Original Article

Fingerprint Based Two-dimensional QSAR Studies on Pyrazolo Quinazolines as CDK2 Inhibitors: A Rational Approach for the Design of Novel Anticancer Agents

Aravinda Pai, Jayashree B.S*

Department of Pharmaceutical Chemistry, Manipal College of Pharmaceutical Sciences, Manipal Academy of Higher Education, Manipal, Karnataka, INDIA.

ABSTRACT

Aim/Background: Quantitative structure activity relationship studies are important ligand-based methods used in drug discovery process. The development of Quantum mechanics and its application in the study of smaller and macromolecules have accelerated their applications in the field of drug discovery. Extensive developments in QSAR studies (either 2D or 3D) could be the result of progressive applications of Quantum mechanics. Latest drug discovery modules depends solely on the advanced computational methods such as 3D QSAR and molecular docking approach. It is believed that a few drug discovery groups from industries still depend on the 2-dimensional QSAR techniques as a part of the lead optimization. Our study aims on learning the application of five important binary fingerprinting techniques based on the quantum mechanics involving 2D QSAR studies for designing novel pyrazolo quinazolines as selective inhibitors of CDK2/CyclinA. **Materials and Methods:** A dataset of 47 analogues of pyrazolo quinazolines were selected with their inhibitory activity on CDK2/Cyclin A. The derivatives were divided into training and test sets. The Kernel based partial regression was run using five important binary fingerprints and statistical significance of each fingerprint was analysed. **Results:** Out of the five fingerprints selected, the fingerprint linear arrived at the optimized 2D QSAR model through the kernel-based regression analysis. The final developed model expressed the importance of the presence of Carboxamide groups on the pyrazole ring that could positively contributes to the inhibitory activity. **Conclusion:** The developed model could be of use to design better analogues with enhanced selectivity and specificity as inhibitors of CDK2/Cyclin A that would provide a clear insight amongst the researchers for the development of novel and potent clinically useful anticancer agents.

Key words: 2D QSAR, Binary fingerprints, Features, Similarity, Kernels, Statistical.

INTRODUCTION

Drug discovery is of concern in the present time for both designing and developing a low molecular weight drug candidate that could target many biological events. The whole process of new drug discovery would require several years and is an expensive affair.[1] Furthermore, only a small proportion of the New Chemical Entity (NCE) get approval from the regulatory agencies and will be eventually brought to the market.[2] This process involves several preclinical as well as clinical trials involving the active role of many clinicians. In the preclinical stage, the main strategies involve the identification and validation of the new drug candidates, identification of the active compounds (hits) and the final transformation of the hits into the lead compounds which could be later on optimized by using lead optimization techniques.

The generation of the hits arises from one of the important screening techniques, High-throughput Screening (HTS).[3] HTS involves a collection of test compounds against a biochemical or cellular targets.[4] Compounds exhibiting a positive response

Submission Date: 03-12-2018;
Revision Date: 29-03-2019;
Accepted Date: 27-06-2019

DOI: 10.5530/ijper.53.3s.100

Correspondence:
Dr. B S Jayashree,
Professor, Department of Pharmaceutical Chemistry, Manipal College of Pharmaceutical Sciences, Manipal Academy of Higher Education (MAHE), Manipal-576 104, Karnataka, INDIA.
Phone: +91-0820-2234567
E-mail: jayashree.sy@gmail.com

www.ijper.org

in screening is considered to be the most suitable hit compound. In the subsequent screening, these hits are concomitantly analyzed based on both the physico-chemical and pharmacological parameters and are later evaluated for their potential to eventually to come out as a lead compound. Over all, the HTS data has their own limitations owing to the generation of both false-positive and false-negative activity outcomes. Typically, a large number of false-positives enables future experiments to carry their importance. However, false-negative measurements become less significant as a consequence of their limited purity and stability in low concentrations during assay.[5] The solution for the compensation for such limitations is facilitated by computational approaches as an alternative tool. Virtual Screening (VS) is a rational, computational, "time-efficient and cost-effective" method analogous to HTS. For instance, VS, a large compound library is build and screened *in silico* against a known set of drug candidates of interest. It is a well reported.[6] tool in modern drug discovery. Many reports.[7] have supported the study and facilitated for relating various 1D, 2D and 3D representations in identifying active hits from the database. Conversely docking[8] method is preferred only when X-ray crystallographic data are available. Ligand-based methods, such as pharmacophore mapping,[9] shape-based screening.[10] and 2D fingerprint similarity.[11] could be the only viable approach when X-ray protein crystallographic structures are not made available. Ligand-based methods are preferred when, the number of test compounds are in large number. In reality, the choice of the virtual screening methods.[12,13] is solely dependent on the data set under consideration. 2D fingerprints are the most robust and popularly used tool in new drug discovery and are of extensive usage in retrieving the hit compounds when compared to that of 3D and docking methods. It provides a platform for running fragmentation studies including atom or bond typing schemes, rules based on bit scaling and various similarity indices. For instance, the cheminformatics package Canvas.[14] utilizes more than 25,000 arrangements of the above four variables. Programmed method of selecting combinations of such variables[15] have paved the way for generating practical guidelines in 2D fingerprint queries. Both the sizes of the datasets and the complexity of the interrelations among variables have precluded a detailed analysis of the type of hits. The present investigation could emerge out as the most appropriate tool in predicting as to how the hits could exhibit specific molecular attributes, chemical diversification and other characteristics apart from the activity itself.

In the present study, we report the application of five important binary fingerprinting techniques in the design of potent CDK2/CYCLIN A inhibitors using kernel based partial least square regression methods. The most ideal model were analyzed based on their regression values. Based on the colour coding, the most appropriate modification sites that could enhance the activity were proposed.

Much of the work on the computational studies of cell cycle related topic has not been fully exploited for many years even though independently substantial work of progressive development has taken place both in computational science as well as in oncology research. In our present study, we are attempting to relate these two subjects of interest. Cell cycle is measured in terms of cell proliferation, growth and cell division after following DNA damage. Its function is to monitor the transition from quiescence (G0) to cell proliferation and through its checkpoints, ensuring the fidelity of the genetic transcript. This mechanism facilitates in cell division and is divided into four phases. The time gap associated with DNA synthesis (S phase) and mitosis (M phase) are separated by gaps of different length called G1 and G2. Progression through the cell cycle is facilitated by a number of Cyclin Dependent Kinases (CDKs) that are complexed with cell cycle specific regulatory proteins called cyclins that facilitate the cell cycle progression. CDKs are a family of serine–threonine kinases that plays a central role in the regulation of cell cycle progression through different phase.[16] They are regulated by the process of phosphorylation and activated by their association with corresponding Cyclin partners.[17] CDK inhibitors have been extensively exploited for the development of potential anti-cancer agents.[18] Individual CDKs tend to phosphorylate the respective substrates at different phases of the cell cycle and are classified as G1 (CDK4 and CDK6-D cyclins, CDK2-cyclin E), S (CDK2-cyclin A, CDK1-cyclin A) and G2/M (CDK1-cyclin B) phase-specific CDKs.[19] The defective functioning of CDKs compromises the normal cell cycle progression.[20] When the cells enter S phase, CDK2/Cyclin A gets phosphorylated and consequently deactivates E2F (Transcription gene in the eukaryotic cell) that deregulation and increase in the levels of E2F transcriptional activity eventually, leads to the cell death mediated by apoptosis. It is noteworthy to mention that, the inhibition of CDK2 could play a crucial role in optimizing the therapeutic benefits mediated by tumor cell apoptosis.[21] Several small molecules of CDK inhibitors are currently under the developmental stage in new drug discovery including the first generation CDK inhibitor, flavopiridol. Flavopiridol is

Binary fingerprint -atom triplets

Description: Three atoms and the corresponding distances separating them. It is nearly an extrapolation of the fingerprint atom pairs. A set of three atoms constitutes the triplets and the distance separating them. There are a possibility arriving of six different patterns to order the atoms in a triplet resulting in a canonicalization so as to ensure that, no further patterns get separated as bits in the fingerprint. $Type_a$-d_{ab}-$Type_b$-d_{bc}-$Type_c$-$d_{ca.}$ The assessment method is same which is given under atom pairs.

