model predictivity. The established QSAR model using 25 molecules in the test set, gave an excellent r^2_{cv} value of 0.756 (>0.5), r^2_m value of 0.962 (>0.5) as well as high slope of regression lines through the origin (k) value of 1.004 ($0.85 \leq k \leq 1.15$), and the non-cross validated correlation coefficient (r^2) values of 0.957 (close to 1), and the calculated ($r^2 - r_o^2 / r^2$) values of -0.048 (<0.1) were obtained. The results of the external validation indicated that the QSAR models possessed a high

accommodating capacity; they may be reliable for being used to predict the activities of new derivatives. Plots of predicted vs. actual pIC_{50} for training and test set are shown in **Figure 3**.

Pharmacophore based 3D-QSAR

Determination of pharmacophore model and validation

Pharmacophore models were developed by selecting highly active compounds using hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic group (H), negatively ionizable (N), positively ionizable (P), and aromatic ring (R) as pharmacophoric features. Common pharmacophores were identified using a tree-based partition algorithm that groups together similar pharmacophores according to their inter site distances.

The dataset were divided into active and inactive sets in order to find the common pharmacophore hypothesis. Molecules with pIC_{50} values higher than 7.00 were considered to be active and those with pIC_{50} values less than 6.5 were considered to be inactive, whereas those pIC_{50} values in between were considered to be moderately active. Reference relative conformational energy (kJ/mol) was included in the score and ligand activity, expressed as pIC_{50} was incorporated with a default weight. On applying the scoring function for five-featured pharmacophore hypotheses, we ranked the hypotheses considering alignment of site points and vectors, volume overlap, selectivity, number of ligands matched, relative conformational energy and activity. A total of 27 different variant hypotheses were generated upon completion of common pharmacophore identification process. The pharmacophore models whose scores ranked in the top1% were selected.⁵² The top model was found to be associated with the five point hypotheses. We have selected the two top-scored hypotheses ADHRR with 3.347 and AAHRR with 3.341 survival scores. Training set compounds were aligned on these hypotheses and analyzed by taking seven PLS (Partial Least Square) factors. The predictivity of each pharmacophore hypothesis was analyzed by the test set compounds. The best QSAR model was

selected from ADHRR hypothesis based on survival score. Hypotheses emerging from this process were subsequently scored with respect to the seventeen inactive compounds, using a weight of 1.0. The hypotheses that survived the scoring process were used to build an atom based QSAR model.

The goodness of such models is measured in terms of coefficient of determination (R^2) and cross-validated correlation coefficient (Q^2). The R^2 values of the model is greater than 0.8, the SD values are lower than 0.3 and a good F-test values. This shows that the model interpret the SAR of this series of training set compounds satisfactorily.

According to Tropsha, a high R^2 is necessary but not sufficient condition for a predictive QSAR model.⁵³ The ADHRR hypothesis has R^2 value of 0.93 and Q^2 of 0.815 indicating that ADHRR is a good predictive hypotheses. In addition, the ADHRR pharmacophore hypothesis has a low RMSE value of 0.258 and the highest Pearson-r value of 0.918 which also supports this hypothesis.

The common pharmacophore model alignment of most active compounds on the five-feature hypothesis ADHRR is shown in **Supplementary Figure 1**. This hypothesis includes five pharmacophoric features with two aromatic rings, one hydrogen bond acceptors, one hydrogen bond donor and one hydrophobic centre. This shows that the active compounds have a common structural framework with the same binding orientation. Plots of actual versus predicted pIC_{50} are shown in the **Figure 4a and 4b** for training set and test set compounds. A summary of the QSAR results of the two top scored hypotheses ADHRR and AAHRR with structure based QSAR are listed in **Table 9**.

External validation

The established QSAR model using 25 molecules in the test set, gave an excellent r^2_{cv} value of 0.816 (>0.5), r^2_m value of 0.692 (>0.5) as well as high slope of regression lines through

the origin (k) value of 0.995 ($0.85 \leq k \leq 1.15$), and the non-cross validated correlation coefficient (r^2) values of 0.930 (close to 1), and the calculated ($r^2 - r_o^2 / r^2$) values of -0.1402 (< 0.1) were obtained. The results of the external validation indicated that the QSAR models possessed a high accommodating capacity; they may be reliable for being used to predict the activities of new derivatives. Hence the hypothesis 1 with one hydrogen bond acceptors (A) and hydrogen bond donors (D) and hydrophobic region and two aromatic rings (R) as pharmacophoric feature was retained for further QSAR studies. In the pharmacophore mapping study, it was found that the major structural factors, affecting the potency of these compounds, are related to the basic scaffold. The two hydrogen bond donor sites, together with the acceptor sites, reflect the importance of the H-bonding and were consistent with the crystallographic structure Jak2 protein.

