Original Article

Bulletin of Pure and Applied Sciences.
Vol.37 C (Chemistry), No.1, 2018: P.53-63
Print version ISSN 0970 4620
Online version ISSN 2320 320X
DOI 10.5958/2320-320X.2018.00008.0

2D and 3D QSAR model generation of CDK4 inhibitors

Divya V[1], Pushpa V.L.[1]*, Sarithamol S, Manoj K.B[1]

Author Affiliations
[1]Department of Chemistry, Sree Narayana College, Kollam, Kerala, India.

Corresponding Author
***Pushpa V.L,** Department of Chemistry, Sree Narayana College, Kollam, Kerala, India.

E-mail: drpushpavl2017@gmail.com

Received on 10th January 2018
Accepted on 24th January 2018

Abstract

2D and 3D QSAR studies were applied on a set of CDK4 inhibitors which were assayed by same method to model and understand inhibitory activities. Among the various 2D QSAR models generated, ten models created by Kernel-Based Partial Least Square (kpls) regression method had high scores with an average value 0.8561. Top ranked model kpls_ linear_15 has a score value of 0.8804, R^2 0.8947 and Q^2 0.8909. This model considers linear fragments and ring closures. Bioactive conformation of the ligand molecules were identified by docking analysis of highly active molecule **2 (**pIC50 =8.796) in the Extra Precision mode with CDK4 homology model and the binding affinity was obtained as XP GScore of –8.072 which is the total contribution of coulomic interactions, binding interactions and van der Waals interations. All optimized structures were aligned and an atom based 