Binary fingerprint –fp linear

Description: Linear fragments + ring closures

A structure for all linear paths restricted to a user-defined number of bonds are distinctly seen and seven by default. The hashing operation is done based on the strings. The description for each linear fragment is dependent on the mode in which it generates an integer bit address. In order to enhance the performance of the hazing pattern of the rings, that is not affecting the number of fragments generated, the default maximum path is invariably expanded from 7 to 14 exclusively for ring closure. This process encodes the information in and around many ring systems that produce only a fraction of the fragments where, it generates only that path which contains 0-14 bonds. Here again, the assessment method is same which is given under atom pairs.

Binary fingerprint –fp radial

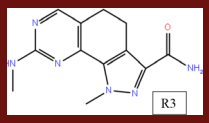
Description: Radial fingerprints are also termed as extended connectivity fingerprints. It considers the surroundings of an atom in terms of its neighboring atoms scaled up to a radius of two bonds. These fingerprints finds extensive applications in structural similarity searching. The radial connectivity of an atom is computed using an advanced version of Morgan algorithm. The extended connectivity could be increased to the multiple bond order, in order to make them more discriminative. A coded atom could be combined with the codes of other neighbouring atoms in order to achieve higher order description. Combined with the codes of its neighbours to the required level of description.

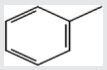
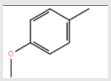
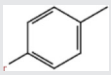
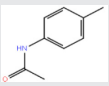
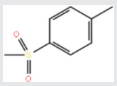
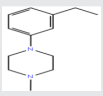
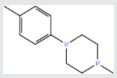
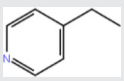
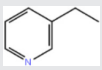
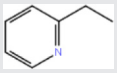
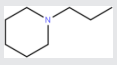
Binary fingerprint –fp dendritic

Description: Linear and branched fragments

The linear and branched features are encoded when, the structure gets separated as fragments containing both linear paths and their intersections possessing a default maximum of five bonds per path. It is worthwhile to note that, dendritic fingerprints incorporate no special

Table 1: Structure and activity of the selected compounds from the literature.



Compound	R3 substitution	CDK2/Cyclin A Activity in micromoles	pact
1 (pq1)		0.051	7.292
14 (pq2)		0.066	7.180
15(pq3)		0.058	7.237
16 (pq4)		0.081	7.092
17(pq5)		0.013	7.886
18 (pq6)		0.038	7.420
19 (pq7)		0.104	6.983
20 (pq8)		0.017	7.770
21 (pq9)		0.021	7.678
22 (pq 10)		0.021	7.678
23 (pq 11)		0.087	7.048
24 (pq 12)		0.209	6.666
25 (pq 13)		0.019	8.037
26 (pq 14)		0.490	5.604
27 (pq 15)		0.004	8.398
28 (pq 16)		0.149	6.827
29 (pq 17)		0.002	8.669

a synthetic flavonoid and is well recognized as a CDK inhibitor owing to its high affinity for CDKs and having, ability to induce cell cycle arrest in a large number of cancer cell lines.[22] Gabriella, *et al*.[23] reported the CDK2/CYCLIN A inhibitory activity of novel pyrazolo quinazolines. The present work is thus based on the extensive literature reports that are available in exhibiting promising anticancer activity against the target enzyme CDK2/CYCLIN A.

MATERIALS AND METHODS

Data set selection

The derivatives of pyrazolo quinazoline are reported as potent CDK2/CYCLIN A inhibitors (48 compounds). In our present simulation study,[23] all the 48 possible structures of pyrazolo quinazolines were drawn and set for energy minimization process with Merck Molecular Force Field (MMFF) using Maestro interface from Schrödinger. The IC_{50} values (uM) were transduced into pIC_{50} ($-\log IC_{50}$) values. The compounds were selected from all the activity ranges with difference of 3 log order and was appropriate for a QSAR study.[24] The dataset was divided into a training set for model development and test set for validation of the external predictivity, containing 33 and 15 compounds respectively. Compounds were classified as high, medium and low active ones based on the data procured. Precautions were taken so that, the compounds from all the three categories of biological activities were present. Compound exhibiting highest activity based on the docking score were taken as a prototype for rest of the other compounds. The compounds selected for the study are represented in Table 1.

Molecular docking studies

Schrödinger software[25] was used to study the protein-ligand interactions using co-crystallized ligand (PDB code: 3 EID, Pyrazolo pyridazine). The refinement of the crystal structure enabled for re-docking the same into the active site of the protein in order to validate the docking procedure.[26] The Root Mean Square Deviation (RMSD) was found to be 0.200 Å. The residues of the co-crystallized inhibitor was considered to be the most apt binding site for the study and was in the range of 5 Å. Simultaneously, the protein structure was prepared using protein preparation wizard and also the ligand preparation was achieved by ligprep option available with Schrödinger software. The grid was created using the x, y and z coordinates of the active site and extended to 10 ×10×10 points with a grid spacing of 0.270 Å. The best pose conformation were clustered with a default 0.5 Å Root-Mean-Squared Deviations (RMSDs).

Two dimensional QSAR studies using binary fingerprints

Importing the minimum energy conformations with their activity

The two-dimensional chemical structures were imported from Maestro interface using the option import structures. All the imported structures were checked for valence errors, invalid structural representation for an accurate model derivation.

Selection and incorporation of molecular descriptors

The molecular descriptors were incorporated using Canvas 2.9 interface. There are mainly four types of descriptors available with Canvas namely, physico-chemical, topological, ligfilter and Qik prop descriptors. There are few important descriptors listed under each main class of descriptors.

Physico chemical descriptors- Molecular weight, ALogP, Molar refractivity and polar surface area.

Topological descriptors-1 path kier alpha, ALogP10, Average eccentricity and Balaban centric.

Ligfilter descriptors-Molecular weight, Number of aliphatic rings, number of aromatic rings and charged acceptor groups.

Qik prop descriptors-computational method mainly based on PM3 properties.

Incorporation of Binary fingerprints

The available five binary fingerprints were added and incorporated using the option binary fingerprints available with the Canvas 2.9. Those models having significant contribution could only be retained for further process. Kernel based partial least square regression using different binary fingerprints.

Binary fingerprint -atom pairs

Description: pairs of atoms is segregated by the type and distance separating them.[27] The fingerprint decides the principle of atom pairs (pairs of atoms) and the distance separating them: $Type_a$-$Type_b$-d_{ab} where, $Type_a \leq Type_b$. The representation is hashed into an integer value and 'd' is the shortest distance between any pair of atoms. The actual contribution of the model as a function of the activity is assessed by kernel based partial least square equation by using regression values (R^2 and Q^2 values). The atoms influencing the model either positively or negatively is assessed by the colour code available on different atoms.

treatment for the rings. Thus, the assessment method remains the same as given under atom pairs.

Model validation

It is an essential component in any QSAR study in order to validate the generated model. For the validation purpose, the total molecules were divided into the training and test sets. The training set was used to generate the model whereas, the test set was used to validate the generated QSAR model. Based on the correlation between the actual and the predicted activity in test sets, the linear regression value was taken as a standard parameter in order to assess the robustness of the generated model. Also, the difference in q^2 and r^2 in all the fingerprint models were taken as a standard parameter to assess the reliability of the generated models. For all the 4 significant 2D QSAR models with the difference between q^2 and r^2 values, were well within the limits (below 0.2).

RESULTS AND DISCUSSION

Molecular docking studies

Molecular docking studies revealed prominent interactions of ligand groups with the key amino acids at the active site of CDK2/Cyclin A. The crystal ligand (pyrazolo pyridazine) shows hydrogen bond acceptor interactions with the active site amino acids leucine 83 and lysine 33, also leucine 83 showed hydrogen bond donor interactions with the secondary amino group of the crystal ligand. The terminal alkyl chain of crystal ligand showed hydrogen bonding donor, acceptor interactions with the active site amino acid aspartic acid 86. The crystal ligand 3D interactions are shown in Figure 1. The 2D interaction of the crystal ligand at the active site is represented in Figure 2.

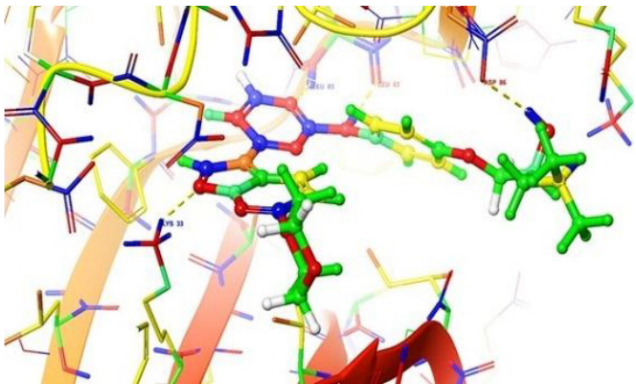


Figure 1: Interaction of Pyrazolo pyridazine at the active site of CDK2/CYCLIN A.

