WORLD JOURNAL OF PHARMACEUTICAL RESEARCH
SJIF Impact Factor 8.453
Volume 14, Issue 16, 511-527. Research Article ISSN 2277–7105

2D AND 3D QSAR MODELING OF 2-AMINOPYRIMIDINE DERIVATIVES OR FLT3 KINASE INHIBITOR FOR THE TREATMENT OF ACUTE MYELOID LEUKAEMIA

[1*]Debapriya Dey, Adarsha Ganguly, Snehasish Koner, Hirak Bhowmik, Debarghya Karforma, Sagar Dey, Subhas Chandra Hazari

GM Mukherjee Lane, Dulmi Nadiha, Purulia, 723102.

Article Received on
24 June 2025,

Revised on 14 July 2025,
Accepted on 03 August 2025,

DOI: 10.20959/wjpr202516-37905

***Corresponding Author**
Debapriya Dey
GM Mukherjee Lane, Dulmi Nadiha, Purulia, 723102.

ABSTRACT

Acute myeloid leukemia (AML) represents an aggressive hematological malignancy. In 2020, the American Cancer Society estimated 19,940 cases in the United States, with an estimated 11,180 deaths, emphasizing the urgent need for the development of novel and more effective therapeutic agents. In this work, thirty-two 2-aminopyrimidine derivatives were collected from the recently published investigation. These compounds were reported to have activity against Fms-like tyrosine kinase 3 (FLT3) enzymes, which is considered as a promising target for the treatment of AML. More importantly, these compounds exhibited activity against both wild type and D835Y mutant enzymes. In this work, we performed ligand-based in silico approaches such as 2D-quantitative structure activity relationship (2D-QSAR) and 3D-QSAR to understand the structural requirements of these 2-aminopyrimidine derivatives for higher biological activity against these enzyme targets. A range of feature selection techniques were employed for developing 2D-QSAR models and the best models were generated with highly satisfactory internal and external predictivity. More significantly, it was observed that limited number of descriptors with high interpretability are capable of developing models, the statistical qualities of which are similar to the models developed after including large number of 2D/3D complex descriptors. Therefore, the 2D-QSAR models presented in this work afford high statistical significance and mechanistic interpretation. The predictive validated 3D-QSAR models were developed with the structures aligned with unsupervised rigid body alignment method. These 3D-QSAR models revealed the importance of steric and electrostatic fields responsible for determining

higher activity against these enzymes. Overall, the current work may be used as important guidelines to design novel FLT3 kinase inhibitors for the treatment of AML.

KEYWORDS: 2D-QSAR, 3D-QSAR, 2-AMINOPYRIMIDINE DERIVATIVES, FLT3 KINASE INHIBITOR, ACUTE MYELOID LEUKAEMIA.

INTRODUCTION

Leukemia are tumors that beginning in cells that would ordinarily form into various kinds of platelets. Most frequently, leukemia begins in early types of white platelets, however a few leukemia. start in other platelet types. Acute myeloid leukemia (AML) begins in the bone marrow (the delicate inward piece of specific bones, where fresh blood cells are made), however most frequently it rapidly moves into the blood, too. It can in some cases spread to different pieces of the body including the lymph hubs, liver, spleen, focal sensory system (mind and spinal line), and balls.[1-3]

Most frequently, AML creates from cells that would transform into white platelets (other than lymphocytes), however now and again AML creates in different kinds of blood-shaping cells.[4] FLT3 is a quality change, or transformation, in leukemia (blood disease) cells. It's the most considered normal hereditary change in Acute myeloid leukemia (AML), a kind of leukemia that beginnings in the bone marrow and frequently moves into the blood.[5] Genomic examinations of Acute myeloid leukemia (AML) have shown that few qualities are repetitively transformed, prompting new genomic groupings, prescient biomarkers, and new restorative targets. Changes of the FMS-like tyrosine kinase 3 (FLT3) quality happen in roughly 30% of all AML cases, with the inward pair duplication (ITD) addressing the most well-known sort of FLT3 transformation (FLT3-ITD; around 25% of all AML cases). FLT3-ITD is a typical driver transformation that presents with a high leukemic weight and gives an unfortunate visualization in patients with AML. The prognostic worth of a FLT3 change in the tyrosine kinase space (FLT3-TKD), which has a lower frequency in AML (roughly 7-10% of all cases), is questionable. Amassing proof shows that FLT3 mutational status advances all through the sickness continuum. This purported clonal development, along with the ID of FLT3-ITD as a negative prognostic marker, features the significance of FLT3-ITD testing at finding and again at backslide. Prior recognizable proof of FLT3 transformations will assist with giving a superior comprehension of the patient's sickness and empower designated therapy that might end up being useful to patients accomplish longer and more solid reductions. Original FLT3 inhibitors produced for clinical use are expansive range,

multi kinase inhibitors; nonetheless, cutting edge FLT3 inhibitors are more unambiguous, more intense, and have less poison levels related with askew impacts. Essential and optional procured protection from FLT3 inhibitors stays a test and gives a reasoning to joining FLT3 inhibitors with different treatments, both traditional and investigational. This audit centers around the neurotic and prognostic job of FLT3 changes in AML, clinical grouping of the sickness, late advancement with cutting edge FLT3 inhibitors, and systems of protection from FLT3 inhibitors.[5,6]