Analysis of 3D-QSAR model

Additional insight into the inhibitory activity can be gained by visualizing the QSAR model in the context of the most and least active compounds. The contribution maps obtained from our result shows how 3D-QSAR methods can identify features that are important for the interaction between each ligand and its target protein. Such maps allow the identification of those positions that require a particular physicochemical property to enhance the bioactivity of a ligand. A pictorial representation of the contours generated for most active and least active decaene derivatives are shown in **Figure 5**. In these representations, the blue cubes indicate favorable regions, while red cubes indicate unfavorable regions for activity. **Figure 5** compares the most significant favorable and unfavorable features that arise when the QSAR model is applied to the most active compound (22b) and the least active compound (28b) for Jak2 inhibition. In the context of most active compound (22b), blue cubes are significantly observed in 2-(pyrrolidin-1-yl) ethanol region (R^1). Docking result also shows that nitrogen from 2-

(pyrrolidin-1-yl) ethanol has formed hydrogen bonding with Glu930 (**Figure 6a**) which is an important residue for the inhibition of Jak2 (**Table 1**). Two methyl group near the pyrrolidine group form van der Waal interaction and pyrrolidine forms the electrostatic interaction. The result is consistent with the blue colour cubes observed in 3D-QSAR result. Hence, both QSAR and docking result confirms that 2-(pyrrolidin-1-yl) ethanol is an important substituent for the activity of the molecule against Jak2. In context of least active compound (28b), red cubes are observed on pyrrolidine (R^4) and substitution of this functional group could be the main factor which could reduce the activity of the molecule. Docking studies also reveals that pyrrolidine has not formed any interaction with Jak2 (**Figure 6b**). Instead, N13 has form hydrogen bonding with Asp994 and it has no interaction with Leu932 or Glu930. Highly electropositive substitution of pyrrolidine in R^4 could pulled the electron density towards the R^4 region and reduce the electron density in benzene ring. Benzene ring shows van der Waal interaction and dimethyl methane region shows electrostatic interaction as few cubes are also seen on these regions. This result clearly reveals that Leu932 and/or Glu930 could be an important residue for the Jak2 inhibition as it is also observed in analyzing the crystal structure of Jak2 inhibitors.

Extensive Pharmacophore based virtual screening

Pharmacophore based virtual screening and e-Pharmacophore based virtual screening

One efficient approach to drug discovery is the virtual screening of the molecular libraries. Pharmacophore based database searching is considered as a type of ligand-based virtual screening, which can be used efficiently to find novel and potential leads for further development. A potent pharmacophore model possesses the chemical functionalities responsible for bioactivities of potential drugs; therefore, it can be used to perform a database search by serving as a 3D query.

The database search studies retrieved all the positive hits and filtered out the inactive compounds. Interpretation of how the pharmacophore maps onto the positive hits can provide an insight into the structural requirements for inhibition of Jak2 and can act as a guide for further modification of the molecules. The generated pharmacophore model (**Supplementary Figure 1 & 2**) was used to screen against the zinc database which contains 2, 31,000 compounds.

For e-pharmacophore, seven pharmacophore sites were predicted for 15 co-crystal ligands. But only five pharmacophore sites were chosen based on the score. **Supplementary Figure 2** illustrates the common pharmacophore sites for co-crystal ligands. Based on these 15 pharmacophore sites virtual screening was performed against Zinc database. Summary of the compound retrieval from e-pharmacophore based virtual screening are shown in **Supplementary Table 1**.

The compounds retrieved from the pharmacophore based screening were subjected to HTVS. Through HTVS, 2140 ligands were identified which bind to the active site of Jak2. For further refinement we placed these compounds for Glide SP docking. From Glide SP, 230 compounds were filtered out and further proceeded for more precise Glide XP docking study. Finally, we identified 27 ligands which interact to the active site residues of Jak2. The chemical structures of these lead molecules are illustrated in **Supplementary Figure 3**. All the five pharmacophore hypotheses sites are present in our identified compounds (**Supplementary Table 2**). The 27 identified compounds were charge with AM1 charge model and redocked to the Jak2 protein using Glide XP. All the compounds were bind to the active site of Jak2. Further, binding energy was calculated for these 27 compounds and MM/GBSA solvation energy (ΔG_{bind}) was ranging from -16.48 to -32.15 kcal/mol suggests strong ligand enzyme interactions.

Conclusion

The present study deals with the comparison of charge calculation methods such as semi empirical (AM1, RM1, PM3, and MNDO) and empirical methods (DFT, HF). Among these methods AM1 has the low average RMSD value and high R^2 value. Therefore, AM1 is the best charge model for our docking calculation. The multi-kinase inhibitor decaene derivatives charge were calculated with AM1 for Glide XP docking study and the conformation generated from Glide XP docking were used for structure based QSAR and the established model has good R^2 value of 0.955, Q^2 value of 0.841 with excellent r^2_{cv} value of 0.756 (>0.5). In pharmacophore based, QSAR model has R^2 value of 0.930 and Q^2 value of 0.815 with r^2_{cv} value of 0.816 (>0.5). From the result, it reveals that structure based QSAR has better model than pharmacophore based QSAR. Extensive virtual screening using pharmacophore hypothesis generated from decaene derivatives and e-pharmacophore model from 15 ligand bound Jak2 proteins crystal structures followed by HTVS, SP and XP against zinc database identified 27 compounds that could be a new potential inhibitor for Jak2. This study may provide a valuable benchmark in QSAR modeling which will further helpful in the discovery of novel potent inhibitors.

Acknowledgement

One of the authors K.D.S. gratefully acknowledges CSIR for providing the Senior Research Fellowship (SRF). The authors acknowledge the infrastructure facilities provided by Department of Bioinformatics, Alagappa University, Post Graduate Diploma in Structural Pharmacogenomics (Funded by UGC, Government of India).