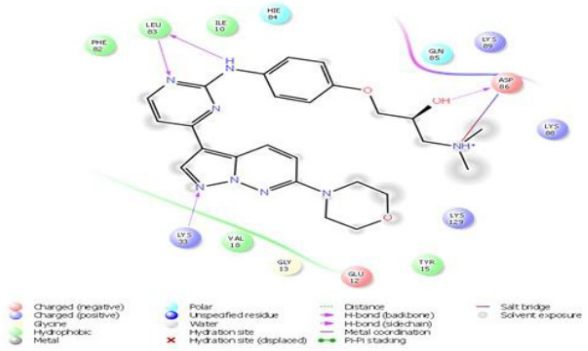


Figure 2: 2D interaction diagram of Pyrazolo pyridazine at the active site of CDK2/CYCLIN A.

Also, the most active analogue pq 36 showe prominent interactions at the active site of CDK2/Cyclin A.

The analogue pq 36 showed hydrogen bond acceptor interactions with the amino acids leucine 83 and aspartic acid 86. Also, the sulphonamide group of pq 36 showed hydrogen bond acceptor interactions at the active site. The 3D interactions of pq 36 with the active site of CDK2/Cyclin A is given in Figure 3 and 2D interactions of the same is represented in Figure 4. The crystal ligand showed the docking score of -12.464 and the most active analogue pq 36 showed docking score of -11.778. The docking scores were presented in the Table 2.

Results of Kernel-based partial least square regression using Binary molecular fingerprints

2D-QSAR models were generated for the series of substituted pyrazoloquinazolines as CDK2/CYCLIN A inhibitors. Molecular modelling was carried out to correlate the results of 2D QSAR, as these methods are mainly based on the two dimensional structure and atom to atom connectivity unlike 3D QSAR that chiefly depends on the 3 dimensional structure based on steric and vanderwaals radius etc.

Statistical analysis of 2D QSAR model genertaed for the finger print "atom triplets"

The fingerprint atom triplets resulted in a statistically significant 2D QSAR model with reliable regression coefficient values and cross-validation coefficient values as represented in Table 4. The model also showed a good predictive accuracy both in the test and the training set molecules as shown in the Table 3. For the assessment of atomic contribution to the model, three molecules were subsquently taken from each active and inactive set atoms that positively contributed to the activity and were colored green and atoms detrimentally contributing to the model were colored red. Figure 5 shows the colour map of 6 molecules chosen under active and

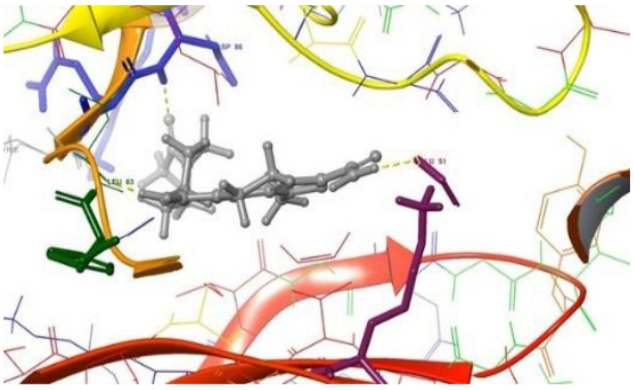


Figure 3: Interaction of the active molecule pq 36 at the active site of CDK2/CYCLIN A.

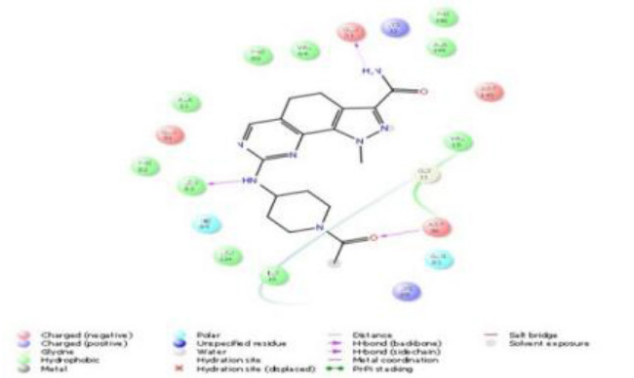


Figure 4: 2D interaction diagram of active molecule pq 36 at the active site of CDK2/CYCLIN A.

inactive molecules. The 2D QSAR model generated is shown in Figure 6. The activity reported vs predicted is represented in the Figures 7 for the test set and in Figure 8 for the training set.

Statistical analysis of 2D QSAR model generated for the finger print "atom pairs"

The fingerprint atom pairs gave a statistically significant 2D QSAR model with an excellent regression coefficient values and cross-validation coefficient values as represented in Table 6. The model also showed good predictive accuracy both in the test as well as in the training set molecules as demonstrated in Table 5. For the assessment of atomic contribution to the model, three molecules were taken from each active and inactive set. Atoms positively contributing to the activity were colored brown whereas, atoms detrimentally contributing to the model were colored blue which is shown in Figure 9. The 2D QSAR model generated is shown in Figure 10. The activity reported vs predicted is represented in the Figure 11 for the test set and Figure 12 for the training set.

Table 2: Docking scores for the molecules taken from literature.

Title	docking score	XP GScore	glide gscore	glide emodel
3eid.final_ligand	−12.464	−12.464	−12.464	110.758
pq36	−11.778	−11.778	−11.778	−62.867
pq5	−11.668	−11.668	−11.668	−68.829
pq28	−11.245	−11.245	−11.245	−54.552
pq17	−11.022	−11.022	−11.022	−45.226
pq26	−10.971	−10.971	−10.971	−56.628
pq15	−10.951	−10.951	−10.951	−54.482
pq25	−10.909	−10.909	−10.909	−69.07
pq34	−10.9	−10.9	−10.9	−56.437
pq42	−10.887	−10.887	−10.887	−56.349
pq44	−10.731	−10.731	−10.731	−63.565
pq29	−10.601	−10.601	−10.601	−59.483
pq8	−10.42	−10.449	−10.449	−64.987
pq16	−10.387	−10.387	−10.387	−58.017
pq38	−10.182	−10.182	−10.182	−71.454
pq30	−9.935	−9.935	−9.935	−63.942
pq40	−9.494	−9.494	−9.494	−68.61
pq32	−9.401	−9.401	−9.401	−81.882
pq14	−9.362	−9.362	−9.362	−60.6
pq19	−8.492	−8.492	−8.492	−55.648
pq43	−8.411	−8.411	−8.411	−64.779
pq20	−8.374	−8.374	−8.374	−50.851
pq13	−8.224	−8.224	−8.224	−54.784
pq33	−8.086	−8.086	−8.086	−46.717
pq21	−8.072	−8.072	−8.072	−55.695
pq35	−7.95	−7.95	−7.95	−56.918
pq7*	−7.856	−7.886	−7.886	−59.517
pq37	−7.841	−7.841	−7.841	−58.622
pq45	−7.625	−7.625	−7.625	−66.931
pq23	−7.152	−7.152	−7.152	−66.832
pq39	−7.088	−7.088	−7.088	−63.207
py27	−6.817	−6.817	−6.817	−82.343
pq9	−6.813	−7.007	−7.007	−69.075
pq4	−6.804	−6.804	−6.804	−57.032
pq11	−6.693	−6.958	−6.958	−59.968
pq22	−6.646	−6.646	−6.646	−65.538
pq46	−6.616	−6.616	−6.616	−79.385
pq10	−6.417	−6.417	−6.417	−72.534
pq6	−6.41	−6.41	−6.41	−83.185
pq1	−6.395	−6.395	−6.395	−59.486
pq3	−6.366	−6.366	−6.366	−63.248
pq2	−6.313	−6.313	−6.313	−69.346
pq18	−5.562	−5.562	−5.562	−56.67
pq47	−5.375	−5.375	−5.375	−79.862
pq12	−5.134	−5.289	−5.289	−62.539
pq41	−5.067	−5.067	−5.067	−72.35
pq24	−4.881	−4.881	−4.881	−69.379

Table 3: Activity prediction tables for the fingerprint atom triplets.