3D QSAR model was established (R^2=0.9651, R^2_{cv}=0.8191 and Q^2=0.7368). This model was assigned large contribution of hydrophobic/ nonpolar features, electron withdrawing features and hydrogen bond donor features to the biological activity. The influences of these structural features were validated by analyzing the docking results. These outcomes will direct advance structural alteration and prediction of new CDK4 inhibitors.

Keywords: QSAR, Homology model, CDK4, Docking, Drug design

1. INTRODUCTION

Cell cycles are regulated by different types of Cyclins and Cyclin Dependent Kinases. Of these, the *holoenzyme* [1] formed by the complexation of G1/S-Specific Cyclin D1 and Cyclin Dependent Kinase 4 and 6 (CDK4/6) are of particular interest because the mutation in its genome level [2–5] or over expression of either the proteins (Cyclin D1 and CDK4) or the holoenzyme is very significant in the proliferation of different types of cancers.

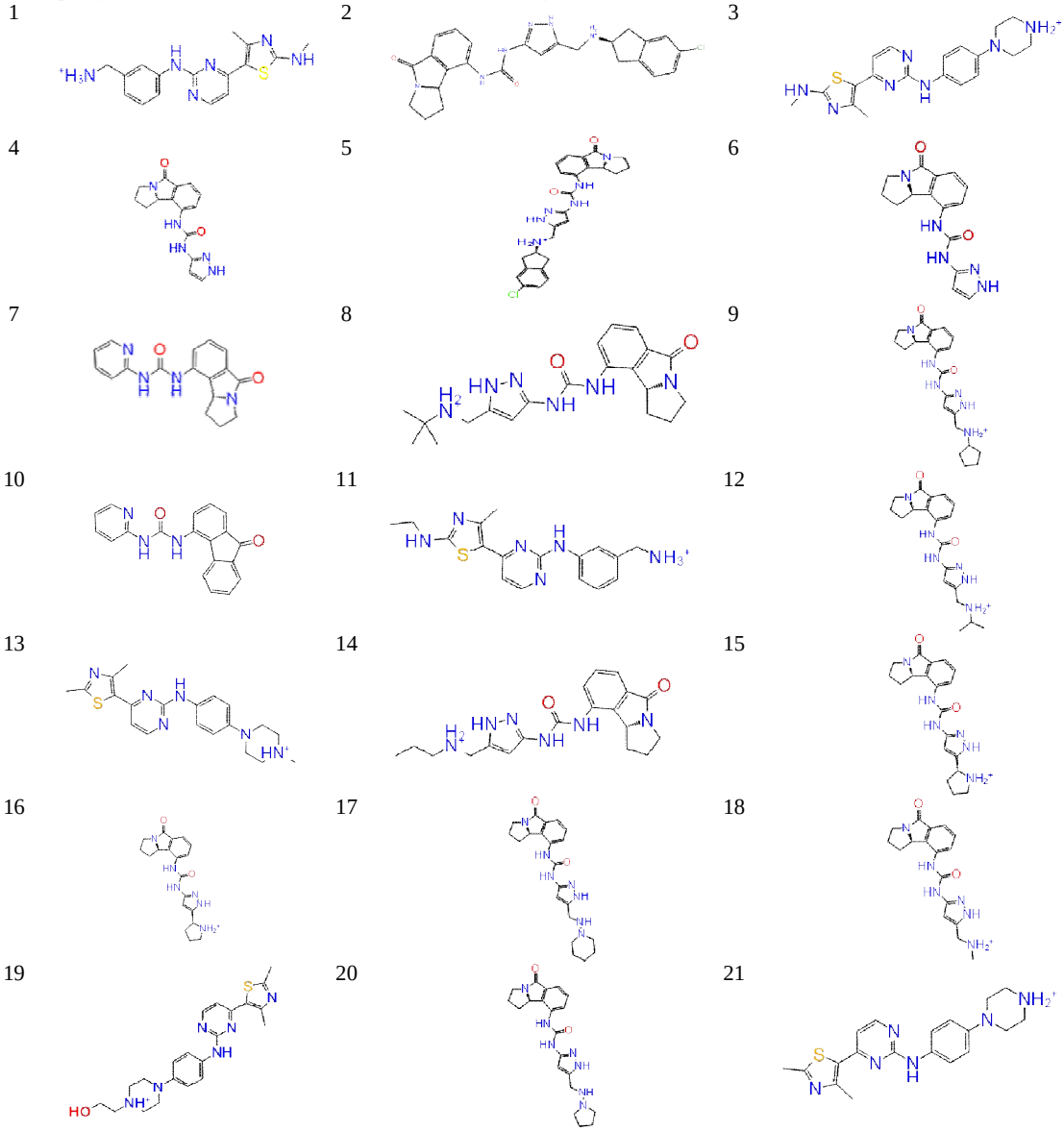
ATP competitive CDK4 inhibitors have shown to cause apoptosis at the G1/S level and thus promote cancer cell death. In the present study, we performed 2D and 3D QSAR studies on 43 ATP competitive CDK4 inhibitors founded on their structures and biological activities to get understanding into the vital structural aspects that influence selective inhibition. 2D QSAR models were generated following a best-practices methodology like generation of descriptors, validate, generate a large

