Pharmacophore optimization of substituted N-phenyl-2,2-dichloroacetamide using molecular modelling studies, design, ADMET prediction, synthesis and evaluation of potential anti-cancer agents.

Shashikant V. Bhandari (drugdesign1@gmail.com)
AISSMS College of Pharmacy
Om G. Nagras
AISSMS College of Pharmacy
Pranali V. Kuthe
AISSMS College of Pharmacy
Aniket P. Sarkate
Dr. Babasaheb Ambedkar Marathwada University
Vrushali P. Pathak
AISSMS College of Pharmacy

Research Article

Keywords: Molecular Modeling, Anticancer, ADMET prediction, QSAR, Docking, Synthesis, MTT assay

Posted Date: May 23rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1644580/v1

Abstract

A current study shows that N-phenyl-2, 2-dichloroacetamide analogues can induce cancer cell apoptosis and inhibit tumor growth by inhibiting the enzyme Pyruvate Dehydrogenase Kinase (PDK), but they have not been proved as clinically promising anticancer agents due to their low biological activity as well as toxicity profile. Hence Two-dimensional (2D), Three-dimensional (3D) and Group based (G) Quantitative Structure Activity Relationship (QSAR) studies were performed to correlate the chemical composition of N-phenyl-2,2-dichloroacetamide analogues with their anticancer activity in order to optimize the pharmacophore requirements for potent PDK inhibitors. New Molecular Entities (NMEs) were designed using the results of all QSAR studies. Binding affinities of designed NMEs were studied on PDK II & PDK I enzymes (PDB code: 2BU8 & 2Q8H respectively) using molecular docking studies. The ADMET properties of designed NMEs were also predicted to ensure their drug like pharmacokinetic profile. Results of molecular modeling studies were cross verified by synthesis and testing the designed compounds by cytotoxicity assay using A549 cell line. Compound D4 exhibited very significant anti proliferative effect. The potency in order was observed as below compounds [Most potent] D4(IC_{50} = 0.167μm/ml) > D5(IC_{50} = 0.191μm/ml) > D8(IC_{50} = 0.218μm/ml) > D9(IC_{50} = 0.238μm/ml) > D11(IC_{50} = 0.290μm/ml) > D12(IC_{50} = 0.409μm/ml) > [Least potent] D14(IC_{50} = 0.635μm/ml). All the designed compounds were found to be more potent as compared to most potent compounds of the selected series. Thus, compound D4 and D5 with potent inhibitory activity in tumor growth inhibition were found be a potential anticancer agent.

1. Introduction

Cancer is the largest cause of death in the world, with approximately 10 million fatalities expected by 2020. Breast (2.26 million instances), lung (2.21 million cases), colon and rectum (1.93 million cases), prostate (1.41 million cases), skin (non-melanoma) (1.20 million cases), and stomach (1.20 million cases) were the most frequent cancers in 2020. (1.09 million cases) [46]. Anticancer drug discovery has undergone a remarkable series of changes over the last decade. The first generation of anti-cancer drugs such as DNA alkylating and cross-linking agents, anti-metabolites, topoisomerase inhibitors and anti-tubulin agents have been traditionally focused on targeting DNA processing and cell division and were almost all cytotoxic agents [1, 2]. In an attempt to avoid unpleasant side effects (bone marrow suppression and gastrointestinal, cardiac, hepatic, and renal toxicities) associated with these conventional anticancer drugs, a new class of anti-cancer drugs known as molecularly targeted agents was being developed that work by targeting a biochemical/metabolic pathway or the protein that is unique to or up regulated in cancer cells [3]. Such agents are typically less toxic than drugs in the older classes and can be given for long-term oral therapy with the objective of treating cancer as a chronic disease. Among all these non-traditional (non-DNA-directed) cancer targets for which pharmacological intervention is feasible, there are none that have generated as much widespread interest, as have the protein kinases inhibitors [4].