c-Kit, a receptor tyrosine kinase, is engaged with intracellular flagging, and the transformed type of c-Kit assumes a urgent part in event of certain tumors. The capacity of c-Kit has prompted the idea that restraining c-Kit kinase movement can be an objective for disease treatment. The promising aftereffects of hindrance of c-Kit for therapy of diseases have been seen in certain malignant growths, for example, gastrointestinal stromal cancer, Acute myeloid leukemia, melanoma, and different cancers, and these outcomes have energized endeavors toward progress of involving c-Kit as a skilled objective for malignant growth treatment. This paper presents the discoveries of past examinations seeing c-Kit as a receptor tyrosine kinase and an oncogene, as well as its quality targets and flagging pathways in typical and disease cells. The c-Kit quality area, protein structure, and the job of c-Kit in typical cell have been examined. Appreciating the sub-atomic instrument fundamental c-Kit-interceded tum orogenesis is thus fundamental and may prompt the recognizable proof of future novel medication targets. The expected instruments by which c-Kit prompts cell change have been portrayed. This study plans to explain the capacity of c-Kit for future malignant growth treatment. Furthermore, it has c-Kit inhibitor drug properties and their capacities have been recorded in tables and shown in schematic pictures. This survey likewise has gathered past examinations that designated c-Kit as an original methodology for malignant growth treatment. This paper further accentuates the upsides of this methodology, as well as the impediments that should be tended to from here on out. At long last, albeit c-Kit is an alluring objective for disease treatment, in light of the results of treatment of patients with c-Kit inhibitors, it is impossible that Kit inhibitors alone can prompt fix. It appears to be that c-Kit transformations alone are not adequate for tumorogenesis, however assume a pivotal part in malignant growth event.[3,7]

AIM AND OBJECTIVE

To develop predictive validated ligand-based *in silico* models (i.e., 2D-QSAR and 3D-QSAR) to understand the structural requirements of some 2‑aminopyrimidine derivatives as inhibitors of FMS-like tyrosine kinase 3 (FLT3) for the treatment of acute myeloid leukemia (AML)

LITERATURE REVIEW - Discovery of novel therapeutic agents is largely known as an exceptionally intricate cycle requiring massive investment. So presently computer-aided drug design (CADD) approaches are utilized to speed-up the productivity of the medication revelation and improvement course. Different methodologies of CADD are now available [8]and these are broadly divided into two categories: (a) ligand-based drug design (e.g., 2D-QSAR, 3D-QSAR) and (b) structure-based drug design (e.g., molecular docking). Currently these CADD approached are widely utilized for lead identification and lead modification purposes to improve the potency and/or to improve pharmacokinetic profiles of novel biologically active compounds. This work demonstrates ligand-based CADD techniques to the structural requirements of some 2‑aminopyrimidine derivatives as inhibitors of FMS-like tyrosine kinase 3 (FLT3) for the treatment of acute myeloid leukemia (AML).

MATERIALS AND METHODS

Dataset collection and structure preparation

Thirty-two 2‑Aminopyrimidine derivatives were collected from the recently published investigation Tong et al.[9] The anticancer Property of these compounds were reported against Fms-like tyrosine kinase 3 (FLT3) is fundamentally associated with cell endurance expansion, as well as the separation of hematopoietic progenitors of lymphoid and myeloid lineages. Its transformations address around 30% of recently analyzed. AML, including inside pair duplication (ITD) changes what's more, tyrosine kinase domain (TKD) mutations,8 commonly related with poor clinical anticipation. The 50% effective concentration (EC_{50} in µM) values of these compounds were converted to pEC_{50} (=-$\log_{10}(EC_{50}/10^6)$) values and were subsequently used for as the response variables for 2D- and 3D-QSAR methods. The SMILES notations of these structures (provided by the authors) were converted to .sdf file using Discovery Studio Visualization tool and then numbered accordingly. These structures were subsequently standardized using Standardizer tool of Chemaxon using following options: the following options: (a) add explicit hydrogen atoms, (b) aromatize, (c) clean 2D, (d) clean 3D, (e) neutralize and (f) strip salts. The standardized

structures were further processed for 2D- and 3D-QSAR analyses.

2D-QSAR analyses

Descriptor calculation- Descriptors of 32 2‑Aminopyrimidine Derivatives were calculated using alvaDescv.2.0.4 (https://www.alvascience.com/alvadesc/)[10] under OCHEM webserver.[11] For calculation of 3D descriptors, the structures of the compounds were geometrically optimized in this web platform using Corina tool. The calculated descriptors of these compounds were then merged with respective pEC_{50} of the compounds to form the dataset for 2D-QSAR model generation.

Dataset division and model development

The dataset was divided into a training and a test set using activity sorting method (using starting point of 2) using SFS-QSAR tool.[12] The models were developed in three stages. In the first stage, some selected descriptors having higher overall interpretability were considered in search of interpretable 2D-QSAR models. In the second stage, we applied 2D descriptors for model generation and finally, in the last stage all descriptors were employed for model generation. The purpose was to understand whether the inclusion of 2D descriptors improved the quality of the model to a considerable extent or not. Similarly, in the third stage it was assessed whether 3D descriptors, the values of which are sensitive to the specific 3D conformation, are truly essential in characterizing the structural requirement of the compounds or not. 2D-QSAR models were developed using two feature selection techniques namely (a) sequential forward selection (SFS) and (b) genetic algorithm (GA). The SFS based model were developed using newly developed open-access Python based SFS-QSAR tool (https://github.com/ncordeirfcup/SFS-QSAR-tool) which implements 'Feature Selector' module of the library Mlxtend (http://rasbt.github.io/mlxtend/). Data treatment was performed by setting variance cut-off of 0.0001 (to remove constant and near-constant descriptors) and correlation cut-off of 0.99 (to eliminate highly inter-correlated descriptors). During model development, four scoring functions of the 'Sequential Feature Selector' module were employed for feature selection, namely: the determination coefficient (R^2), the negative mean absolute error (*NMAE*), the negative mean Poisson deviance (*NMPD*), and the negative mean gamma deviance (*NMGD*). No cross-validation was performed during feature selection. GA-based models were developed using open access tool Genetic Algorithm v.4.1_2 (accessed from https://dtclab.webs.com/software-tools).[13] In contrast to SFS, which is considered as non-stochastic feature selection method, GA follows stochastic algorithms to

generate randomized models and it employs techniques such as cross-over and mutations to improve the fitting of the independent variables with the response variable. Similar to SFS-QSA, descriptor pre-treatment was carried out during development of GA-based models. Maximum five descriptors were initially allowed in the 2D-QSAR models.