number of QSAR models and apply models for prediction [6]. Atom based 3D QSAR study was employed for finding out the significance of electrostatic features, steric factors, non-bonding interactions etc of the same data set. For finding out the bioactive conformation of the ligand molecule, several structure-based designing strategies including homology modeling of CDK4 and docking analysis of highest biologically active molecule (**2**) were carried out. Assuming the docked form of the molecule **2** as the active conformation, all the prepared molecules were aligned on it and went for Atom based 3D QSAR modeling.

2. EXPERIMENTAL

2.1. Data Set

3D structures of 43molecules having inhibitory activity towards CDK4 in terms of IC50 [7–10] were drawn and prepared by using the ligprep module of Schrodinger software suits. The molecules were randomly divided into training sets (33 molecules 75%) and test sets (11 molecules, 25%) for developing the robust 2D and 3D QSAR models (Figure 1).



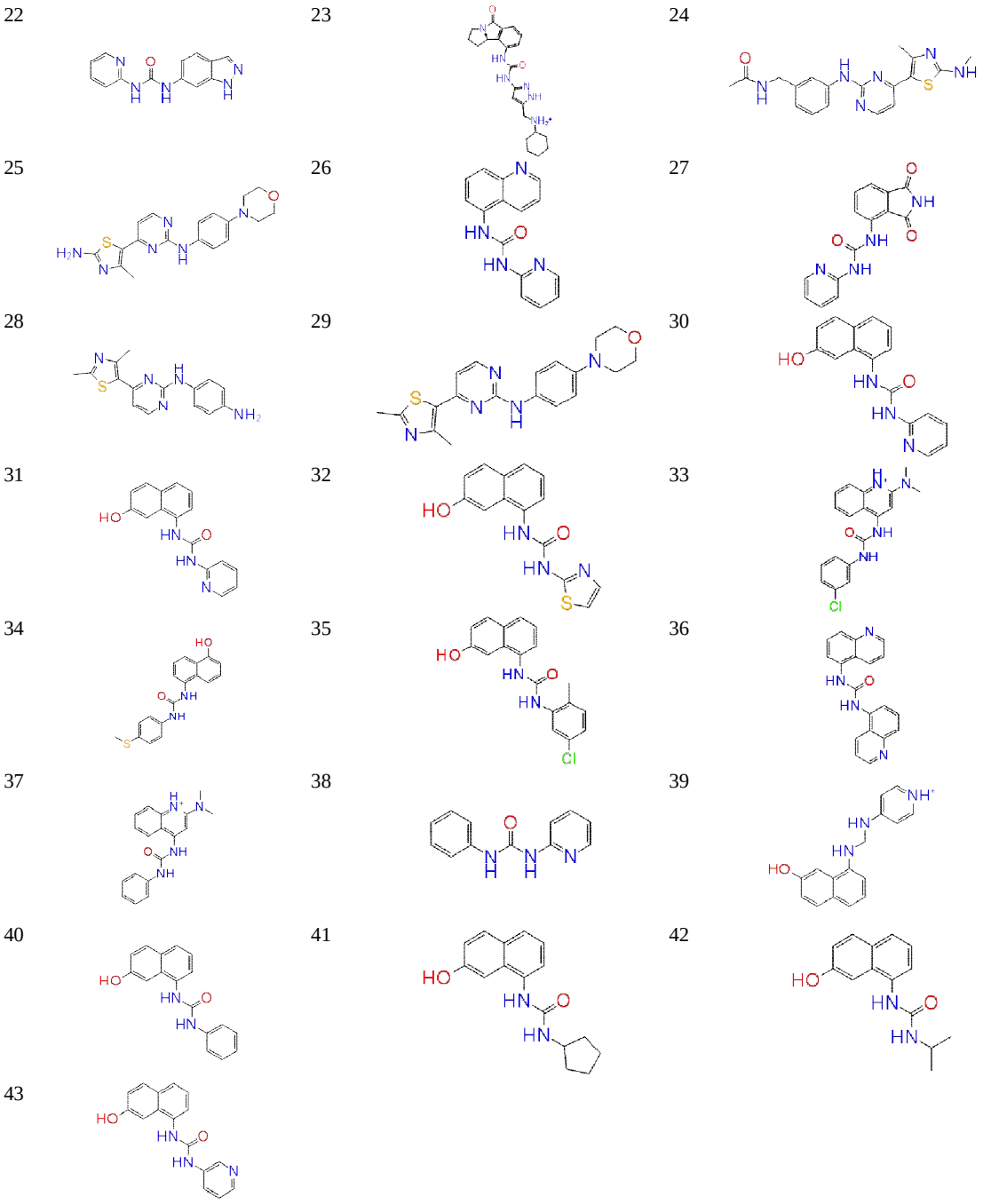


Fig. 1: Structure of molecules with CDK4 inhibitory activity

2.2. 2D QSAR study

AutoQSAR is a fully automated QSAR model generator from Schrodinger Software suit. By considering all the 1D, 2D and 3D structural data as input and computed descriptors and fingerprints and best suitable QSAR modelsare also generated using CANVAS module of Schrodinger Software suit. It adopts different approaches like Kernal- based partial least squares, Naïve Bayes and ensemble-based recursive partitioning for model generations. Different QSAR models are ranked basedon their performance. Diverse machine learning approaches were executed for training individualistically using these descriptors and fingerprints. Those descriptors were eliminated which

were having the same value for all compounds since they do not provide any significant data. Various descriptors with substantial collinearities were also eliminated since they do not provide any useful contributions to the development of QSAR. The maximum allowed correlations between any pair of independent variables were adjusted to 0.80.Reduction was achieved by performing hierarchical, agglomerative clustering on the absolute Pearson correlation matrix of the descriptors, which is a proper similarity matrix. 50 different random splits of learning set into 75% training set and 25% test set and went for QSAR model generations. A total of 400 models are generated.A quality score is assigned to each model based on its performance on both its training set and test sets and are sorted by decreasing score and the top ten retained.

2.3. Modeling and 3D QSAR studies

Conformations of ligands in the bioactive form, its orientation in the lattice and alignments are the main contributing factors for the 3DQSAR analysis [11]. Since no inhibitor bound CDK4 protein crystal structure was reported, a homology model of CDK4 was constructed using the crystal structure of CDK2 (1GIH; 2.8Å resolution) complexed with CDK4 inhibitor, **10** [12]. Knowledge based model building method was adopted by which insertions were constructed and gaps were closed with segments from known structures and returned to a single model of the structure. The inhibitor molecule bound with the template protein structure was also selected while building the model.