Structure	Model Set	pact	Pred(pact)	Error
pq42	training	9	9.014	0.014
pq44	training	9	9.054	0.054
pq17	training	8.699	8.533	−0.166
pq36	test	8.699	8.386	−0.313
pq45	training	8.699	8.651	−0.048
pq38	test	8.523	8.202	−0.321
pq15	training	8.398	8.445	0.047
pq46	training	8.398	8.395	−0.003
py27	test	8.301	7.826	−0.476
pq29	training	8.301	8.357	0.055
pq47	training	8.301	8.316	0.015
pq26	training	8.222	8.238	0.017
pq32	training	8.155	8.126	−0.029
pq28	test	8.097	8.26	0.163
pq40	training	7.921	7.927	0.006
pq5	training	7.886	7.889	0.003
pq8	training	7.77	7.763	−0.006
pq13	test	7.721	8.037	0.316
pq24	training	7.699	7.691	−0.008
pq9	training	7.678	7.684	0.006
pq10	training	7.678	7.705	0.027
pq23	training	7.638	7.625	−0.014
pq39	test	7.602	7.811	0.209
pq37	training	7.495	7.521	0.026
pq6	training	7.42	7.426	0.006
pq33	training	7.409	7.404	−0.005
pq1	test	7.292	7.354	0.062
pq25	test	7.284	7.587	0.303
pq3	training	7.237	7.216	−0.02
pq43	training	7.237	7.191	−0.045
pq2	test	7.18	7.358	0.178
pq4	training	7.092	7.097	0.005
pq11	training	7.06	7.048	−0.012
pq18	test	7.013	7.178	0.165
pq7	training	6.983	6.982	−0.001
pq16	test	6.827	6.764	−0.063
pq20	training	6.757	6.798	0.041
pq12	training	6.68	6.666	−0.014
pq21	training	6.529	6.535	0.007
pq34	test	6.455	6.473	0.019
pq41	training	5.95	5.922	−0.028
pq22	test	5.863	7.289	1.427
pq19	training	5.755	5.785	0.03
pq14	training	5.604	5.554	−0.05
pq16	training	6.827	6.764	−0.063
pq30	test	8.222	8.255	0.033
pq31	training	7.229	7.254	0.024
pq34	training	6.455	6.473	0.019
pq35	test	5.339	5.391	0.052

Table 4: Finger print (atom triplets), Kernel based partial least square regression data of training and test sets.

KPLS FACTORS	SD	R^2
1	0.3125	0.8913
2	0.1733	0.9676
3	0.1089	0.9876
4	0.06748	0.9954
5	0.04461	0.9980
Training set		
KPLS FACTORS	**RMSE**	Q^2
1	0.5063	0.7065
2	0.4493	0.7689
3	0.4202	0.7979
4	0.4247	0.7935
5	0.4317	0.7866

Test set

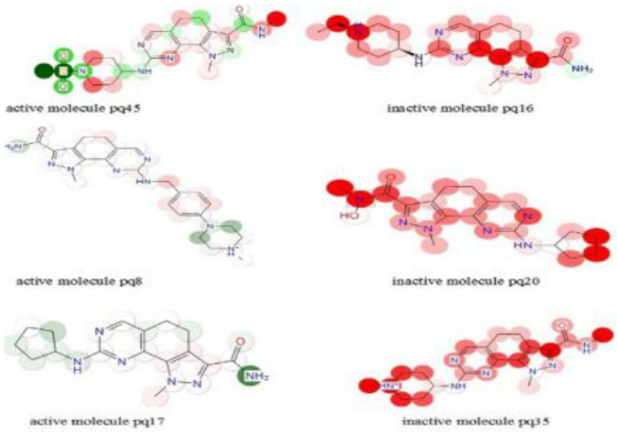


Figure 5: Colour map of fingerprint-atom triplets. Active molecules are shown in predominant green colour and inactive molecules are sown in predominant red colour.

Statistical analysis of 2D QSAR model generated for the finger print "dendritic"

The fingerprint fp dendritic gave a statistically significant 2D QSAR model with good regression coefficient values and cross-validation coefficient values as represented in Table 8. The model also showed good predictive accuracy both in the test as well as in the training set molecules as demonstrated in Table 7. The evaluation of the atomic contribution to the model was considered using three molecules from each active and inactive sets. The positively contributing factors for eliciting the activity were colored yellow and green coloured for those which are detrimentall to the model as represented in Figure 13. The 2D QSAR model generated is shown in Figure 14. The activity reported vs predicted is repre-

sented in the Figures 16 for test set and in Figure 15 for the training set.

Statistical analysis of 2D QSAR model generated for the finger print "linear"

We have arrived at a statistically significant 2D QSAR model with a minimum difference in the regression coefficient and cross-validation coefficient values as depicted in the Table 10. They displayed an optimim predictive values both in the test and training sets as represented in Table 9. The evaluation of the atomic contribution to the model was considered using three molecules from each active and inactive sets. The positively contributing factors for eliciting the activity were colored red and yellow coloured for those which are detrimental to the model as represented in Figure 17. The 2D QSAR model generated is shown in Figure 18. The activity reported vs predicted is represented in the Figure 20 for the test set and Figure 19 for the training set.

Statistical analysis of 2D QSAR model generated for the finger print "radial"

The fingerprint radial could not produce a statisticallly reliable 2D QSAR model, hence the statistics were not presented.

Consolidated 2D QSAR model generated from the 5 fingerprints

From our studies, 2D QSAR models derived from 4 fingerprint models, generated a concise and an efficient SAR model. This model directed for 3 important modifications that could be incorporated so as to enhance the activity and also slectivity towrads the enzyme, CDK2/

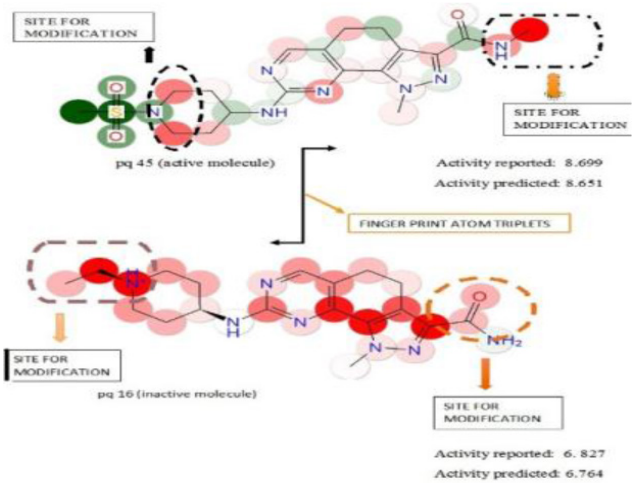


Figure 6: 2D QSAR model generated for the fingerprint atom triplets. Proposed substitution sites for the possible enhancement of inhibitory activity is presented.

Table 5: Activity prediction tables for the fingerprint atom pairs.

Structure	Model Set	pact	Pred(pact)	Error
pq42	training	9	9.045	0.045
pq44	test	9	9.321	0.321
pq17	training	8.699	8.481	−0.218
pq36	training	8.699	8.691	−0.008
pq45	training	8.699	8.575	−0.124
pq38	test	8.523	8.571	0.048
pq15	training	8.398	8.498	0.1
pq46	test	8.398	9.021	0.623
py27	training	8.301	8.346	0.045
pq29	training	8.301	8.358	0.057
pq47	training	8.301	8.412	0.111
pq26	test	8.222	7.607	−0.615
pq32	training	8.155	8.102	−0.053
pq28	training	8.097	8.078	−0.019
pq40	training	7.921	7.936	0.015
pq5	training	7.886	7.898	0.012
pq8	test	7.77	6.929	−0.84
pq13	training	7.721	7.779	0.058
pq24	training	7.699	7.663	−0.036
pq9	training	7.678	7.699	0.021
pq10	training	7.678	7.694	0.016
pq23	test	7.638	7.724	0.086
pq39	test	7.602	7.638	0.036
pq37	training	7.495	7.509	0.015
pq6	training	7.42	7.367	−0.053
pq33	test	7.409	7.301	−0.108
pq1	training	7.292	7.375	0.083
pq25	training	7.284	7.319	0.035
pq3	training	7.237	7.124	−0.113
pq43	training	7.237	7.15	−0.086
pq2	training	7.18	7.143	−0.038
pq4	training	7.092	7.111	0.019
pq11	test	7.06	7.55	0.49
pq18	training	7.013	7.139	0.126
pq7	test	6.983	6.461	−0.522
pq16	training	6.827	6.808	−0.019
pq20	training	6.757	6.78	0.023
pq12	training	6.68	6.576	−0.104
pq21	test	6.529	7.191	0.662
pq34	test	6.455	6.456	0.001
pq41	training	5.95	5.913	−0.037
pq22	training	5.863	5.879	0.016
pq19	training	5.755	5.78	0.025
pq14	training	5.604	5.688	0.085
pq35	training	5.339	5.328	−0.01
pq16	training	6.827	6.808	−0.019
pq30	training	8.222	8.258	0.036
pq30	test	8.222	8.258	0.036
pq31	training	7.229	7.223	−0.006
pq31	test	7.229	7.223	−0.006
pq34	training	6.455	6.456	0.001
pq35	test	5.339	5.328	−0.01