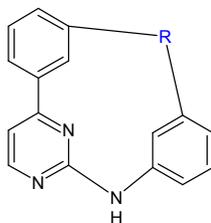
References

1. K. Yamaoka, P. Saharinen, M. Pesu, V. E. Holt, O. Silvennoinen and J. J. O'Shea, *Genome Biol*, 2004, 5, 253.
2. R. M. Bandaranayake, D. Ungureanu, Y. Shan, D. E. Shaw, O. Silvennoinen and S. R. Hubbard, *Nat Struct Mol Biol*, 19, 754-759.
3. J. J. O'Shea, M. Gadina and R. D. Schreiber, *Cell*, 2002, 109 Suppl, S121-131.
4. C. Guilluy, J. Bregeon, G. Toumaniantz, M. Rolli-Derkinderen, K. Retailleau, L. Loufrani, D. Henrion, E. Scalbert, A. Bril, R. M. Torres, S. Offermanns, P. Pacaud and G. Loirand, *Nat Med*, 16, 183-190.
5. K. K. Griendling, D. Sorescu and M. Ushio-Fukai, *Circ Res*, 2000, 86, 494-501.
6. H. Ohtsu, G. D. Frank, H. Utsunomiya and S. Eguchi, *Antioxid Redox Signal*, 2005, 7, 1315-1326.
7. C. Yan, D. Kim, T. Aizawa and B. C. Berk, *Arterioscler Thromb Vasc Biol*, 2003, 23, 26-36.
8. R. M. Touyz, *Braz J Med Biol Res*, 2004, 37, 1263-1273.
9. Y. Taniyama and K. K. Griendling, *Hypertension*, 2003, 42, 1075-1081.
10. M. Ushio-Fukai, R. W. Alexander, M. Akers, Q. Yin, Y. Fujio, K. Walsh and K. K. Griendling, *J Biol Chem*, 1999, 274, 22699-22704.
11. N. R. Madamanchi, S. Li, C. Patterson and M. S. Runge, *Arterioscler Thromb Vasc Biol*, 2001, 21, 321-326.
12. C. Walz, B. J. Crowley, H. E. Hudon, J. L. Gramlich, D. S. Neuberg, K. Podar, J. D. Griffin and M. Sattler, *J Biol Chem*, 2006, 281, 18177-18183.
13. M. Sattler, S. Verma, G. Shrikhande, C. H. Byrne, Y. B. Pride, T. Winkler, E. A. Greenfield, R. Salgia and J. D. Griffin, *J Biol Chem*, 2000, 275, 24273-24278.
14. S. Verstovsek, *Hematology Am Soc Hematol Educ Program*, 2009, 636-642.
15. K. S. N.Vaddi and V.Gupta, *Med Lett Drugs their*, **2012**, 54, 27-8.
16. R. A. Mesa, *IDrugs*, 2010, 13, 394-403.
17. A. Ostojic, R. Vrhovac and S. Verstovsek, *Drugs Today (Barc)*, 2011, 47, 817-827.
18. C. A. Zerbini and A. B. Lomonte, *Expert Rev Clin Immunol*, 2012, 8, 319-331.
19. A. Kirabo, P. P. Sayeski, *Pharmaceuticals* 2010, 3, 3478-3493.

20. Y. Nakaya, K. Shide, H. Naito, T. Niwa, T. Horio, J. Miyake, K. Shimoda, *Blood Cancer J.* 2014, 4:e174
21. L. B. Schenkel, X. Huang, A. Cheng, H. L. Deak, E. Doherty, R. Emkey, Y. Gu, H. Gunaydin, J. L. Kim, J. Lee, R. Loberg, P. Olivieri, J. Pistillo, J. Tang, Q. Wan, H. L. Wang, S. W. Wang, M. C. Wells, B. Wu, V. Yu, L. Liu, *J Med Chem*, 2011, 54, 8440-50.
22. K. C. Tsai, S. H. Wang, N. W. Hsiao, M. Li and B. Wang, *Bioorg Med Chem Lett*, 2008, 18, 3509-3512.
23. A. D. William, A. C. Lee, S. Blanchard, A. Poulsen, E. L. Teo, H. Nagaraj, E. Tan, D. Chen, M. Williams, E. T. Sun, K. C. Goh, W. C. Ong, S. K. Goh, S. Hart, R. Jayaraman, M. K. Pasha, K. Ethirajulu, J. M. Wood and B. W. Dymock, *J Med Chem*, 54, 4638-4658.
24. A. D. William, A. C. Lee, K. C. Goh, S. Blanchard, A. Poulsen, E. L. Teo, H. Nagaraj, C. P. Lee, H. Wang, M. Williams, E. T. Sun, C. Hu, R. Jayaraman, M. K. Pasha, K. Ethirajulu, J. M. Wood and B. W. Dymock, *J Med Chem*, 55, 169-196.
25. A.D.Bochevarov, E. Harder, T. F. Hughes, J. R. Greenwood, D. A. Braden, D. M. Philipp, D. Rinaldo, M. D. Halls, J. Zhang, R. A. Friesner, *Int. J. Quantum Chem*, 2013, 113, 2110-2142
26. Suite 2012: Jaguar, version 7.9, Schrödinger, LLC, New York, NY, 2012.
27. Suite 2012: Maestro, version 9.3, Schrödinger, LLC, New York, NY, 2012.
28. Suite 2012: LigPrep, version 2.5, Schrödinger, LLC, New York, NY, 2012.
29. S. L. Dixon, A. M. Smondyrev, E. H. Knoll, S. N. Rao, D. E. Shaw and R. A. Friesner, *J Comput Aided Mol Des*, 2006, 20, 647-671.
30. S. L. Dixon, A. M. Smondyrev and S. N. Rao, *Chem Biol Drug Des*, 2006, 67, 370-372.
31. Suite 2012: Phase, version 3.4, Schrödinger, LLC, New York, NY, 2012.
32. K. Dhanachandra Singh, M. Karthikeyan, P. Kirubakaran and S. Nagamani, *J Mol Graph Model*, 2011, 30, 186-197.
33. K. D. Singh and K. Muthusamy, *Acta Pharmacol Sin*, 2013, 34, 1592-1606.
34. A. Golbraikh and A. Tropsha, *J Mol Graph Model*, 2002, 20, 269-276.
35. P. Lu, X. Wei and R. Zhang, *Eur J Med Chem*, 2010, 45, 3413-3419.
36. A. Basu, K. Jasu, V. Jayaprakash, N. Mishra, P. Ojha and S. Bhattacharya, *Eur J Med Chem*, 2009, 44, 2400-2407.