Statistical analyses of the models

The predictivity of the final 2D-QSAR model was justified with a range of statistical parameters including $R^{2,}$ adjusted R^2 (R^2_A), *F*-statistics and internal cross-validation parameter Q^2. The external validation parameter R^2_{Pred} was used to check external predictivity of the generated models. Apart from these, the parameters related to rm2 matrices such as rm^2_{LOO} and Δrm^2_{LOO} were used as internal validation parameter whereas rm^2_{test} and $\Delta\ rm^2_{test}$ were used as external validation parameters.[14-16] The best QSAR models were also checked for inter-collinearity among its descriptors. Furthermore, *Y*-randomization test was conducted with 1000 runs to generate 1000 models with randomized response variables. To confirm that the model was not developed by chance, a parameter named $^cR^2_p$ was calculated.[17] The higher value of $^cR^2_p$ implies that the original model was not developed by chance.

Applicability domain of the models

In this work, applicability domain of the models was estimated by Williams plot, which is a plot between leverage values and standardized coefficients. If the leverage value of any data-point is larger than the hat value (which is 3*p/n, where p is number of model's descriptors + 1 and n is the number of data-points in the training set), it is considered as a structural outlier. On the other hand, if standardized residual of the data-point is > ±2.5, it is considered as response outlier.[18]

3D-QSAR analyses

Alignment methods

3D-QSAR analyses were performed with open-source software named Open3DQSAR.[19] alignment methods was used in the current work and this is ligand-based alignment.[20] Ligand-based alignment were performed using newly developed open-source tool named Openpharmacophore though the codes of this tool was slightly modified to obtain the aligned structures as well as to screen the ligands against the developed pharmacophore. In ligand-based alignment, the 'LigandBasedPharmacophore' feature of this tool was utilized that allow the lower energy conformations of the ligands to align with the help of rdMolAlign module of rdkit. In this work, the 3D structures of the ligands were first minimized using

'obminimize' function of OpenBabel tool using steepest descent method. The minimized structures were submitted for ligand-based alignment. The developed pharmacophore was then used for screening of all dataset molecules using EmbedLib. MatchPharmacophore and EmbedLib. EmbedPharmacophore functions of rdkit.

Model development

The 3D-QSAR models were developed using open-source software named Open3DQSAR using training and test sets combinations.[19,21] The detailed methodologies of this tool have been discussed earlier. Briefly, this software uses a carbon and volume-less positively (+1) charged probe for calculating steric and electrostatic fields of the query chemicals, respectively. The pre-treatment of the fields was carried out after setting a smart region definition (SRD)-cut-off level of 2.0 and also by removing *N*-level variables. The Open3DQSAR uses SRD for variable grouping, based on the closeness of variables in 3D space, as well as two different variable selection algorithms, namely: Fractional Factorial Design-based variable selection (FFD-SEL), and the Uninformative Variable Elimination-based partial least square (UVE-PLS). The predictive quality of the 3D-QSAR-based PLS models generated from the compounds was examined using the same statistical metrics as the ones referred to above for the 2D-QSAR models.[19]

RESULTS AND DISCUSSIONS

2D-QSAR modelling

Two different methods namely SQS-MLR and GA-MLR were applied separately for feature selection to construct 2D-QSAR linear models. Summary of statistical results obtained in the 2D-QSAR models developed with wild- and mutant-FLT3 inhibitors are presented in Table 1 and Table 2 respectively.

Table 1: Summary of the results obtained from 2D-QSAR analyses performed with selected interpretable descriptors, 2D descriptors (0D-2D) and all (0D-3D) descriptors.

Method	Interpretable descriptors			All 2D Descriptors			All 2D & 3D descriptors		
	Scoring	Q^2_{LOO}	R^2_{Pred}	Scoring	Q^2_{LOO}	R^2_{Pred}	Scoring	Q^2_{LOO}	R^2_{Pred}
SFS-MLR	R^2	0.926	0.843	R^2	0.921	0.757	R^2	0.934	0.817
	NMAE	0.917	0.823	NMAE	0.911	0.921	NMAE	0.874	0.926

	NMPD	0.925	0.842	NMPD	0.920	0.791	NMPD	0.934	0.817
	NMGD	0.925	0.840	NMGD	0.920	0.790	NMGD	0.932	0.766
GA-MLR	-	**0.913**	**0.968**	-	**0.934**	**0.942**	-	**0.919**	**0.965**

It can be observed from Table 1 (wild type) and Table 2 (mutant type) that starting from interpretable descriptors the overall predictivity of the models failed to increase when more complex descriptors were included to generate the models. Therefore, the models generated with interpretable descriptors are likely to provide both high statistical qualities and mechanistic interpretation. GA-MLR generated the most predictive models with interpretable descriptors in both wild type and mutant type model. In case of the wild type interpretable descriptor model Q^2_{LOO} and R^2_{Pred} values appeared as 0.913 and 0.968 while in Mutant type interpretable descriptor model Q^2_{LOO} and R^2_{Pred} values appeared as 0.966 and 0.866, signifying that the models were not overfitted. These GA-MLR models were thus further analyzed to understand their statistical robustness and these were also used to understand structural requirements of the models.

Table 2: Summary of the results obtained from 2D-QSAR analyses performed with selected interpretable descriptors, 2D descriptors (0D-2D) and all (0D-3D) descriptors.