Table 6: Finger print (atom pairs), Kernel based partial least square regression data of training and test sets.		
KPLS FACTORS	**SD**	R^2
1	0.3717	0.8493
2	0.2180	0.9497
3	0.1437	0.9788
4	0.1032	0.9894
5	0.07598	0.9944
Training set		
KPLS FACTORS	**RMSE**	Q^2
1	0.4095	0.7879
2	0.3921	0.8154
3	0.3896	0.8178
4	0.4057	0.8024
5	0.4124	0.7958

Test set

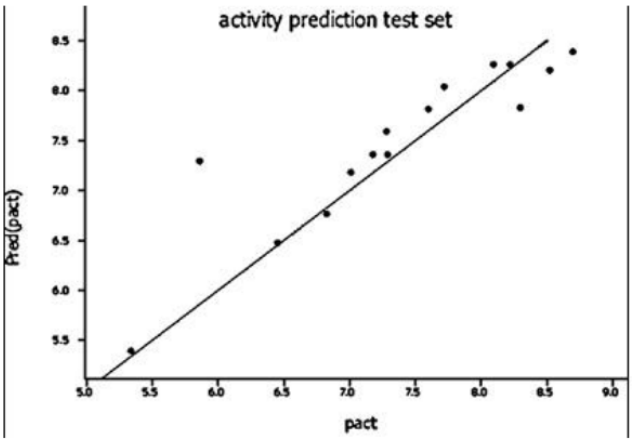


Figure 7: Actual versus predicted activity in test set.

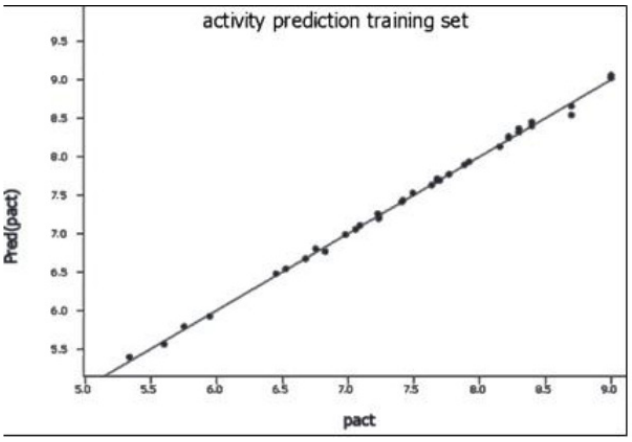


Figure 8: Actual versus predicted activity in training set.

Table 7: Activity prediction tables for the fingerprint dendritic.				
Structure	**Model Set**	**pact**	**Pred(pact)**	**Error**
pq42	test	9	8.543	−0.457
pq44	training	9	9.448	0.448
pq17	training	8.699	8.429	−0.27
pq36	training	8.699	8.637	−0.062
pq45	training	8.699	8.324	−0.375
pq38	training	8.523	8.665	0.142
pq15	training	8.398	8.505	0.107
pq46	test	8.398	8.868	0.47
py27	training	8.301	8.282	−0.019
pq29	training	8.301	8.346	0.045
pq47	test	8.301	7.807	−0.494
pq26	test	8.222	8.016	−0.206
pq32	training	8.155	8.098	−0.057
pq28	test	8.097	8.113	0.016
pq40	training	7.921	7.564	−0.357
pq5	test	7.886	8.12	0.233
pq8	training	7.77	7.798	0.028
pq13	training	7.721	7.763	0.042
pq24	test	7.699	8.044	0.345
pq9	training	7.678	7.699	0.021
pq10	training	7.678	7.704	0.026
pq23	training	7.638	7.643	0.005
pq39	training	7.602	7.556	−0.046
pq37	training	7.495	7.43	−0.065
pq6	test	7.42	7.879	0.459
pq33	training	7.409	7.345	−0.064
pq1	training	7.292	7.37	0.077
pq25	training	7.284	7.345	0.061
pq3	test	7.237	7.598	0.362
pq43	test	7.237	7.351	0.114
pq2	training	7.18	7.179	−0.002
pq4	training	7.092	7.104	0.012
pq11	training	7.06	7.141	0.081
pq18	test	7.013	7.23	0.217
pq7	training	6.983	7.01	0.027
pq16	test	6.827	6.831	0.004
pq20	test	6.757	6.992	0.235
pq12	training	6.68	6.65	−0.03
pq21	training	6.529	6.586	0.057
pq41	training	5.95	6.315	0.365
pq22	training	5.863	5.84	−0.023
pq19	training	5.755	5.786	0.031
pq14	training	5.604	5.617	0.014
pq35	test	5.339	5.282	−0.057
pq16	training	6.827	6.831	0.004
pq30	training	8.222	8.155	−0.067
pq31	training	7.229	7.211	−0.018
pq34	training	6.455	6.457	0.002

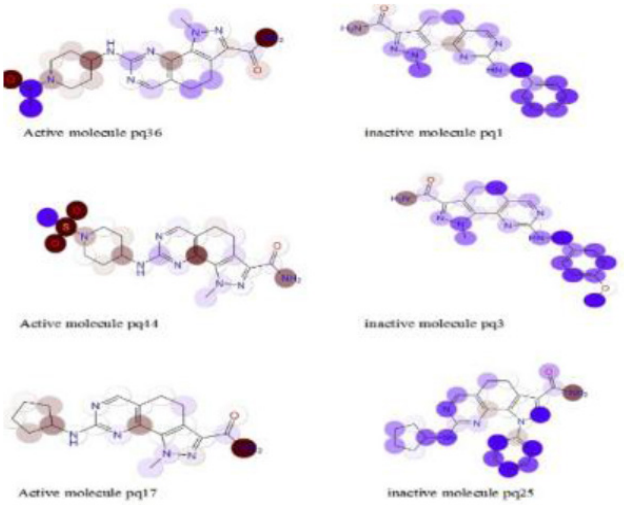


Figure 9: Colour map for the fingerprint-atom-pairs. Active molecules are shown in predominant brown colour and inactive molecules are sown in predominant blue colour.

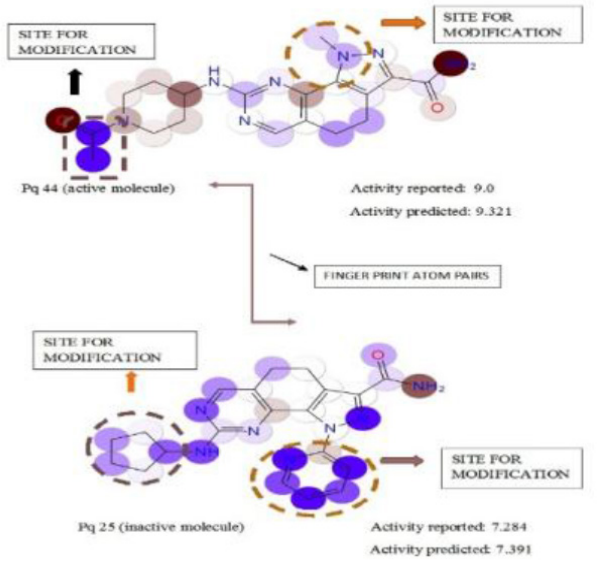


Figure 10: 2D QSAR model generated for the fingerprint atom pairs. Proposed substitution sites for the possible enhancement of inhibitory activity is presented.