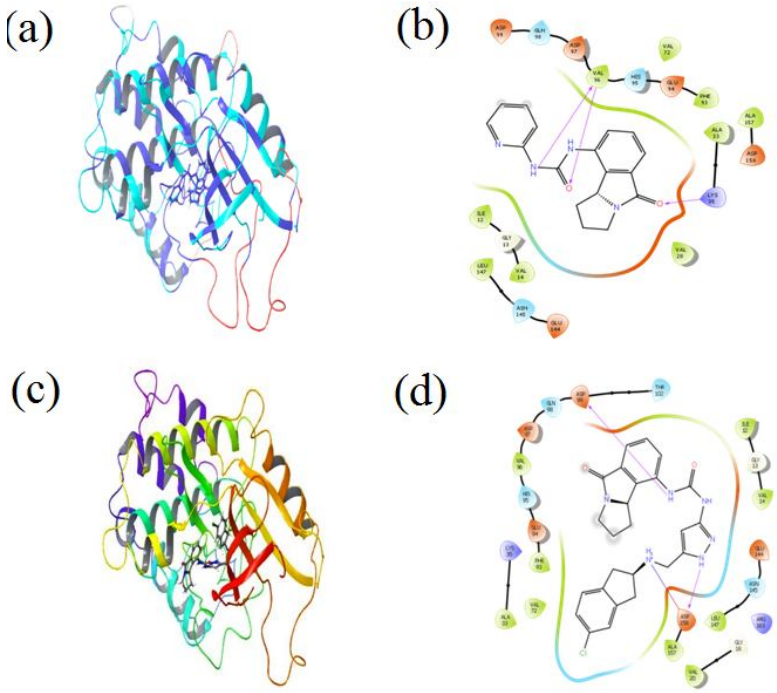


Fig. 2: (a) Homology model of CDK4 with natural ligand, 10 (b) Ligand interaction diagram of modeled CDK4 and natural ligand,10 (c) Homology model of CDK4 with compound 2 and (d) ligand interaction diagram of modelled CDK4 and compound 2.

Molecular alignment is crucial for the construction of reliable 3D QSAR models. The sensitivity of alignment based 3D-QSAR was determined by the superimposition of bioactive conformations of molecules on each other [13]. In this study, docking based alignment was adopted since this method is comparatively accurate without much subjective influence. Docked conformation of **2** was selected as a template for alignment since it had a high biological activity (Figure 2). Docking was performed in extra precision mode. Receptor Grid was generated around the centroid of workspace ligand by selecting the molecule. The scaling factor corresponding to van der Waals radii of non-polar receptor atoms was adjusted to 1.0 and the partial atomic charge cut off was specified below 0.25. All other

prepared molecules were aligned on the bioactive conformation of the template molecule, **2.** Figure 3 represents the alignment of molecules on the bioactive conformation of **2**.

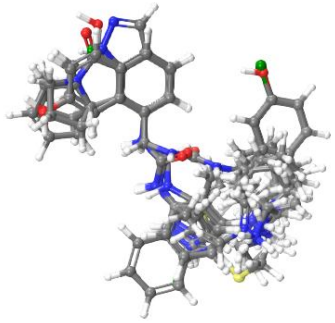


Fig. 3: The alignment of molecules on the bioactive conformation of **2**.

Both internal and external cross validation techniques were used for developing model since the consistency of the model and prediction of biological activity of those molecules which were not included in the training set were dependent on its statistical implication. Same molecules were considered for training and test set as that of 2D QSAR model. An atom based QSAR model was developed based on these molecules.

The 3D grid covering the space occupied by the ligands was specified as 1Å .maximum number of partial least square factors was chosen as 3. With an aim to get better predictions of test set molecules, we have to avoid unproductive variables. This had to be achieved by eliminating variables with l**t level**l less than 2. Since the number of training molecules was 33, leave one out (LOO) cross validation statistics was used for internal validation the QSAR model.

3. RESULTS AND DISCUSSIONS

3.1. 2D QSAR study

Based on 2D structural information and biological activity of 43 structurally prepared molecules, 497 physiochemical and topological descriptors along with a variety of Canvas fingerprints were generated and based on these independent variables, models were built. A large number of models were built and validated using different random training and test sets. The quality of the model was evaluated by analyzing the score. Top ranked ten models were listed in Table 1. From the analysis, it was found that Kernel-Based Partial Least Square (kpls) regression method using both descriptors and fingerprints generates high scored model.

The robustness of the model can be assessed by its score value. As the accuracy level approaches to 1, corresponds to perfect predictions. The score value can be calculated as

$$Score_M = accuracy_{test}\ (1.0 - |accuracy_{train} - accuracy_{test}|)$$

In the present study, all the top ranked ten models have score values greater than 0.82 and has an average value of 0.8561. Scatter plot of top scored model is presented in Figure 4.

The correlation coefficients $R^2$of the models between the observed and predicted activity of the training set have an average greater than 0.85 and that of the test sets also have an average Q^2 value 0.8675.The average RMSE and standard deviations of all these models were 0.4897 and 0.5222 respectively. And all these statistical results were largely established for high performance. Top ranked model, kpls_linear_15shows the highest score (0.8804), R^2(0.8947) and Q^2 (0.8909)and does considerably better than all other fingerprint types. This finding advocates that this model is a robust fingerprint method that has utmost chance of perfect execution with any settings.