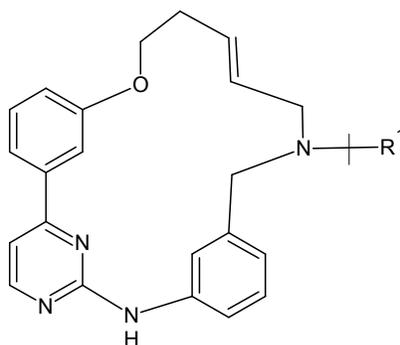
37. R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin and D. T. Mainz, *J Med Chem*, 2006, 49, 6177-6196.
38. T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks, *J Med Chem*, 2004, 47, 1750-1759.
39. R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J Med Chem*, 2004, 47, 1739-1749.
40. Suite 2012: Glide, version 5.8, Schrödinger, LLC, New York, NY, 2012.
41. M. P. Jacobson, D. L. Pincus, C. S. Rapp, T. J. Day, B. Honig, D. E. Shaw and R. A. Friesner, *Proteins*, 2004, 55, 351-367.
42. M. P. Jacobson, R. A. Friesner, Z. Xiang and B. Honig, *J Mol Biol*, 2002, 320, 597-608.
43. Suite 2012: Prime, version 3.1, Schrödinger, LLC, New York, NY, 2012.
44. N. K. Salam, R. Nuti and W. Sherman, *J Chem Inf Model*, 2009, 49, 2356-2368.
45. K. Loving, N. K. Salam and W. Sherman, *J Comput Aided Mol Des*, 2009, 23, 541-554.
46. J. J. Irwin and B. K. Shoichet, *J Chem Inf Model*, 2005, 45, 177-182.
47. X. Hou, J. Du, J. Zhang, L. Du, H. Fang and M. Li, *J Chem Inf Model*, 2013, 53, 188-200.
48. K. Onodera, K. Satou and H. Hirota, *J Chem Inf Model*, 2007, 47, 1609-1618.
49. M. S. Park, A. L. Dessal, A. V. Smrcka and H. A. Stern, *J Chem Inf Model*, 2009, 49, 437-443.
50. R. Wang, X. Fang, Y. Lu, C. Y. Yang and S. Wang, *J Med Chem*, 2005, 48, 4111-4119.
51. R. Wang, X. Fang, Y. Lu and S. Wang, *J Med Chem*, 2004, 47, 2977-2980.
52. S. L. Dixon, A. M. Smondyrev and S. N. Rao, *Chem Biol Drug Des*, 2006, 67, 370-372.
53. A. Tropsha, *QSAR Comb Sci*, 2010, 29, 476-488.

Table 1: Compounds selected (Amide containing linkers) for 3D-QSAR study and their measured biological activity and predicted activity.



Compound No	R	Experimental pIC ₅₀ value	Predicted pIC ₅₀ value	Pharm set
26a		5.76	6.05	Training
26b		5.00	4.93	Training
26c		5.00	5.05	Training
26e		5.48	5.61	Training
26g 1		6.82	6.7	Training
26h		7.14	6.92	Training
26i		5.77	6.29	Test
26j		6.64	6.54	Test

Table 2: Compounds selected (Amide substituent) for 3D-QSAR study and their measured biological activity and predicted activity.



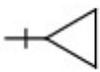
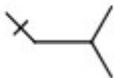
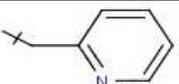
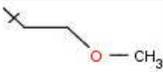
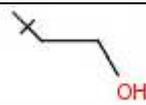
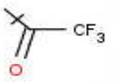
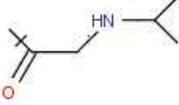
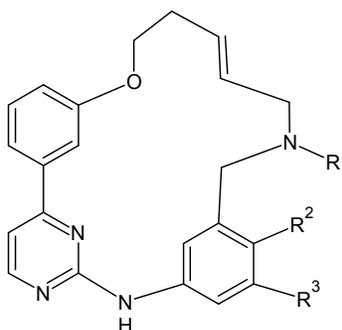
Compound No	R1	Experimental pIC ₅₀ value	Predicted pIC ₅₀ value	Pharm set
26g 2	H	6.82	6.69	Training
26h 2	CH ₃	7.14	6.73	Training
26k		7.05	6.8	Training
26l		5.82	5.86	Training
26m		5.82	5.96	Training
26n		5.00	5.21	Training
26o		5.87	6.18	Test
26p		5.80	6.29	Training
26q		6.14	5.94	Test
26f		5.00	5.12	Training
26r		6.92	6.84	Training

Table 3: Compounds selected (Amine and Aromatic Ring Substitutions) for 3D-QSAR study and their measured biological activity and predicted activity.