Method	**Interpretable descriptors**			**All 2D Descriptors**			**All 2D & 3D descriptors**		
	Scoring	**Q^2_{LOO}**	**R^2_{Pred}**	**Scoring**	**Q^2_{LOO}**	**R^2_{Pred}**	**Scoring**	**Q^2_{LOO}**	**R^2_{Pred}**
SFS-MLR	R^2	0.964	0.769	R^2	0.930	0.82	R^2	0.980	0.646
	NMAE	0.955	0.7498	NMAE	0.932	0.809	NMAE	0.979	0.649
	NMPD	0.964	0.769	NMPD	0.930	0.76	NMPD	0.980	0.646
	NMGD	0.964	0.7691	NMGD	0.930	0.763	NMGD	0.980	0.646
GA-MLR	-	**0.966**	**0.866**	-	**0.943**	**0.922**	-	**0.956**	**0.941**

Detailed statistical results of these two models are presented in Table 3 and the observed vs predicted plots of these models are shown in Figure 1.

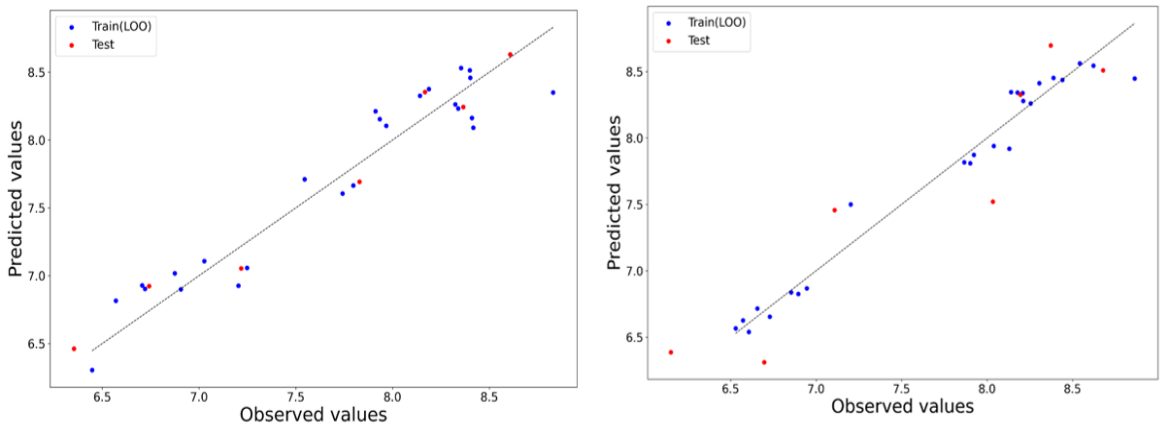

Figure 1: Plots of observed vs predicted activity of 2D-QSAR (A) Model-1 and (B) Model-2.

Table 3: Equations and statistical results of the final 2D-QSAR models.

Model	St-QSAR model	Statistical results
1.(mutant Type)	PIC50 = 6.1468(+/-0.2006) +0.184(+/-0.0234) F02[C-N] -0.8829(+/-0.1275) nRNHR -0.4436(+/-0.092) nR06 -0.0315(+/-0.0225) CATS2D_05_AL +0.3717(+/-0.0312) CATS2D_08_LL	$N_{training}$ = 25, *R2 :0.97*, R^2_{Adj} = 0.969, $F(19,5)$ =155.4, Q^2_{LOO} = 0.966, *MAE* = 0.078, $r_m{}^2_{LOO}$ = 0.952, $\Delta r_m{}^2_{LOO}$ = 0.0094, N_{test} = 7, R^2_{Pred} = 0.866, $r_m{}^2_{test}$ = 0.822, $\Delta r_m{}^2_{test}$ = 0.037
2.(wild type)	PIC50(-log10{BA/10^9} = 6.422(+/-1.8606) +2.8925(+/-0.4988) MDEC-34 -0.2994(+/-0.1535) AMW -1.0358(+/-0.2789) MDEC-24 -0.4597(+/-0.1122) B03[N-O] +0.094(+/-0.044) CATS2D_07_DL	$N_{training}$ = 25, *R2 : 0.94*, R^2_{Adj} = 0.92, $F(18,5)$ =31.727, Q^2_{LOO} = 0.913, *MAE* = 0.161, $r_m{}^2_{LOO}$ = 0.87, $\Delta r_m{}^2_{LOO}$ = 0.0725, N_{test} = 7, R^2_{Pred} = 0.968, $r_m{}^2_{test}$ = 0.931, $\Delta r_m{}^2_{test}$ = 0.027

($N_{training}$: Number of data-points present in the training set, N_{test}: Number of data-points present in the test set)

Apart from ensuring satisfactory internal and external predictivity, a number of studies are required to be performed to check the suitability of 2D-QSAR models. Intercorrelation among descriptors was checked to find that Model-1 and Model-2 have maximum intercollinearity of 0.777 and 0.677, respectively. These values suggest that the descriptors of each regressionbased model are independent of each other. Next, cR_p^2 values of these Model 1 and 2 were found as 0.876 and 0.844, respectively indicating both models were unique in nature.

The applicability domain of these models was explained by Williams plots and these are presented in Figure 2. In Model-1, no structural outlier was obtained whereas this model was

found to have one response outlier. On the other hand, Model-2 contained one structural outlier and one response outlier. No compound was however removed since the structural outlier found on the model was predicted satisfactorily by the model. The relative significance of the descriptors of Model 1 and 2 are presented in Figure 3 with respect to their standardized coefficients.