Table 1: Statistical parameters of 2D QSAR models

Model Code	Score	S.D	R^2	RMSE	Q^2	MW(Null Hypothesis)
kpls_linear_15	0.8804	0.4693	0.8947	0.4693	0.8909	0.5264
kpls_molprint_2D_50	0.8801	0.5216	0.8762	0.3431	0.9342	0.4234
kpls_dentritic_17	0.8691	0.5231	0.8767	0.4985	0.8546	0.515
kpls_dentritic_49	0.866	0.5522	0.8632	0.4673	0.8701	0.5475
kpls_molprint_2D_15	0.8609	0.5455	0.8577	0.4728	0.8892	0.5264
kpls_linear_49	0.8532	0.5136	0.8816	0.5105	0.8439	0.5475
kpls_dentritic_20	0.852	0.4882	0.8885	0.5056	0.869	0.1848
kpls_molprint_2D_49	0.8394	0.5614	0.8586	0.5432	0.8238	0.5475
kpls_dentritic_15	0.839	0.5082	0.8765	0.5295	0.861	0.5264
kpls_molprint_2D_24	0.8206	0.5384	0.8671	0.5554	0.8371	0.2262

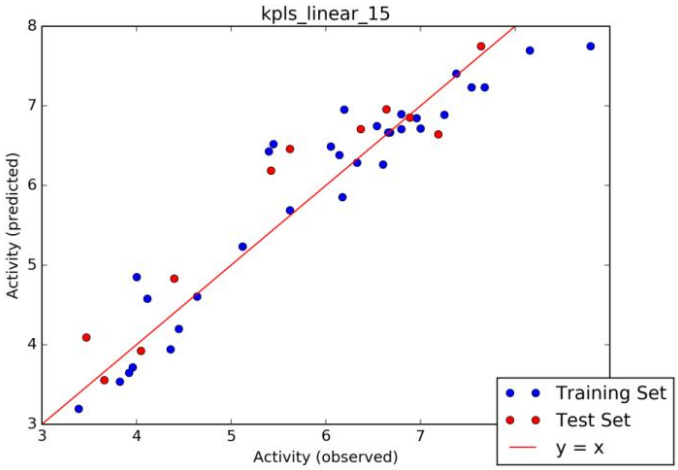


Fig. 4: kpls linear 2D QSAR model

3.2. Modelling and 3D QSAR studies

Homology model of CDK4 was constructed in the ligand bound form considering CDK2 as a template using the Prime module of Schrodinger software suitIle12, Val 20, Ala 33, Val 77, Phe 93, Glu 94, His 45, Val 96, Gln 98, Asp 99, Thr 102, Glu 143, Leu 147, Ala 157 and Asp158 are the binding region residues [14].

Both Intrinsic forces and extrinsic forces considerably stimulate the bioactive conformation of the molecule [15–16]. Consistency of 3D QSAR model was governed by the determination of bioactive conformations. In order to find out the bioactive conformation of the ligands, docking procedure was adopted. The molecule, **2** with highest biological activity in terms of pIC50 (8.796) was selected for flexible docking with CDK4 in its active site in which the natural ligand **10** was bound. Figure 2(a) and (b) represents the bound conformation and the ligand interaction diagram of natural ligand **10** (pIC50=7.00) and the modelled CDK4. The bound molecule formed two hydrogen bonds with Val 96 of CDK4 and another hydrogen bond with Lys 35. The natural ligand is a diaryl urea derivative, **10** with specific inhibition towards CDK4. Figure 2(c) and (d) characterizes the docked conformation of ligand, **2** (pIC50=8.7959) which had a high biological activity in terms of pIC50 with modelled CDK4. Compound **2** forms two hydrogen bonds with CDK4 at residues Asp99 and Asp158; both were in the hinge region. A salt bridge was also formed between Asp158 and the ligand molecule. These two hydrogen bonding connections together with salt bridge contributes

positively towards its high XPGScore of –8.072, docking score of –7.826 with glide energy of –58.781. Glide XP GScore is the total contribution of Coulombic interactions, binding interactions and van der Waals interactions [17]. Desolvation energy as well as strain energy caused by either ligand molecule or the protein obstructs binding. These penalties were also considered for calculating the XP GScore. It is symbolized as follows:

$$\text{XP GScore} = E_{coul}+E_{vdW}+E_{bind}+E_{penalty}$$
$$E_{bind} = E_{hyd\_enclosure}+E_{hb\_nn\_motif}+E_{hb\_cc\_motif}+E_{PI}+E_{hb\_pair}+E_{phobic\_pair}$$
$$E_{penalty} = E_{desolv}+E_{ligand\_strain}.$$

Observed biological activity and predicted biological activities predicted by both 2D QSAR and 3D QSAR methodology are summarized in Table 2 and the scatter plot is shown in Figure 5.