Pyruvate Dehydrogenase kinase (PDK) enzyme is a new era in the research area of potent, selective and non-toxic anticancer agents. Pyruvate Dehydrogenase Kinase (PDK) plays a key role in cancer cellular proliferation and apoptosis resistance and hence it has proved to be an effective target for development of anticancer agents [20, 21]. Number of compounds have been reported as orally effective PDK inhibitors, but none has proved a promising anticancer agent [4, 5, 6] (Fig. 1)

Dichloroacetamide (DCA) a small molecule is an orally effective classical mitochondrial PDK inhibitor [1 – 7]. In 1973, it was found to be capable of activating mitochondrial Pyruvate Dehydrogenase (PDH) by inhibiting PDK [1, 8] and since then it has been used mainly to treat acquired and congenital forms of lactic acidosis [9 – 13] as well as other mitochondrial diseases [2, 14 – 19]. Recently, it has been reported that DCA can induce cancer cell apoptosis, reduce cellular proliferation and suppress tumor growth [20, 21]. In DCA-treated cancer cells, the cell metabolism is shifted from abnormal cytoplasmic glycolysis to mitochondrial glucose oxidation through inhibition of PDK and the expression of the K^+ channel Kv1.5 is increased through NFAT1 (Nuclear Factor Activation T-cells)-dependent mechanism which is suppressed in cancer [20]. Finally, cytochrome C and Apoptosis Inducing Factor (AIF) are released from mitochondria to cytoplasm, which activate caspases and induces apoptosis [20]. It has also been reported that DCA could induce apoptosis of endometrial cancer cells [22], inhibit metastatic breast cancer cell growth *in vitro* and *in vivo* [23], induce apoptosis and arrest cell-cycle in colorectal cancer cells, inhibits neuroblastoma and helps to radiate prostate cancer cells *in vitro* [24] by similar mechanism.

However, DCA has not been marketed as a promising clinical anticancer agent due to its very low biological activity (IC_{50} > 1000μM) and major toxic effects on CNS like neurotoxicity as well as peripheral neuropathy as it blocks myelin formation and hence modification of DCA nucleus became necessary [12]. Thus, in an attempt to increase its potency, various derivatives of DCA were synthesized out of which N-phenyl-2, 2-dichloroacetamide and its analogues were found to be much more potent than DCA [25].

One of the main objectives in current drug design scenario is the prediction of new potent and selective drug like biologically active compounds on the basis of previously synthesized ones in an attempt to save the cost of blind synthetic process, manual labor as well as time. Molecular Modeling studies is an approach which is used to narrow down a library containing an extraordinarily high number of random molecules into a smaller list of the potentially effective inhibitors. Thus, we have focused our aim on molecular modeling of anticancer agents containing N-phenyl-2, 2-dichloroacetamide nucleus with simultaneous goal of their enhanced performance against PDK enzyme. For achieving this aim and optimizing the pharmacophore requirements, we first carried out Quantitative Structure–Activity Relationships (QSARs). QSAR studies help us to correlate the physicochemical properties of molecules with their biological activities, which can enlighten their key structural components which are important for significant biological activity. One of the important roles of QSAR methodology is to predict the biological activity of in-silico designed Chemical Entities [26]. In the present study, we have discussed the best possible practices for developing robust and externally predictive QSAR models and

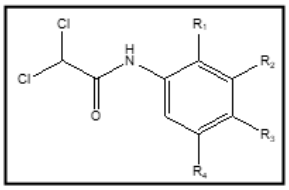
successful applications of QSAR studies for developing NMEs exhibiting anticancer activity. 2D, 3D & G-QSAR studies were carried out using predictive QSAR modeling method [30] and NMEs were designed using the results of best QSAR model of all three studies.

Designed NMEs were subjected to molecular docking analysis to determine their binding mode and interactions with PDK target. PDK is available in four different isoforms out of which DCA binds to PDK II with more affinity compared to others (PDK II = PDK IV > PDK I >>> PDK III) [3, 41]. Since apart from PDK II, DCA also binds with PDK I as all isoforms are present in the human body in different organs [3]; it is necessary to study the binding mode of NMEs with all available crystal protein structures of PDK having DCA as co-crystallized ligand. As standard binding mode of DCA with PDK II (PDB code: 2BU8) & PDK I (PDB Code: 2Q8H) have been reported in literature, docking studies were performed with both of these receptor protein structures. Rational drug design should take both pharmacokinetic and metabolic information into consideration and the information should be incorporated with molecular, biochemical and pharmacological data to provide to ensure drug like pharmacokinetic profile. Hence, prediction of pharmacokinetic properties (ADMET) was used as the last screen to sort out those designed NMEs which follow Lipinski's rule and other ADMET ideal ranges to ensure their drug like pharmacokinetic profile.

2. Chemistry

2.1. Biological dataset

For this study, total 32 molecules of N-phenyl-2,2-dichloroacetamide series reported for anticancer activity on non-small cell lung cancer cell line (A549) (Table 1) has been chosen to develop QSAR models [25].