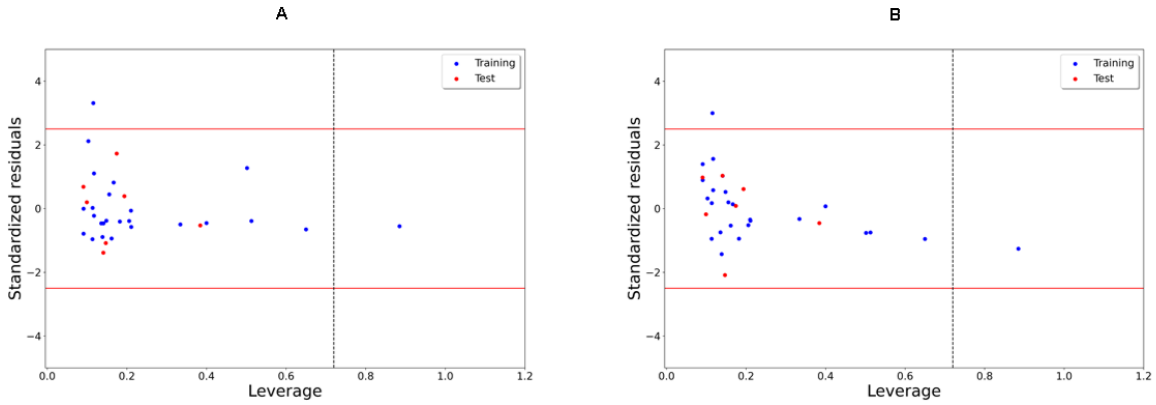

Figure 2: Williams plot describing the applicability domains of (B) Model-1 and (A) Model2.

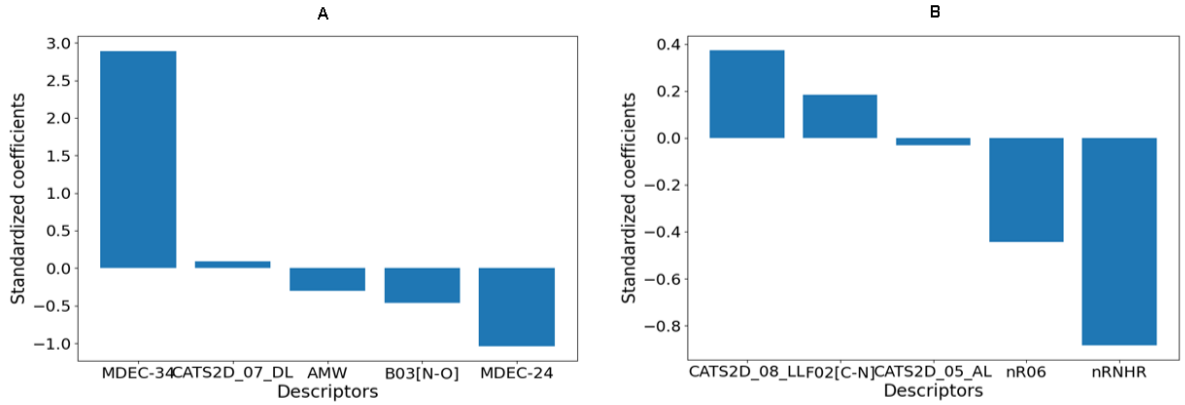

Figure 3: The relative significances of the molecular descriptors appeared in the 2D-QSAR models.

Model 2 (A) for the 2D-QSAR model[22] generated with the wild-type FLT3 kinase, MDEC-34 appeared as the most significant model. This descriptor signifies molecular distance edge between all tertiary and quaternary carbons.[23] The positive correlation between MDEC-34 and response variable indicates that higher value of this descriptor is favoured for higher biological potency against wild-type FLT3 kinase. For example, the most potent compound SN32 (pIC_{50}=8.830) has MDEC-34 value of 1.946 whereas the lowest active compound

(pIC_{50}=6.354) has MDEC-34 value of 1.639. Another molecular distance edge descriptor named MDEC-24 was found as second most influential descriptor of the model. This descriptor stands for molecular distance edge between all secondary and quaternary carbons. However, in contrary to MDEC-34, MDEC-24 is negatively correlated to response variable and therefore low value of this descriptor is required for higher biological activity. One of the most potent compounds **SN31** (pIC_{50}=8.607) has a value of 1.486 whereas one of the least potent compound **SN3** (pIC_{50}=6.570) has a value of 1.768 for this parameter.

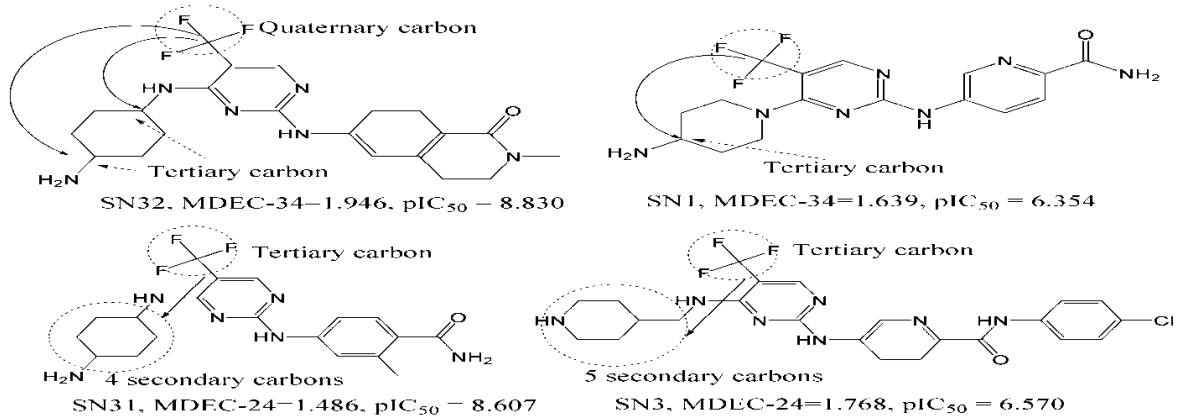

Figure 4: The explanations of descriptors MDEC-34 and MDEC-24 with respect to highly active and low active derivatives.