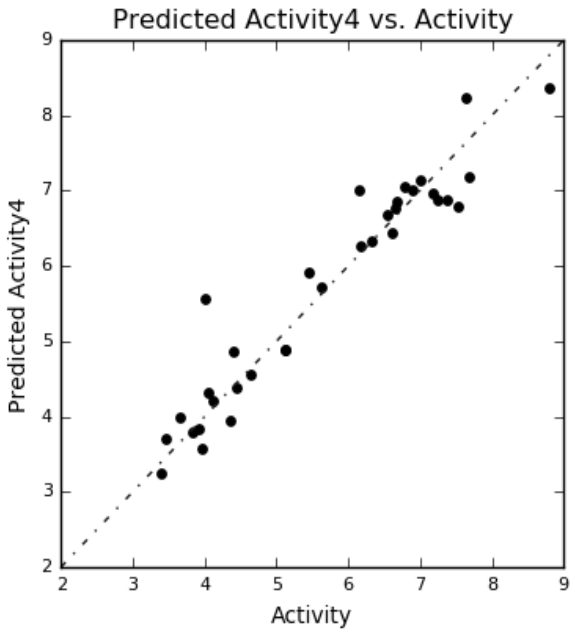


Fig. 5: Scatter plot of Atom Based 3D QSAR model

Predictiveness of the model can be assessed by foretelling the activity of an internal test set and external test set of molecules. The robustness of the model was depend on the R^2_{cv} and R^2_{cv}=($PRESS_0$-PRESS)/($PRESS_0$) where $PRESS_0$ is the mean of the observed biological activity while PRESS is the sum of the squares of the differences between the predicted and the observed activity values [18]. Q^2 is the Predictive correlation coefficient is a measure of predictive ability of the derived QSAR model and is calculated by Q^2 [19]. Q^2= (SD – PRESS) / SD where SD is the sum of squared deviations between the biological activities of the test set molecules and the mean activity of the training set molecules, while PRESS is the sum of squared deviations between the observed and the predicted activities of the test set molecules. A greater F value (Fischer Statistic value) infers that a more noteworthy correlation has been reached. This value can be considered as a degree of the statistical implication of the regression model [20]. It is counted as a norm to define whether a more complex model is significantly better than the other.

Table 2: Experimental and predicted inhibitory activities of CDK4 inhibitors

Title	Set	Y(Obs)	2D QSAR		3D QSAR	
			Y(Pred)	Error	Y(Pred)	Error
1	test	6.6383	6.9581	0.3198	6.6423	0.004
2	train	8.7959	7.7489	–1.047	8.357	–0.4389
3	train	8.1549	7.6977	–0.4572	7.2029	–0.9520
4	train	7.6778	7.2336	–0.4342	7.1773	–0.5005
5	test	7.6383	7.7489	0.1106	8.2422	0.6039
6	train	7.5376	7.2336	–0.304	6.7842	–0.7534
7	train	7.3768	7.4071	0.0303	6.8681	–0.5087
8	train	7.2518	6.8874	–0.3643	6.8692	–0.3826
9	test	7.1871	6.6438	–0.5433	6.9593	–0.2278
10	train	7	6.716	–0.284	7.1362	0.1362
11	train	6.9586	6.8466	–0.112	6.7248	–0.2338
12	test	6.8861	6.8536	–0.0324	7.0125	0.1264
13	train	6.7959	6.8946	0.0987	6.534	–0.1884
14	train	6.7959	6.7073	–0.0886	7.0568	0.2609
15	train	6.6778	6.6646	–0.0132	6.8627	0.1849
16	train	6.6576	6.6646	0.007	6.7753	0.1177
17	train	6.6021	6.2655	–0.3366	6.4314	–0.1607
18	train	6.5376	6.7453	0.2077	6.6921	0.1545
19	test	6.3665	6.7073	0.3407	6.4845	0.1180
20	train	6.3279	6.2884	–0.0395	6.3393	0.0114
21	train	6.1938	6.9522	0.7583	6.2408	0.0469
22	train	6.1739	5.8557	–0.3182	6.2738	0.0999
23	train	6.1427	6.381	0.2383	7.0082	0.8655
24	train	6.0506	6.4888	0.4382	6.1108	0.0602
25	train	5.6198	5.6885	0.0687	5.2709	–0.3489
26	test	5.6198	6.4599	0.8401	5.7231	0.1033
27	train	5.4337	6.5224	1.0787	5.9143	0.4706
28	test	5.4202	6.1872	0.767	5.7141	0.2939
29	train	5.3979	6.4291	1.0312	5.7598	0.3618
30	train	5.1192	5.2328	0.1136	4.8807	–0.2385
31	train	5.1192	5.2328	0.1136	4.8807	–0.2385
32	train	4.6383	4.6038	–0.0343	4.5532	–0.0851
33	train	4.4337	4.197	–0.2467	4.3804	–0.0633
34	test	4.3979	4.8321	0.4341	4.8762	0.4783
35	train	4.3565	3.9429	–0.4137	3.9519	–0.4046
36	train	4.1135	4.5766	0.4631	4.2018	0.0883
37	test	4.0458	3.9236	–0.1222	4.3203	0.2745
38	train	4	4.8517	0.8517	5.5669	1.5669
39	train	3.9586	3.7166	–0.242	3.5797	–0.3789
40	train	3.9208	3.647	–0.2738	3.8295	–0.0913
41	train	3.8239	3.5352	–0.2887	3.8021	–0.0218
42	test	3.6576	3.5555	–0.1021	3.9973	0.3397
43	train	3.4685	4.0912	0.6227	3.7151	0.2466