Sr. No	Compound	R_1	R_2	R_3	R_4	IC_{50} (µm/ml)	pIC_{50}
1.	5c	-H	-Cl	-H	-Cl	4.40	-0.64
2.	3e	-H	-I	-H	-H	4.76	-0.68
3.	5b	-H	-Cl	-Cl	-H	5.09	-0.71
4.	3q	-H	$-SO_2Ph$ ($NHCOCHCl_2$)	-H	-H	5.89	-0.77
5.	3m	-H	$-SO_2CF_3$	-H	-H	6.53	-0.81
6.	3d	-H	-Br	-H	-H	7.80	-0.89
7.	4i	-H	-H	$-SO_2CF_3$	-H	8.50	-0.93
8.	4h	-H	-H	$-SCF_3$	-H	10.56	-1.02
9.	3g	-H	-C≡CH	-H	-H	11.22	-1.05
10.	3j	-H	$-CF_3$	-H	-H	12.18	-1.09
11.	5d	-Cl	-H	-H	-Cl	12.38	-1.092
12.	4e	-H	-H	-I	-H	12.54	-1.10
13.	4d	-H	-H	-Br	-H	13.74	-1.14
14.	3l	-H	$-SCF_3$	-H	-H	14.07	-1.15
15.	3c	-H	-Cl	-H	-H	14.78	-1.17
16.	3k	-H	$-OCF_3$	-H	-H	15.05	-1.18
17.	3h	-H	$-NO_2$	-H	-H	18.04	-1.26
18.	4c	-H	-H	-Cl	-H	20.37	-1.31
19.	3i	-H	$-OCH_3$	-H	-H	25.78	-1.41
20.	4f	-H	-H	$-NO_2$	-H	34.42	-1.54
21.	2d	-Br	-H	-H	-H	36.99	-1.57
22.	3a	-H	$-CH_3$	-H	-H	41.05	-1.61
23.	4a	-H	-H	$-CH_3$	-H	48.92	-1.69
24.	2c	-Cl	-H	-H	-H	50.15	-1.70
25.	3f	-H	-C≡N	-H	-H	66.53	-1.82
26.	4b	-H	-H	-F	-H	68.41	-1.84
27.	4g	-H	-H	$-OCH_3$	-H	80.88	-1.91
28.	2b	-F	-H	-H	-H	87.53	-1.94
29.	1	-H	-H	-H	-H	130	-2.11
30.	2f	$-NHCOCHCl_2$	-H	-H	-H	130.94	-2.12
31.	2e	$-NO_2$	-H	-H	-H	197.56	-2.30
32.	DCA	-	-	-	-	1011	-3.0048

DCA = Dichloroacetate

$pIC_{50} = \log 1/IC_{50}$

Table-1 Selected series of compounds containing N-phenyl-2, 2-dichloroacetamide pharmacophore

2.2. **Molecular modeling tools**

All QSAR studies were performed using VLife Molecular Design Suite 3.5 & 4.0 [31]. Molecules were optimized by Merck Molecular Force Field (MMFF) energy minimization method [32].

2.3. Two-dimensional QSAR (2D QSAR) studies

2.3.1. Experimental design for 2D QSAR

Dataset of 32 molecules was divided into multiple training and test sets by using manual data selection method and no. of sets were generated using different combinations of molecules in training and test sets such that it covers each molecule in different set every time in an attempt to ensure robustness of QSAR model and increase its predictive ability [27]. From these sets, training & test sets which followed all model evaluation parameters were subjected to Y-randomization test (Fig. 2.).

Evaluation parameters of Y-randomization test:

n = Number of molecules

df = Degree of freedom (n-k-1) (higher is better)

k = Number of descriptors in a model ($\leq n/5$)

r^2 = Coefficient of determination (>0.7)

q^2 = Cross validated r^2 (>0.5)

SEE = Standard error of estimate (smaller is better)

Pred_r^2 = r^2 of external test set (>0.5)

F-test = Statistical significance of the model (higher is better for same descriptors and compounds)

Best_ran_r^2 = Highest r^2 value in the Y-randomization test (as low as compared to r^2)

Best_ran_q^2 = Highest q^2 value in the Y-randomization test (as low as compared to q^2)

Z score = It is calculated by the Y-randomization test (higher is better)

Alpha = Statistical significance parameter by randomization test (<0.01)

Three models were selected which satisfied the results of Y-randomization test and were named as Training set-A, Training set-B and Training set-C. These models were subjected to two times for external validation by splitting them into two test sets viz. test set a_1, a_2 for training set-A, test set-b_1, b_2 for training set-B and test set c_1, c_2 for training set-C in order to avoid the chance correlated results. Only those models which satisfied both the test sets were selected for design of NMEs. We have ensured that selected training and test sets also satisfied the following criteria:

a. Representative points of the test set must be close to those of the training set;
b. Representative points of the training set must be close to representative points of the test set;
c. Training set must have wide chemical and biological diversity.