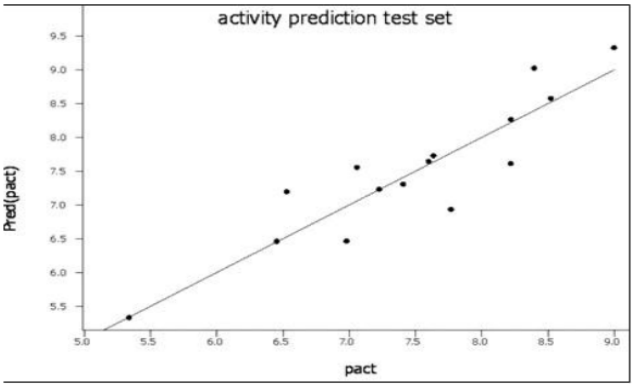


Figure 11: Actual versus predicted activity in test set.

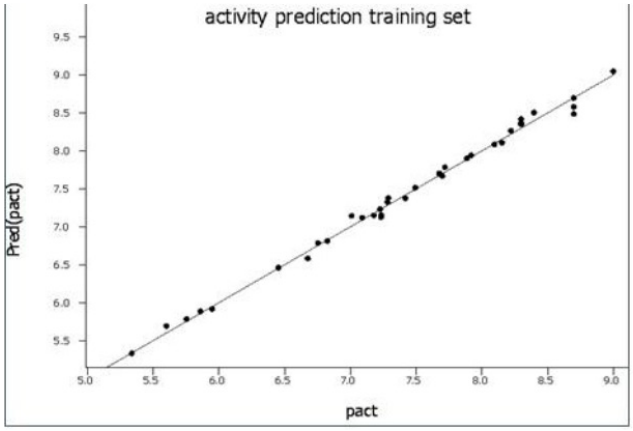


Figure 12: Actual versus predicted activity in training set.

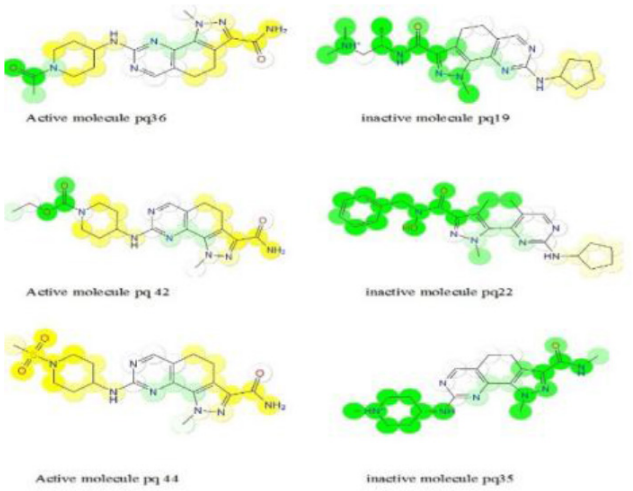


Figure 13: Colour maps for the fingerprint-dendritic. Active molecules are shown in predominant yellow colour and inactive molecules are sown in predominant green.

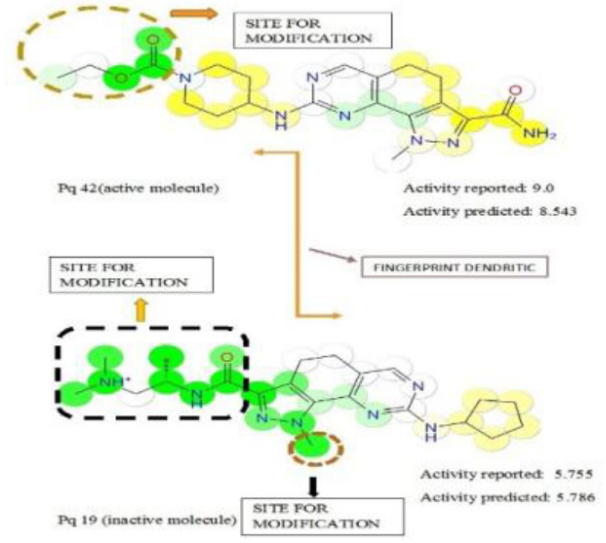


Figure 14: 2D QSAR model generated for the fingerprint dendritic. Proposed substitution sites for the possible enhancement of inhibitory activity is presented.

Table 8: Finger print (dendritic), Kernel based partial least square regression data of training and test sets.

KPLS FACTORS	SD	R^2
1	0.3828	0.8431
2	0.2007	0.9581
3	0.1302	0.9828
4	0.06019	0.9965
5	0.03205	0.9990
Training set		
KPLS FACTORS	**RMSE**	Q^2
1	0.4543	0.7422
2	0.4379	0.7605
3	0.4543	0.7422
4	0.4602	0.7355
5	0.4783	0.7143

Test set

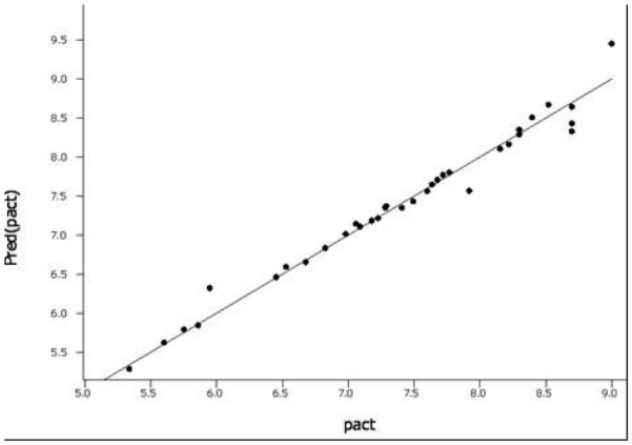


Figure 15: Actual versus predicted activity in training set.

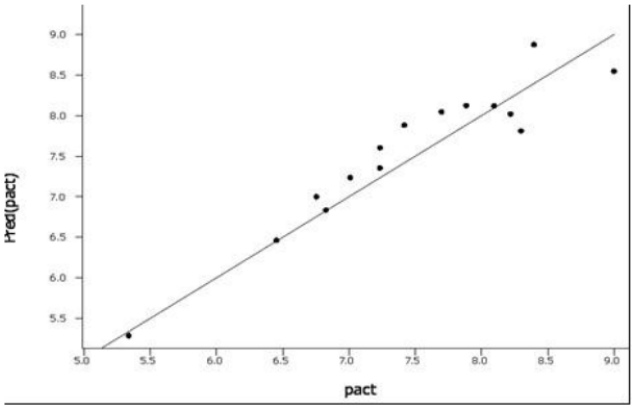


Figure 16: Actual versus predicted activity in test set.

Table 9: Activity prediction tables for the fingerprint Linear.

Structure	Model Set	pact	Pred (pact)	Error
pq42	training	9	8.7	-0.3
pq44	test	9	9.694	0.694
pq17	training	8.699	8.527	-0.172
pq36	training	8.699	8.699	0
pq45	training	8.699	8.747	0.049
pq38	test	8.523	8.567	0.044
pq15	training	8.398	8.526	0.128
pq46	training	8.398	8.803	0.405
py27	training	8.301	8.3	-0.001
pq29	test	8.301	8.264	-0.037
pq47	training	8.301	7.924	-0.377
pq26	training	8.222	8.332	0.11
pq32	test	8.155	9.134	0.979
pq28	training	8.097	8.084	-0.013
pq40	training	7.921	7.437	-0.484
pq5	training	7.886	7.888	0.002
pq8	training	7.77	7.727	-0.043
pq13	training	7.721	7.747	0.026
pq24	test	7.699	7.451	-0.248
pq9	test	7.678	7.71	0.032
pq10	training	7.678	7.666	-0.012
pq23	training	7.638	7.623	-0.016
pq39	training	7.602	7.614	0.012
pq37	test	7.495	7.558	0.063
pq6	training	7.42	7.427	0.007
pq33	training	7.409	7.45	0.041
pq1	test	7.292	7.271	-0.021
pq25	training	7.284	7.234	-0.05
pq3	training	7.237	7.264	0.028
pq43	training	7.237	7.47	0.233
pq2	training	7.18	7.159	-0.022
pq4	training	7.092	7.146	0.054
pq11	test	7.06	7.349	0.289
pq18	training	7.013	6.998	-0.015
pq7	training	6.983	6.888	-0.095
pq16	training	6.827	6.868	0.041
pq20	training	6.757	6.803	0.046
pq12	training	6.68	6.707	0.027
pq21	training	6.529	6.52	-0.009
pq34	test	6.455	6.346	-0.108
pq41	training	5.95	6.167	0.217
pq22	training	5.863	5.924	0.061
pq19	training	5.755	5.826	0.071
pq14	training	5.604	5.571	-0.033
pq35	training	5.339	5.243	-0.095
pq16	test	6.827	6.868	0.041
pq30	training	8.222	8.364	0.143
pq30	test	8.222	8.364	0.143
pq31	training	7.229	7.264	0.035
pq31	test	7.229	7.264	0.035
pq34	test	6.455	6.346	-0.108
pq35	test	5.339	5.243	-0.095