Table 3: Results of Atom based 3D QSAR

# Factors	4
Standard Deviation	0.3052
R^2	0.9651
R^2_{CV}	0.8191
R^2 Scramble	0.6922
Stability	0.91
F	145.2
P	5.55E–15
RMSE	0.75
Q^2	0.7368
Pearson R	0.8696
Hydrogen bond donor	0.090997
Hydrophobic/non-polar	0.577229
Positive ionic	0.02014
Electron withdrawing	0.31158

From the analysis it is found that R^2 for regression is 0.9651and cross validated R^2 computed from predictions obtained by a leave one out (LOO) procedure is 0.8191. High values of both R^2 and R^2_{cv} are indicators of the high predictive ability of the model. Stability nearer to one (0.91) indicates the model predictions should be stable. Greater value of variance ratio (F=145.2) specifies high statistical significance. Greater degree of the confidence of the model was indicated by smaller value of P (5.55E–15).Q^2 is the Predictive correlation coefficient is a measure of predictive ability of the derived QSAR model and in the present study; it is found that Q^2 is 0.7368. This result was pointing to fact that the proposed model has best predictive ability.

The field contributions of factors of Hydrophobic or nonpolar influence were 57.7% and 31.1% respectively signifying that the hydrophobic or nonpolar fields have greater impact on the model.

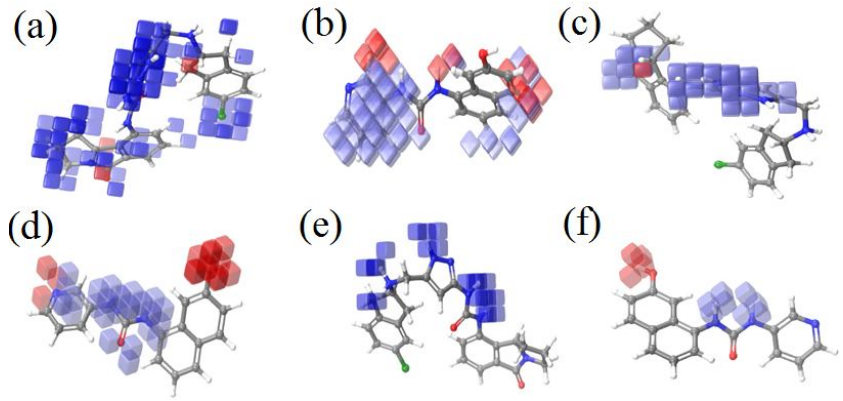


Fig. 6: Contour maps of atom based QSAR model in combination with compound 2 and 43 (a) and (b) hydrophobic/ nonpolar feature, (c) and (d) Electron withdrawing effect and (e) and (f) hydrogen bond donor feature

In atom based QSAR, a molecule is treated as a set of overlapping van der Waals spheres [21]. The consequence of spatial arrangement of structural features on CDK4 can be visualized in the Contour plot analysis. Figure 6 (a) and (c) represents the contour maps of atom based QSAR model of compound **2** of highest activity (pIC50=8.7959) and Figure 6(b) and (d) represent that of compound **43**(pIC50=3.4685). Here, blue cubes represent positive coefficient towards the inhibition and the red cubes represent negative coefficient towards inhibition of CDK4. From Figure 6(a) which represents the hydrophobic or non-polar feature, the blue cubes around benzopyrrolizin–5–one and the pyrazole