2.3.2 Uni-column statistics

The comparative statistical parameters of training and test sets created by manual data selection method are reported in **Table 2**. The minimum and maximum values in both training and test set should be compared in a way that-

i. The maximum of the test should be less than max of training set.
ii. The minimum of the test should be greater than min of training set.

It shows that the test set is interpolative i.e., derived within the min-max range of the training set. The mean and standard deviation of the training and test set provides insight to the relative difference of mean and point density distribution (along mean) of the two sets. Standard deviation of Training set A, B and C with test set a_1, a_2; b_1, b_2 and c_1, c_2 respectively were found to be nearly close to each other. This showed that even though the selected molecules in training or test sets are different, but the distribution pattern with respect to the biological activity of molecules in both the selection methods is quite similar.

Para-meters	Training Set B+C	Test set a_1	Test set a_1	Training Set A+C	Test set b_1	Test set b_2	Training Set A+B	Test set c_1	Test set c_2
Avg.	0.9653	0.8400	-1.0479	-0.9858	-0.8363	-0.9305	-0.9127	-0.8722	-0.8103
Max	0.0128	-0.4389	-0.6421	0.0128	-0.6675	-0.8311	-0.002	-0.4412	-0.2203
Min	-1.5923	-1.2149	-1.1155	-1.5923	-0.2331	-0.1089	-1.6212	-1.3274	-1.3788
S.D.	0.5629	0.5134	0.5928	0.5868	0.5626	0.5437	0.5549	0.6034	0.5273

Avg- Average

S.D.- Standard deviation

Table 2 Uni-Column statistics for training sets and test sets

2.3.3 Descriptor selection

Various 2D descriptors (a total of 338) were calculated and preprocessing of them was carried out by removing invariable columns.

It has been reported that there is high probability of chance correlation between the observed and predictive activity; especially when no. of descriptors is comparable or more than the no. of compounds in dataset for any QSAR analysis [31]. Thus, reduction in no. of descriptors is a very important step which is required to avoid the occurrences of chance correlation and inclusion of irrelevant descriptors in final QSAR model. We applied combinations of different descriptor selection methods viz. forward, forward-backward, genetic algorithm, simulated annealing etc. as well as different QSAR methods on same molecule sets and finally considered the results of forward variable selection method with Multiple Linear Regression (MLR) after comparing all results to improve performance as well as predictability of QSAR model.

2.3.3.1 Correlation matrix

It is very popular and crucial technique used for QSAR studies [28]. We have considered the correlation between descriptor with activity as well as their inter-correlation i.e., descriptor-descriptor correlation. We have shown only those descriptors contributed for the selected series of compounds in 2D QSAR studies; which show either direct or inverse correlation with biological activity.

DESCRIP-TORS	$Xlog_{10}P$	T_N_N_3	SssSE_index	T_N_Cl_4	T_O_O_2	SddssS(sulphate)count	T_O_Cl_6	T_2_S_0
	Common (set I & II)		Descriptor set - I			Descriptor set – II		
$Xlog_{10}P$	1.00	-	-	-	-	-	-	-
SssSE_index	-0.407	-	1.00	-	-	-	-	-
T_N_N_3	-0.138	1.00	0.067	-	-	-	-	-
T_N_Cl_4	0.275	-0.093	0.092	1.00	-	-	-	-
T_O_O_2	-0.396	0.176	0.137	-0.188	1.00	-	-	-
SddssS(sulphate) Count	0.125	-0.083	0.121	-0.002	-0.21	1.00	-	-
T_O_Cl_6	0.275	-0.092	0.087	-0.017	-0.013	-0.151	1.00	-
T_2_S_0	0.135	-0.075	0.183	0.009	0.142	0.103	0.213	1.00
pIC_{50}	0.805	-0.428	0.147	0.358	-0.122	0.347	0.358	0.332

Table 3 Correlation matrix of descriptors (2D QSAR; set-I & set-II)

2.3.3.2 Fitness plot

Correlation coefficient cannot give information about data spread between the descriptor and activity. There may be some descriptors showing chance correlation with activity because each variable selection method is based upon correlation between descriptor and activity and not on the type of data spread. To avoid above said pitfall, the proper observation of fitness plot between descriptor and activity is needed [**Fig 3**].