Ring descriptor nr06 (number of 6-membered rings) appeared as the third most significant descriptor of the model and this descriptor was found to be negatively correlated with the biological activity. The third most contributing descriptor of the model is 2D-atom pairs descriptor named B03[N-O] (Presence/absence of N – O at topological distance 3). This binary descriptor counts the presence of topological distance between nitrogen and atom. Since this descript is also negatively correlated it clearly suggests that higher value of this descriptor is found in less active compounds. In Figure X, the explanation of this descriptor is presented. Average molecular weight was also found as a negatively correlated descriptor and it signifies that higher average molecular weight is detrimental to higher biological activity. The last descriptor CATS2D_07_DL belongs to Chemically Advanced Template Search (CATS)[24] descriptor that counts the number of hydrogen bond donor and lipophilic features located at a topological distance of 7. Higher value of this descriptor was found in the higher active compounds.

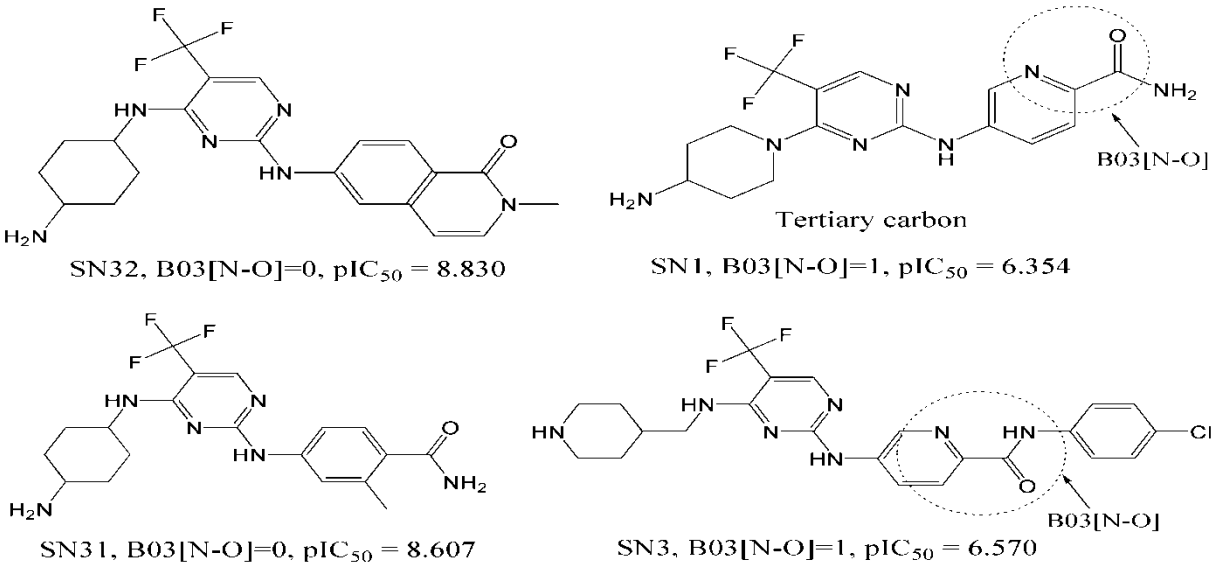

Figure 5: The explanations of descriptors B03[N-O] with respect to highly active and low active derivatives.

The most significant descriptor of the 2D-QSAR model for mutant FLT3 was found as nRNHR, which is a functional group counts descriptor signifying number of secondary amines (aliphatic). The higher value of this descriptor is found to be detrimental for higher activity. The second contributing descriptor of the model is nR06 that is a ring descriptor representing the number of six-membered ring in the compounds. Higher value of this descriptor was found as a negative contributor towards higher biological activity. The third most influential descriptor of the model is CATS2D_08_LL, which stands for topological distance of 8 between two lipophilic features. This is the most significant descriptor which is positively correlated to biological activity. The higher value of the descriptor was often among highly active compounds of the dataset. Another descriptor which is found in the model F02[C-N] that stands for Frequency of C – N at topological distance 2 , this descriptor is also positively correlated to biological activity indicating that high value of the descriptor is favorable for higher Biological activity. Finally CATS2D_05_AL descriptor appear as negatively correlated to the biological activity. This descriptor signifies topological distance of 5 between a lipophilic and a hydrogen bond acceptor feature.

3D-QSAR analysis

In order to further understand how these compounds may interact with the receptor, 3D-QSAR analysis was performed with 32 compounds that was aligned by rigid body molecular alignment method. It is well known that unlike 2D-QSAR, 3D-QSAR methodology largely depends on the bioactive conformers of the ligands and their alignment. The aligned

conformations were randomly divided into 25 training and 7 test set compounds and the 3D-QSAR model was generated using two techniques – (a) FFD-SEL and (b) UVE-PLS. The statistical qualities of the 3D-QSAR models developed with wild- and mutant-type FLT3 are depicted in Table 4, from which it may be inferred that the aligned conformations satisfactorily characterize the experimental activity exhibited by the compounds present in the current dataset. Evidently, UVE-PLS technique was more successful in deducing more predictive 3D-QSAR model as compared to the FFD-SEL method. For example, the wild-type model was produced with Q^2_{LOO} of 0.724 and R^2_{Pred} of 0.762. Below, we explain the 3D-QSAR model developed with wild-type FLT3 inhibitors with respect to the generated electrostatic and steric contour maps (Figure 4 and 5, respectively).

Table 4: Statistical results of 3D-QSAR models.