Compound No	R ₁	R ₂	R ₃	Experimental pIC ₅₀ value	Predicted pIC ₅₀ value	Pharm Set
26h 3	Me	H	H	7.14	6.97	Training
27a	Me	OCH ₃	H	6.55	6.69	Test
27b	Me	Cl	H	7.09	6.75	Training
27c	Me	OCF ₃	H	6.82	7.07	Training
27d	Me		H	7.21	7.31	Training
27e	Me		H	7.00	6.95	Test
27g	Me		H	6.96	6.77	Test
27h	Me		H	6.85	7.07	Training
27i	Me		H	7.31	7.22	Training
27j	Me		H	7.32	7.24	Training
27k	Me		H	7.25	7.29	Training

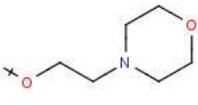
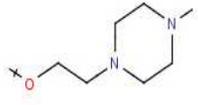
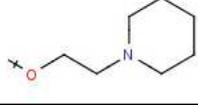
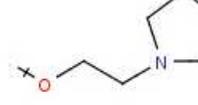
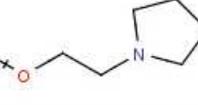
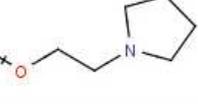
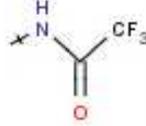
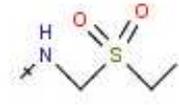
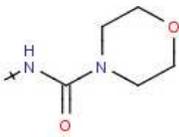
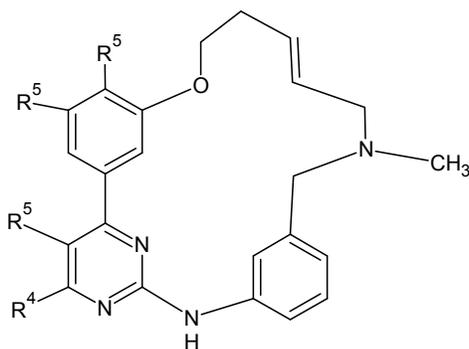
27n	Me		H	6.92	6.97	Test
27o	Me		H	6.66	6.82	Training
27p	Me		H	7.21	7.08	Test
27l			H	6.66	6.71	Training
27m			H	6.82	7.02	Test
27q	Me		OMe	6.68	6.95	Test
27r	Me	H		7.11	7.52	Training
27s	Me	H	NH ₂	7.30	7.48	Training
27t	Me	H		7.75	8.03	Training
27u	Me	H		7.36	7.32	Training

Table 4: Compounds selected (Aromatic Ring Substitutions) for 3D-QSAR study and their measured biological activity and predicted activity.



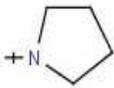
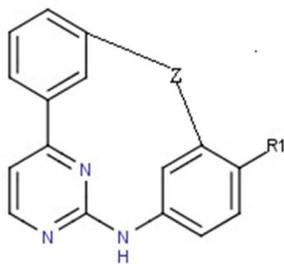
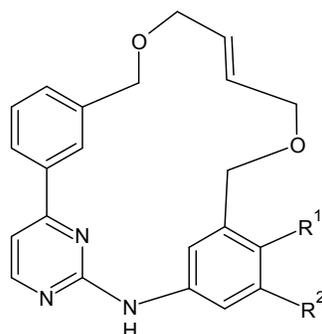
Compound No	R ₄	R ₅	R ₆	R ₇	Experimental pIC ₅₀ value	Predicted pIC ₅₀ value	Pharm Set
26h 4	H	H	H	H	7.14	6.61	Training
28a	CH ₃	H	H	H	5.00	5.01	Training
28b		H	H	H	5.00	5.21	Training
28c	H	CH ₃	H	H	6.75	6.84	Training
28d	H	H	F	H	6.80	6.57	Training
28e	H	H	H	OMe	6.29	6.48	Test

Table 5: Compounds selected (suitable linkers) for 3D-QSAR study and their measured biological activity and predicted activity.



Compound No	-Z-	R ₁	Experimental pIC ₅₀ value	Predicted pIC ₅₀ value	Pharm Set
16a		H	5.92	6.28	Test
16b		H	6.64	6.33	Training
16c		OCH ₃	6.89	7.26	Training
16d		OCH ₃	6.59	6.58	Test
16e		OCH ₃	7.16	7.42	Training
16f		OH	7.28	7.20	Test

Table 6: Compounds selected (Solubilizing Groups) for 3D-QSAR study and their measured biological activity and predicted activity.