moiety contribute positively towards its hydrophobic nature whereas the carbonyl group in the 5th position of benzopyrrolizine contributes negatively. Influence of chlorine is remarkable towards inhibitory activity as it is covered with blue contours. Blue contours which favor better activity were absent in the case of **43** and this is supported by its low pIC50. Introduction of carbons, halogens and C–H in the regions favouring hydrophophic/ nonpolar regions may enhance the inhibitory activity of the molecule. Other significant component that impacts on the activity is the electron withdrawing character which includes different types of hybridization, hydrogen bond acceptor features, л-electron donation features etc and is shown in Figure.6 (c) and (d). Presence of diamide linkage, pyrazole moiety and benzopyrrolizin–5–one contribute positively towards the biological activity of the compound **2**whereas the presence of hydroxyl group in the naphthyl ring of **43** has a negative electron withdrawing effect and thus it has a lower bioactivity. Insertion of non-ionic nitrogens and oxygens may intensify pIC50 value. Hydrogens attached to polar atoms which are categorized as hydrogen bond donor attributes also contribute significantly towards the inhibitory activity (Figure 6(e) and (f)). High activity of **2** is also supported by the presence of hydrogen bonding groups as it is specified by the blue contour areas. Compound 43 lacks hydrogen bonding feature at certain specific areas. Nitrogen in the pyrrazole ring forms a hydrogen bond with Asp158 as per the docking analysis. Also, hydroxyl group (–OH) at the 7th position of the naphthyl ring reduces the inhibitory effect.

4. CONCLUSIONS

A set of molecules with CDK4 specific inhibition was examined to relate pIC50 values against CDK4 to the molecular structure. High predictive ability of 2D and 3D QSAR models was substantiated by leave one out (LOO) and internal and external validation procedures. The performance of KPLS approach was resulted in 10 best predictive models with an average score value of 0.8561. 3D QSAR results proved that electron withdrawing, hydrogen bonding as well as hydrophobic non-polar interactions affects considerably the biological activity. The scrutiny of the 3D contour maps permitted us to spot areas of identified inhibitors that require a physiochemical features to increase activity. Docking study of compound **2** support these results.

The biologically active conformation of the inhibitor molecules was provided by the docking study of the homology model of CDK4. This helped in the substantiation of QSAR models with respect to the protein environment and give exquisite perceptions for the structure activity explanations. With the aid of this information, we can design new active molecules with enhanced pharmacological properties.

Acknowledgement

We extend our sincere thanks to Application Scientists of Schrodinger, Bangalore for providing technical assistance for completing this work. We are grateful to University Grant Commission for sanctioning grant under FDP for doing this work.

REFERENCES

[1] Roberts, Charles J. Sherr and James M., GENES & DEVELOPMENT, **2004,** 2699–2711.
[2] Peters, Marcia Hall and Gordon, 67–106.
[3] Xavier Grana, E. Premkumar Reddy, Oncogene, **1995,** 211–219.
[4] Taya, Y., Mol. Cells, 5, **1995,** 191–195.
[5] Weinberg, R. A., Cell, 81, **1995,** 323–330.
[6] Dixon SL, Duan J, Smith E, Von Bargen CD et.al., Future Med. Chem., , **2016,** 1825–1839.
[7] Teruki Honma, Kyoko Hayashi, Tetsuya Aoyama, Noriaki Hashimoto et. al., J. Med. Chem., 44, **2001,** 4615–4627.
[8] Teruki Honma, Kyoko Hayashi, Tetsuya Aoyama, Noriaki Hashimoto et. al., J. Med. Chem., 44, **2001,** 4615–4627.

[9] Teruki Honma, Takashi Yoshizumi, Noriaki Hashimoto, Kyoko Hayashi et.al., J. Med. Chem., 44, **2001,** 4628–4640.
[10] Campbell McInnes, Shudong Wang et.al. Chemistry & Biology, 11, **2004,** 525–534.
[11] Joseph Rebehmed, Florent Barbault et.al. J. Comput. Aided Mol. Des., **2008,** 831–841.
[12] Ikuta, M., Kamata, K., Fukasawa, K., Honma, T. et.al. J. Biol. Chem., **2001,** 27548–27554.
[13] Jitender Verma, Vijay M. Khedkar and Evans C. Coutinho., Current Topics in Medicinal Chemistry, **2010,** 95–115.
[14] Campbell McInnes C, Shudong Wang et.al. Chemistry and Biology, **2004,** 525–534.
[15] Akamatsu, M., Curr.Top. Med. Chem., **2002,** 1381–1394.
[16] Kim, K.H., BlackieAcademic & Professional: Glasgow, UK, **1995,** 291–331.
[17] Richard A. Friesner, Robert B. Murphy, Matthew P. Repasky et.al., J. Med. Chem.,**2006,** 6177–6196.
[18] Dunteman, G.H., SAGE Publications Ltd: London, **1989,** 15–22.
[19] Marshall, G.R., Springer Publications, London, UK, **1998,** 80–116.
[20] Archdeacon, T.J., Univ of Wisconsin Press: USA, **1994,** 160–177.
[21] Dixon, S L., Journal of Computer–Aided Molecular Design, **2006,** 647–671.