Parameter	Wild type		Mutant-type	
	FFD-SEL	UVE-PLS	FFD-SEL	UVE-PLS
PC	3	2	3	2
N_{Tr}	25	25	25	25
F-test	207.817	195.846	199.15	214.55
R^2/SDEP	0.967/0.125	0.946/0.160	0.966/0.138	0.951/0.165
Q^2LOO/SDEP	0.687/0.389	0.724/0.365	0.841/0.298	0.823/ 0.315
Q^2_{LTO} /SDEP	0.683/0.392	0.723/0.366	0.834/0.305	0.819/0.318
Q^2_{LMO}/SDEP	0.705/0.375	0.754/0.046	0.814/0.322	0.803/0.0213
N_{Ts}	7	7	7	7
R^2_{Pred}/SDEP	0.851/0.307	0.762/0.389	0.785/ 0.412	0.767/0.43

The contributions of steric and electrostatic fields found in the best 3D-QSAR models were 0.49 and 0.51, respectively. Therefore, steric effects mainly contributed in the model development. For Figure 4 in the most active compound (B) there are a Nitrogen atom (N) in the structure and N is an electronegative atom and the other hand for the least active compound (C) there are nitrogen atom (N) present in the structure but that inserted into electropositive field. For most of the active compounds, the electronegative residues are drifted away from electropositive field. Secondly for the best active compound there are no electronegative compound present in the structure but in the other hand in least active compound there is nitrogen atom which is again a electronegative compound and it is close to electropositive field. Thirdly, for the best compounds more electron rich aromatic ring is inserted into electronegative field and for the least active compound less electron rich aromatic compound ring is inserted into electronegative field.

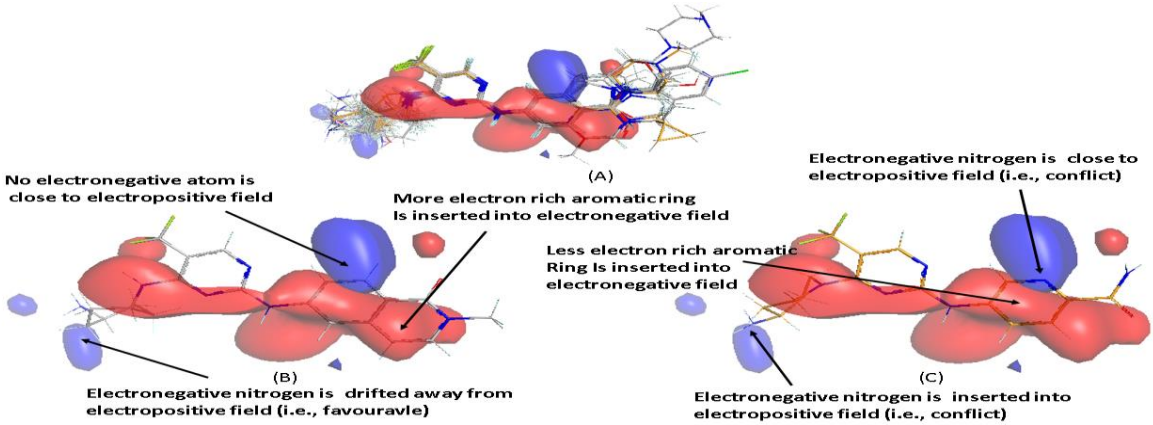

Figure 4. Contour maps obtained from the best 3D-QSAR model (Blue: Electropositive favorable; Red: Electronegative favorable). (A) aligned conformations of all compounds, (B) the best active compound SN32 and (C) the least active compound SN1.

When we comparing the difference between best active compound and the least active compound, we found that the best active compound has a cyclohexyl chain and that chain is properly inserted into the steric favorable moiety. On the other hand for least active compound piperidylamine not properly inserted into steric favorable. Secondly in the most active compound there is a aromatic group which is more close to the steric favorable group that's why it is more active and in the other hand there are no such groups present close to the steric favorable group.

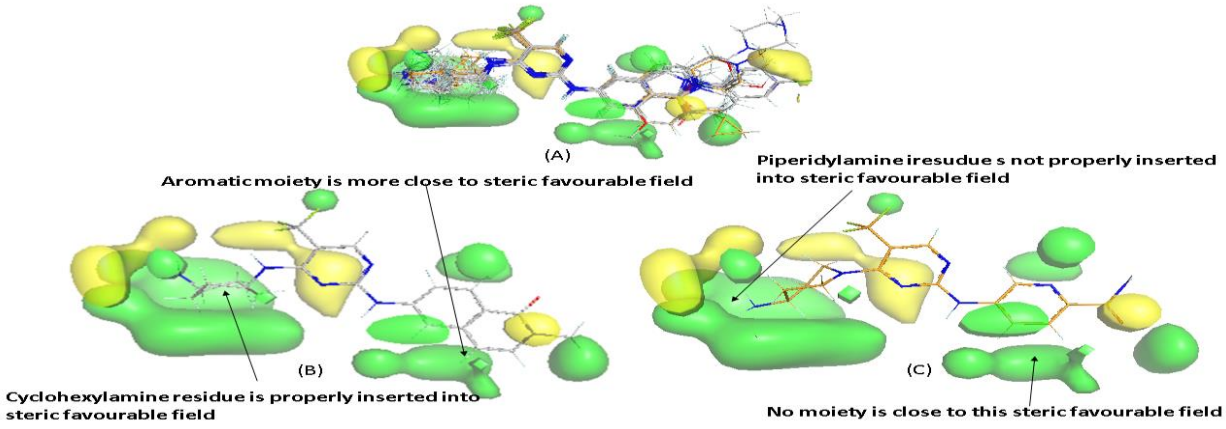

Figure 5. Steric favorable fields of the model where Green for steric favorable fields and yellow is for steric unfavorable [(A) All compounds, (B) the most active compound, (C) the least active compound].