Compound No	R ₁	R ₂	Experimental pIC ₅₀ value	Predicted pIC ₅₀ value	Pharm Set
17a		H	7.22	7.13	Test
17c		H	7.08	7.12	Test
17d		H	7.39	7.12	Training
17e		H	7.08	6.76	Training
17f		H	6.92	7.07	Training
17g		H	6.82	6.79	Training
17h		H	7.11	7.02	Test
21a		H	6.66	7.32	Test
21b		H	7.62	7.56	Training

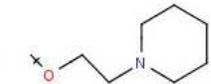
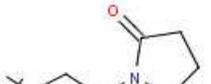
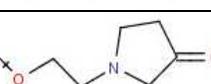
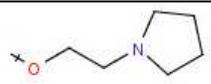
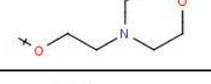
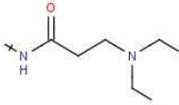
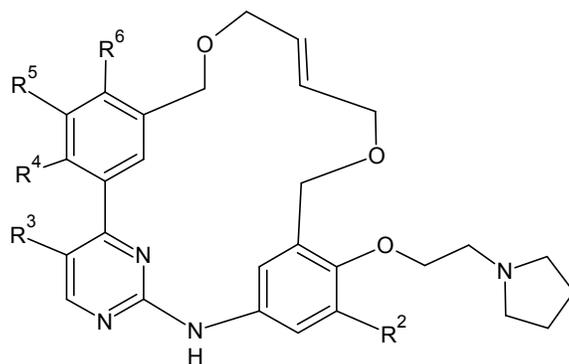
21c		H	7.64	7.57	Training
21d		H	7.82	7.81	Training
21e		H	7.66	7.84	Training
21f		H	7.66	7.56	Test
21h	H		7.32	7.24	Training
21i	H		7.02	7.23	Test
21k	H		7.46	7.46	Training
21l	H		7.36	7.33	Training

Table 7: Compounds selected (Aromatic Ring Substitutions with Small Groups) for 3D-QSAR study and their measured biological activity and predicted activity.



Compound No	R ₂	R ₃	R ₄	R ₅	R ₆	Experimental pIC ₅₀ value	Predicted pIC ₅₀ value	Pharm Set
22a	OCH ₃	H	H	H	H	7.44	7.41	Training
22b	H	CH ₃	H	H	H	8.16	7.61	Test
22c	H	F	H	H	H	7.77	7.51	Test
22d	H	H	OCH ₃	H	H	6.48	6.14	Training
22e	H	H	H	F	H	7.62	7.52	Test
22f	H	H	H	H	OCH ₃	7.72	7.57	Training
22g	H	H	H	H	F	7.60	7.54	Test

Table 8. Analysis of essential residues in Jak2 protein for its inhibition using 15 ligand bound proteins crystal structures reported in PDB.

PDB ID	Leu932	Glu930	Glu898	Phe995	Asp994
2B7A	+	+	-	-	-
2W1I	+	+	-	-	-
2XA4	+	-	-	-	-
3E62	+	+	-	-	-
3E63	+	+	-	-	-
3E64	+	+	-	-	+
3FUP	+	+	-	-	-
3IOK	+	-	-	-	-
3JY9	+	+	+	+	+
3KCK	+	+	+	+	-
3KRR	+	-	-	-	-
3LPB	+	-	-	-	-
3Q32	+	-	-	-	-
3TJC	+	+	-	-	-
3TJD	+	+	-	-	-

Table 9. Summary of QSAR analysis of Pharmacophore based and structure based methods.

External validation values	ADHRR	AAHRR	Structure based QSAR
R^2	0.930	0.896	0.957
Q^2	0.815	0.724	0.841
SD	0.226	0.275	0.179
F	208.300	135.200	292.700
RMSE	0.258	0.315	0.190
Pearson	0.918	0.867	0.931
$r^2-r_0^2/r^2$	-0.140	-0.215	-0.048
r_m^2	0.692	0.614	0.962
r_{pred}^2	0.998	0.998	0.790
r_{cv}^2	0.816	0.825	0.756
K	0.995	0.995	1.004

Figure Legend

Figure 1. Schematic representation of the workflow.

Figure 2. Correlation co-efficient (R^2) for the experimental value and predicted ΔG_{bind} , and average RMSD of the 15 co-crystal ligands after docking.

Figure 3. Graph of actual versus predicted pIC_{50} of the training set and the test set using Structure based QSAR. (a) Training set and (b) Test set.

Figure 4. Graph of actual versus predicted pIC_{50} of the training set and the test set using Pharmacophore based QSAR. (a) Training set and (b) Test set.

Figure 5. Pictorial representation of the cubes generated using the QSAR model. Blue areas indicate favourable regions, while red areas indicate unfavourable regions for the activity. The structure based 3D-QSAR model visualized in the context of most active compound (a) and least active compound (b). The Pharmacophore based 3D-QSAR model visualized in the context of most active compound (c) and least active compound (d).

Figure 6. The binding mode of most active compound (a) and least active compound (b) inside the catalytic site of Jak2 protein.