Table 10: Finger print (linear), Kernel based partial least square regression data of training and test sets.		
KPLS FACTORS	**SD**	R^2
1	0.4847	0.7492
2	0.3276	0.8888
3	0.213	0.9544
4	0.1778	0.9692
5	0.1598	0.9760
Training set		
KPLS FACTORS	**RMSE**	Q^2
1	0.3418	0.8256
2	0.3693	0.8279
3	0.3117	0.8775
4	0.2985	0.8876
5	0.3000	0.8864

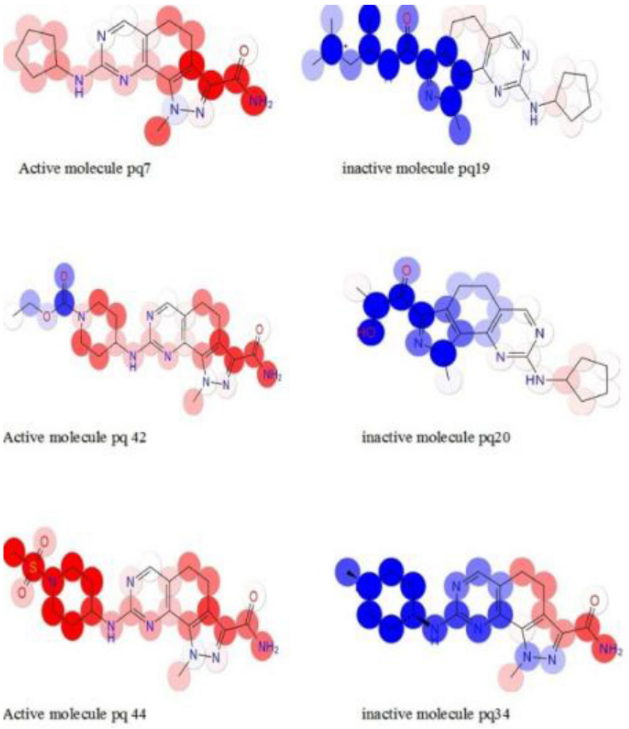


Figure 17: Colour maps for the fingerprint-linear. Active molecules are shown in predominant red colour and inactive molecules are shown in predominant blue colour.

CYCLIN A. The important modifications proposed are as follows:

The retention of the carboxamide group on the pyrazole ring is quite crucial to predict the activity. Any substitutions in the carboxamide nitrogen would tend to decrease the activity in all the 5 fingerprint models.

Substitution of piperidine with an amino linkage to the quinazoline moeity is essential for the inhibitory activity.

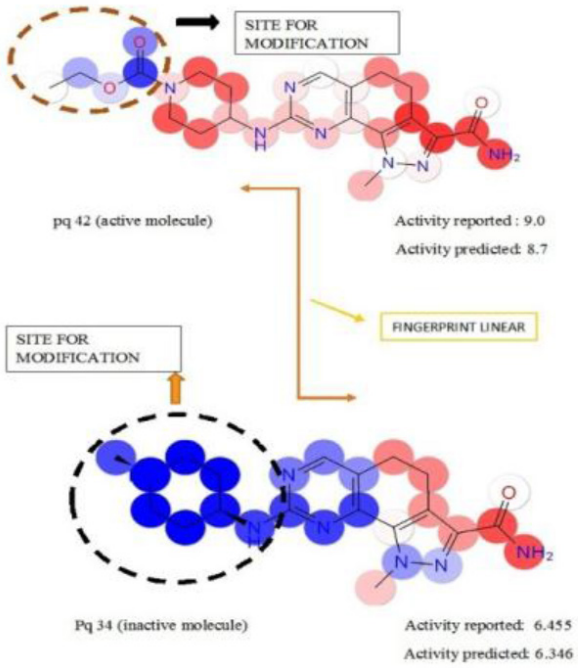


Figure 18: 2D QSAR model generated for the fingerprint linear. Proposed substitution sites for the possible enhancement of inhibitory activity is presented.

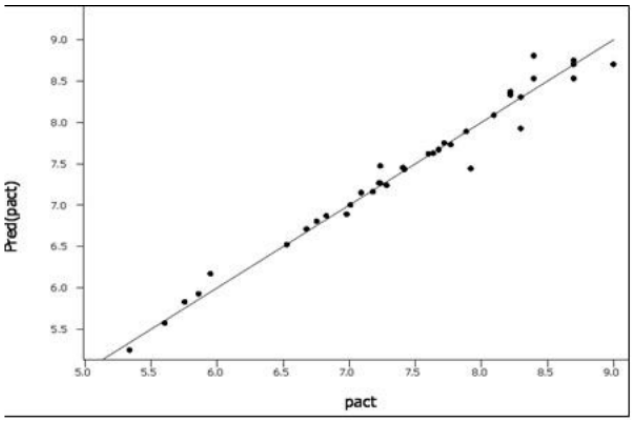


Figure 19: Actual versus predicted activity in training set.

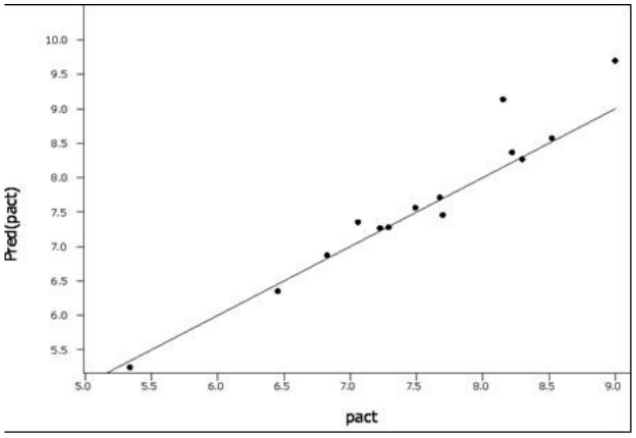


Figure 20: Actual versus predicted activity in test set.

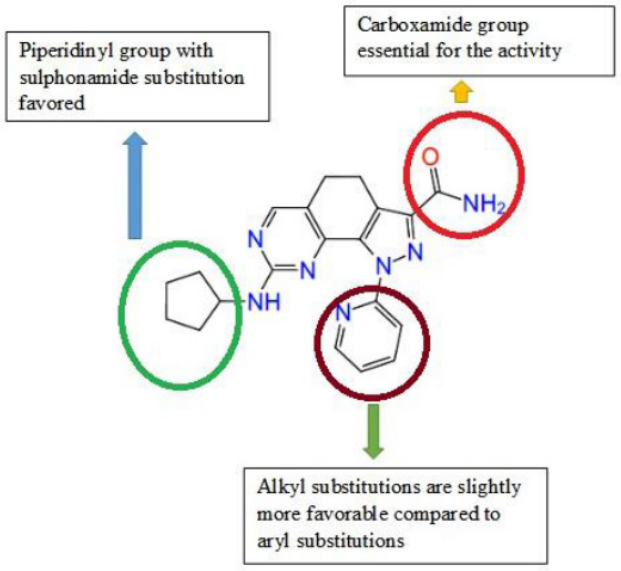


Figure 21: Consolidated SAR of all the 4 fingerprint models.

However, the substitution on piperdinyl nitrogen with sulphonamido group could be responsible for the enhanced activity from our fingerprint models. Whereas, the substitution of a carboxamide for sulphonamide could have dentrimental effect on the activity.

Further, alkyl substitutions are more favored compared to the aryl substitutions on the nitrogen of the pyrazole moiety with a slight increase in the activity. The consolidated SAR model is presented in Figure 21.