CONCLUSION

In the current investigation, various *in silico* analyses have been performed one by one in a systematic manner to understand the structural requirement of 2‑Aminopyrimidine Derivatives as anti-cancer agents. Within the scope of this work, we attempted to understand possible binding mechanism of these compounds against FLT3 Kinase Inhibitors with High Selectivity over c-KIT. Even though the compounds of the dataset had high structural similarities, these were found to have large variations in their enzyme inhibitory activities. The current investigation was invested to systematically explain the variations obtained in the anticancer activity of these compounds with respect to their structural attributes. In this work, we have performed 2D QSAR and 3D QSAR modelling to know the structural requirements of the possible compound which may have a strong anticancer property against Acute Myeloid Leukemia.[25] The future scope of this investigation includes structure-based design to understand whether the receptor-ligand interactions comply with the information obtained from the ligand-based design. In order to do so, compounds may be docked with the homology models of wild-type FLT3 and mutant-type FLT3 provided by Tong et al.[9] Structure-based pharmacophores developed with the protein complexes may further help in obtaining pharmacophore aligned structures to be used for 3D-QSAR model development. Lastly, MD simulations may be carried out to develop dynamic behaviors (i.e., binding stability) of these compounds against the proposed homology model. Overall, the current work may be used as important guidelines to design novel FLT3 kinase inhibitors for the treatment of AML.

REFERENCES

1. Chopra, M. and S.K. Bohlander, *The cell of origin and the leukemia stem cell in acute myeloid leukemia.* Genes, Chromosomes and Cancer, 2019; **58**(12): p. 850-858.
2. Siveen, K.S., S. Uddin, and R.M. Mohammad, *Targeting acute myeloid leukemia stem cell signaling by natural products.* Molecular Cancer, 2017; **16**(1).
3. Swaminathan, M. and E.S. Wang, *Novel therapies for AML: a round-up for clinicians.* Expert Review of Clinical Pharmacology, 2021; **13**(12): p. 1389-1400.
4. Döhner, H., et al., *Acute Myeloid Leukemia.* New England Journal of Medicine, 2015; **373**(12): p. 1136-1152.
5. Lam, S.S.Y. and A.Y.H. Leung, *Overcoming Resistance to FLT3 Inhibitors in the Treatment of FLT3-Mutated AML.* International Journal of Molecular Sciences, 2020; **21**(4).

6. Daver, N., S. Venugopal, and F. Ravandi, *FLT3 mutated acute myeloid leukemia: 2021 treatment algorithm.* Blood Cancer Journal, 2021; **11**(5).
7. Eguchi, M., et al., *Mechanisms Underlying Resistance to FLT3 Inhibitors in Acute Myeloid Leukemia.* Biomedicines, 2020; **8**(8).
8. Zhao, L., et al., *Advancing computer-aided drug discovery (CADD) by big data and data-driven machine learning modeling.* Drug Discovery Today, 2020; **25**(9): p. 1624-1638.
9. Tong, L., et al., *Identification of 2-Aminopyrimidine Derivatives as FLT3 Kinase Inhibitors with High Selectivity over c-KIT.* Journal of Medicinal Chemistry, 2022. **65**(4): p. 3229-3248.
10. Mauri, A., *alvaDesc: A Tool to Calculate and Analyze Molecular Descriptors and Fingerprints*, in *Ecotoxicological QSARs*. 2020. p. 801-820.
11. Sushko, I., et al., *Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information.* Journal of Computer-Aided Molecular Design, 2011. **25**(6): p. 533-554.
12. Halder, A.K., A.H.S. Delgado, and M.N.D.S. Cordeiro, *First multi-target QSAR model for predicting the cytotoxicity of acrylic acid-based dental monomers.* Dental Materials, 2022; **38**(2): p. 333-346.
13. Ambure, P., et al., *"NanoBRIDGES" software: Open access tools to perform QSAR and nano-QSAR modeling.* Chemometrics and Intelligent Laboratory Systems, 2015; **147**: p. 1-13.
14. Halder, A.K., *Finding the structural requirements of diverse HIV-1 protease inhibitors using multiple QSAR modelling for lead identification.* SAR and QSAR in Environmental Research, 2018; **29**(11): p. 911-933.
15. Tetko, I.V., V.Y. Tanchuk, and A.E.P. Villa, *Prediction of n-Octanol/Water Partition Coefficients from PHYSPROP Database Using Artificial Neural Networks and E-State Indices.* Journal of Chemical Information and Computer Sciences, 2001; **41**(5): p. 1407-1421.
16. Pratim Roy, P., et al., *On Two Novel Parameters for Validation of Predictive QSAR Models.* Molecules, 2009; **14**(5): p. 1660-1701.
17. Ojha, P.K. and K. Roy, *Comparative QSARs for antimalarial endochins: Importance of descriptor-thinning and noise reduction prior to feature selection.* Chemometrics and Intelligent Laboratory Systems, 2011; **109**(2): p. 146-161.
18. Sahigara, F., et al., *Comparison of Different Approaches to Define the Applicability Domain of QSAR Models.* Molecules, 2012; **17**(5): p. 4791-4810.

19. Tosco, P. and T. Balle, *Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields.* Journal of Molecular Modeling, 2010; **17**(1): p. 201-208.
20. Tosco, P., T. Balle, and F. Shiri, *Open3DALIGN: an open-source software aimed at unsupervised ligand alignment.* Journal of Computer-Aided Molecular Design, 2011; **25**(8): p. 777-783.
21. Verma, J., V. Khedkar, and E. Coutinho, *3D-QSAR in Drug Design - A Review.* Current Topics in Medicinal Chemistry, 2010; **10**(1): p. 95-115.
22. Fang, C. and Z. Xiao, *Receptor-based 3D-QSAR in Drug Design: Methods and Applications in Kinase Studies.* Current Topics in Medicinal Chemistry, 2016; **16**(13): p. 1463-1477.
23. *Handbook of Molecular Descriptors*. Methods and Principles in Medicinal Chemistry. 2000.
24. Reutlinger, M., et al., *Chemically Advanced Template Search (CATS) for Scaffold-Hopping and Prospective Target Prediction for 'Orphan' Molecules.* Molecular Informatics, 2013; **32**(2): p. 133-138.
25. Bullinger, L., K. Döhner, and H. Döhner, *Genomics of Acute Myeloid Leukemia Diagnosis and Pathways.* Journal of Clinical Oncology, 2017; **35**(9): p. 934-946.