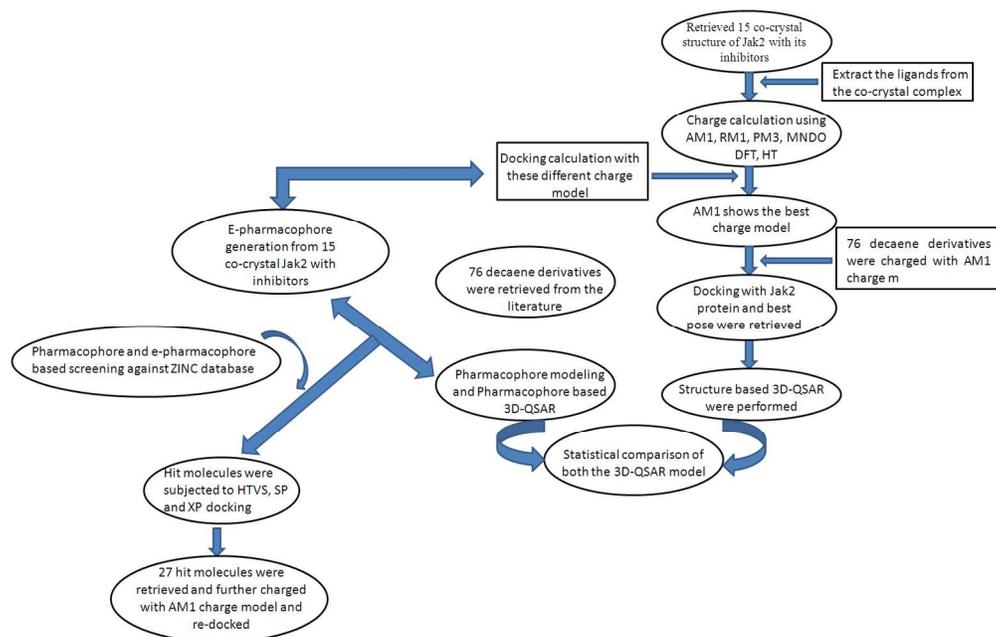


Figure 1. Schematic representation of the workflow.
447x284mm (300 x 300 DPI)

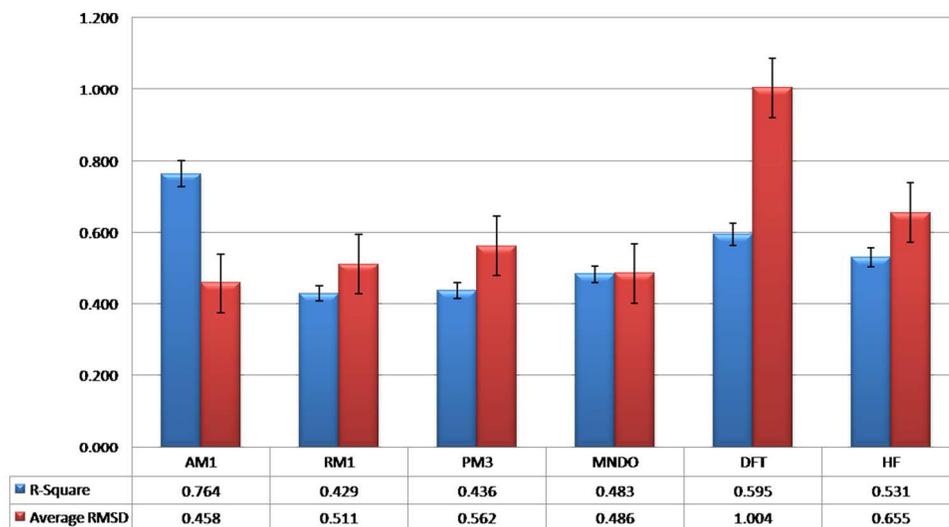


Figure 2. Correlation co-efficient (R^2) for the experimental value and predicted ΔG_{bind} , and average RMSD of the 15 co-crystal ligands after docking.
214x116mm (300 x 300 DPI)

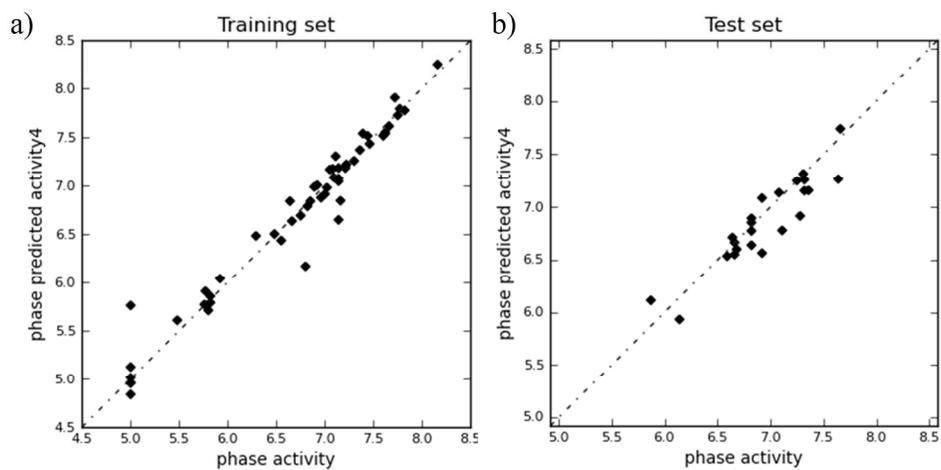


Figure 3. Graph of actual versus predicted pIC50 of the training set and the test set using Structure based QSAR. (a) Training set and (b) Test set.
230x115mm (300 x 300 DPI)