CONCLUSION

A 2D QSAR based study for a series of novel pyrazolo quinazolines were performed using Kernel-based regression using five prominent binary fingerprinting methods. The contributions from each one of the fingerprints models towards the activity was influenced by multiple factors. Thus, in our present study, the fingerprint linear generated a statistically significant 2-dimensional structure activity relationship model with the most appropriate statistical values. The features responsible for the success of linear fingerprint are mainly attributed to their two-dimensional structure and substitution pattern. The structural features of compounds showed the fusion between bicyclic quinazoline moieties and a monocyclic pyrazole moiety with various substitutions on both quinazoline as well as pyrazole ring. However, the limitations of this study revealed that, the fingerprint radial could not reach reliable statistical values. This limitation could be attributed to the atom typing and molecular descriptors selected. Further, suitable fingerprints should be incorporated based on their molecular structure, linear connectivity, nature and branching of the ring systems.

ACKNOWLEDGEMENT

The authors would like to acknowledge the facilities provided by Manipal College of Pharmaceutical Sciences and Manipal Academy for Higher Education in executing the present research work. Schrödinger Inc. USA for the software support provided to carry out molecular docking and 2D QSAR studies.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ABBREVIATIONS

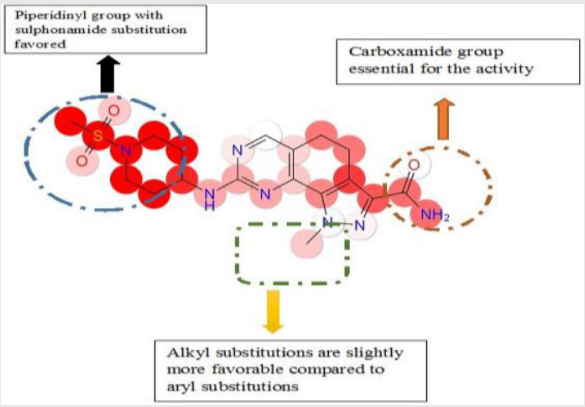
CDK: Cyclin dependent Kinase; **2D:** Two dimensional; **3D:** Three Dimensional; **QSAR:** Quantitative structure-activity relationship.

REFERENCES

1. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, *et al*. How to improve R and D productivity: The pharmaceutical industry's grand challenge. Nature Reviews Drug Discovery. 2010;9(3):203.
2. Bajorath J. Rational drug discovery revisited: Interfacing experimental programs with bio-and chemo-informatics. Drug Discovery Today. 2001;6(19):989-95.
3. Mestres J. Virtual screening: A real screening complement to highthroughput screening. Biochem Soc Trans. 2002;30(4):797-9.
4. Smith A. Screening for drug discovery: The leading question. Nature. 2002;418(6896):453.
5. Bajorath J. Integration of virtual and high-throughput screening. Nature Reviews Drug Discovery. 2002;1(11):882.
6. Walters WP, Stahl MT, Murcko MA. Virtual screening-an overview. Drug Discovery Today. 1998;3(4):160-78.
7. Dixon SL, Merz KM. One-dimensional molecular representations and similarity calculations: Methodology and validation. Journal of Medicinal Chemistry. 2001;44(23):3795-809.
8. Miller MD, Kearsley SK, Underwood DJ, Sheridan RP. FLOG: A system to select 'quasi-flexible'ligands complementary to a receptor of known three-dimensional structure. Journal of Computer-Aided Molecular Design. 1994;8(2):153-74.
9. Clark DE, Jones G, Willett P, Kenny PW, Glen RC. Pharmacophoric pattern matching in files of three-dimensional chemical structures: Comparison of conformational-searching algorithms for flexible searching. Journal of Chemical Information and Computer Sciences. 1994;34(1):197-206.
10. Putta S, Lemmen C, Beroza P, Greene J. A novel shape-feature based approach to virtual library screening. Journal of Chemical Information and Computer Sciences. 2002;42(5):1230-40.
11. Flower DR. On the properties of bit string-based measures of chemical similarity. Journal of Chemical Information and Computer Sciences. 1998;38(3):379-86.
12. Cramer RD, Poss MA, Hermsmeier MA, Caulfield TJ, Kowala MC, Valentine MT. Prospective identification of biologically active structures by topomer shape similarity searching. Journal of Medicinal Chemistry. 1999;42(19):3919-33.
13. Zhang Q, Muegge I. Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: ranking, voting and consensus scoring. Journal of Medicinal Chemistry. 2006;49(5):1536-48.
14. Schrödinger Release: Canvas, Schrödinger, LLC, New York, NY. 2017;4.

15. Sastry M, Lowrie JF, Dixon SL, Sherman W. Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments. Journal of Chemical Information and Modeling. 2010;50(5):771-84.
16. Sielecki TM, Boylan JF, Benfield PA, Trainor GL. Cyclin-dependent kinase inhibitors: Useful targets in cell cycle regulation. Journal of Medicinal Chemistry. 2000;43(1):1-8.
17. Morgan DO. Cyclin-dependent kinases: engines, clocks and microprocessors. Annual Review of Cell and Developmental Biology. 1997;13(1):261-91.
18. Malumbres M, Barbacid M. Mammalian cyclin-dependent kinases. Trends in Biochemical Sciences. 2005;30(11):630-41.
19. Wadler S. Perspectives for cancer therapies with cdk2 inhibitors. Drug Resistance Updates. 2001;4(6):347-67.
20. Lowe EDI, Tews KY, Cheng NR, Brown S, Gul MEM, Nobel SJ, *et al*. Specificity determinants of recruitment peptides bound to phospho-CDK2/cyclin A. Johnson Biochemistry. 2002;41(52):15625-34.
21. Sherr CJ. The Pezcoller lecture: Cancer cell cycles revisited. Cancer Research. 2000;60(14):3689-95.
22. Cotesta S, Giordanetto F, Trosset JY, Crivori P, Kroemer RT, Stouten PF, *et al*. Virtual screening to enrich a compound collection with CDK2 inhibitors using docking, scoring and composite scoring models. Proteins: Structure, Function and Bioinformatics. 2005;60(4):629-43.
23. Traquandi G, Ciomei M, Ballinari D, Casale E, Colombo N, Croci V, *et al*. Identification of potent pyrazolo [4,3-h] quinazoline-3-carboxamides as multi-cyclin-dependent kinase inhibitors. Journal of Medicinal Chemistry. 2010;53(5):2171-87.
24. Li YP, Weng X, Ning FX, Ou JB, Hou JQ, Luo HB, *et al*. 3D-QSAR studies of azaoxoisoaporphine, oxoaporphine and oxoisoaporphine derivatives as anti-AChE and anti-AD agents by the CoMFA method. Journal of Molecular Graphics and Modelling. 2013;41:61-7.
25. Huey R, Morris GM, Olson AJ, Goodsell DS. A semiempirical free energy force field with charge-based desolvation. Journal of Computational Chemistry. 2007;28(6):1145-52.
26. Schrödinger Release Glide, Schrödinger, LLC, New York, NY. 2017;4.
27. Muddukrishna BS, Pai V, Lobo R, Pai A. Application of two-dimensional binary fingerprinting methods for the design of selective Tankyrase I inhibitors. Molecular Diversity. 2018;22(2):359-81.

PICTORIAL ABSTRACT



SUMMARY

- Much of the work on the computational studies of cell cycle related topic has not been addressed for many years even though independently substantial work of progressive development has taken place both in computational science as well as in oncology research. In our present study, we are attempting to relate these two subjects of interest.
- Our study aims at learning the application of five important binary fingerprinting techniques based on the quantum mechanics involving 2D QSAR studies for designing novel pyrazolo quinazolines as selective inhibitors of CDK2/CyclinA.
- A dataset of 47 analogues of pyrazolo quinazolines were selected with their inhibitory activity on CDK2/Cyclin A. The derivatives were divided into the training and test sets. The Kernel based partial regression was run using five important binary fingerprints and statistical significance of each fingerprint was analysed.
- Out of the five fingerprints selected, the fingerprint linear arrived at the optimized 2D QSAR model through the kernel-based regression analysis. The final developed model expressed the importance of the presence of carboxamide groups on the pyrazole ring that could positively contributes to the inhibitory activity.
- The developed model could be of use to design better analogues with better selectivity and specificity as inhibitors of CDK2/Cyclin A that would provide a clear insight amongst the researchers for the development of novel and potent clinically useful anticancer agents.

Cite this article: Pai A, Jayashree BS. Fingerprint Based Two-dimensional QSAR Studies on Pyrazolo Quinazolines as Cdk2 Inhibitors: A Rational Approach for the Design of Novel Anticancer Agents. Indian J of Pharmaceutical Education and Research. 2019;53(3 Suppl 2):s299-s312.