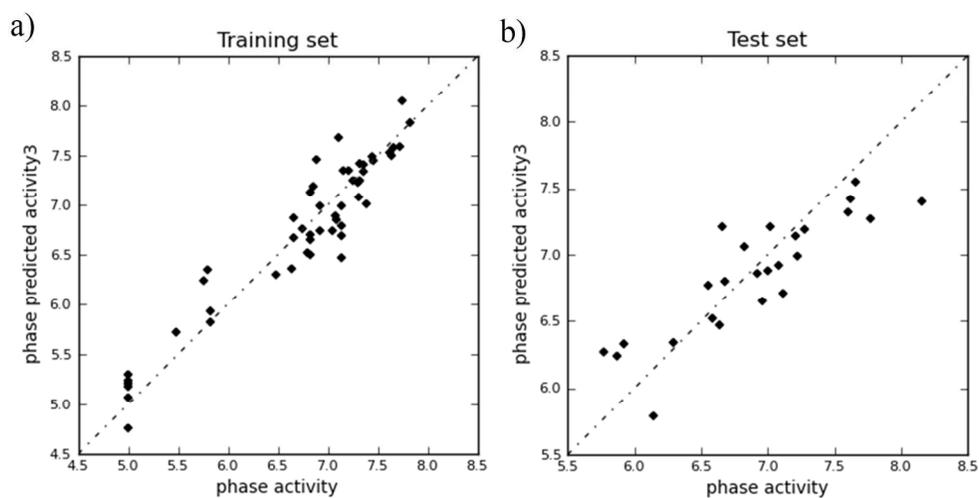


Figure 4. Graph of actual versus predicted pIC50 of the training set and the test set using Pharmacophore based QSAR. (a) Training set and (b) Test set.
215x111mm (300 x 300 DPI)

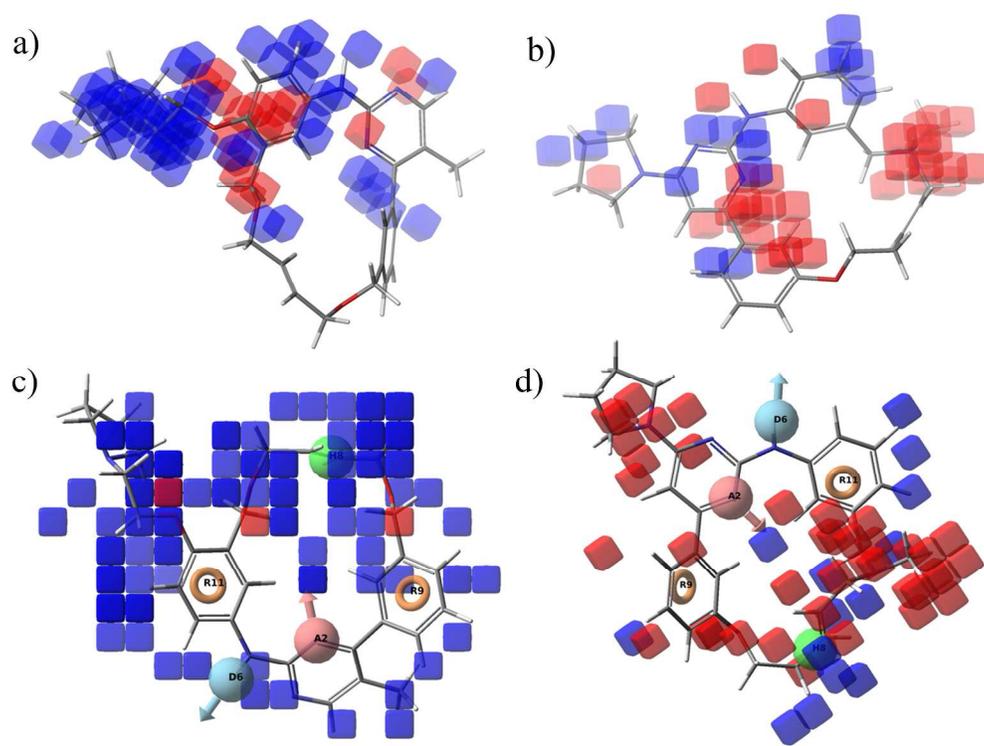


Figure 5. Pictorial representation of the cubes generated using the QSAR model. Blue areas indicate favourable regions, while red areas indicate unfavourable regions for the activity. The structure based 3D-QSAR model visualized in the context of most active compound (a) and least active compound (b). The Pharmacophore based 3D-QSAR model visualized in the context of most active compound (c) and least active compound (d).

249x191mm (300 x 300 DPI)

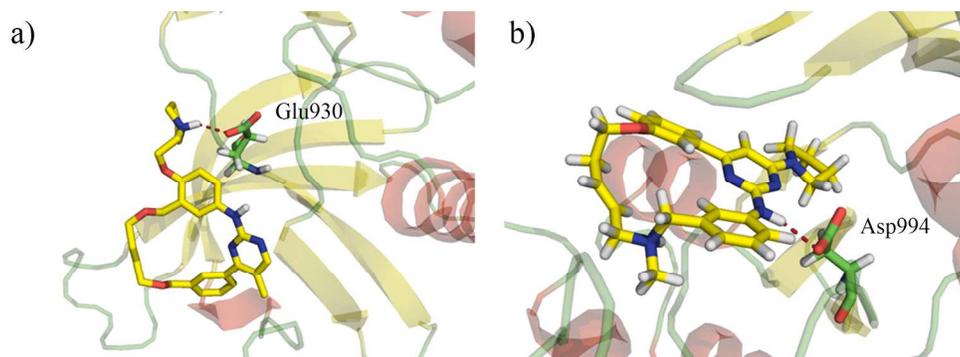


Figure 6. The binding mode of most active compound (a) and least active compound (b) inside the catalytic site of Jak2 protein.
217x79mm (300 x 300 DPI)