The following are few important points that we have taken into consideration while selecting proper descriptors for QSAR model generation:

- We have ensured that percentage distribution of data points on both sides of best fit line should be nearly 50-50%. (We preferred slope value more than 0.15)

- In case of topological descriptor, no. of occurrences of particular data point was observed in fitness plot which gave information about the frequency of occurrence of each particular substituent in series. Thus, although the particular descriptor shows good correlation with activity as well as comes in the QSAR final model result, but we cannot take it into final consideration unless and until it shows well spread fitness plot. In conclusion, we can say that careful observation and right analysis of a fitness plot helped us to reduce no. of descriptors.

2.3.3.3 Variance

Another significant way to find out unimportant descriptors is by using information of variance of descriptors [39]. There were some descriptors which showed consistently high variance even if there was small change in physicochemical properties and vice versa. After close analysis of the output of our study, we conclude that we should focus more on correlation between descriptors and activity instead of considering descriptors of highest variance, as final results rely more on correlation than on variance.

The algorithm we followed for variable reduction is as follows –

1. Define appropriate correlation cutoff value between descriptor & activity, which is mentioned as $A_{max.}$ Remove all descriptors which have value less than A_{max}
2. Define appropriate cross correlation cutoff value between descriptor-descriptor, which is defined as $C_{max.}$ Remove all descriptors which have values larger than $C_{max.}$
3. Define variance cutoff value for descriptor which is mentioned as V_{max}. The descriptors having variance value less than V_{max} were removed.

We observed that this algorithm reduced no. of descriptors nearly up to 50%. After this, we applied manual variable selection method and multiple linear regression (MLR) as it is ensured that each remaining descriptor is significantly contributing for QSAR model. The only thing we have to find out is no. of descriptors in final equation should be as low as possible which must be contributing highly and should be seen in the structural features of the reported compounds of the series as well [29].

2.4 **Three-dimensional QSAR (3D QSAR) studies**

2.4.1 Alignment of molecules

Proper alignment of molecules is the prerequisite for studying 3D QSAR as well as in almost all the fields of drug discovery for getting reliable results. So, after optimization we carried out alignment of all molecules using MolSign which also serves as the basic tool to identify the common pharmacophore features as well as the individual molecular feature.

The color scheme for identification of various chemical features is as follows:

Hydrogen bond donor: magenta color

Hydrogen bond acceptor: Buff color

Hydrophobic: Orange color

Aliphatic: Orange color

Negative ionizable: green color

Positive ionizable: Violet color

The larger tessellated spheres are indicative of the common pharmacophore features identified in the molecules and the smaller solid features are of the individual molecules.

2.4.2 3D QSAR by SA-kNN-MFA

3D QSAR studies were performed by generation of numerous models by taking same molecules in the respective training and test sets as in 2D QSAR by using k-Nearest Neighbor-Molecular Field Analysis (kNN–MFA) methodology with Simulated Annealing (SA) variable selection method as it has been reported as the more relevant and suitable method to perform 3D QSAR [36, 37, 40]. kNN–MFA requires suitable alignment of given set of molecules after optimization which had already been carried out by MolSign, but it was again carried out to generate a folder of aligned molecules to proceed for 3D QSAR by atom-based alignment which gives alignment based on each and every individual atom of the pharmacophore. Molecular alignment was used to visualize the structural diversity in the given set of molecules. It was followed by generation of common rectangular grid around the molecules. Steric and electrostatic interaction energies were computed at the lattice points of the grid using a methyl probe of charge +1. These interaction energy values at the grid points are considered for relationship generation using kNN method and utilized as descriptors for obtaining distances within this method. Resulting set of aligned molecules was then used to build 3D QSAR model.

2.5 **Group based QSAR (G-QSAR) studies**

Group based QSAR (G-QSAR) is a new approach to investigate structure activity relationship based on molecular fragments of a set of molecules that significantly enhances the use of QSAR as an approach for new molecule design. As a predictive tool for activity, G-QSAR is significantly superior to conventional 2D and 3D QSAR as in this method, every molecule of the data set is considered as a set of fragments, the fragmentation scheme being either template based or user defined. The descriptors are evaluated for each fragment and a relationship between these fragment descriptors is formed with the activity of the whole molecule. Unlike conventional QSAR, with the G-QSAR, we get critically important site-specific clues within a molecule where a particular descriptor needs to be modified. G-QSAR studies were also performed by generation of multiple models of the same training & test sets as in 2D & 3D QSAR by using MLR analysis. G-QSAR was carried out using template-based fragmentation scheme and forward variable selection method.

2.6 Design of New Molecular Entities (NMEs) containing N-phenyl-2, 2-dichloroacetamide pharmacophore

The information obtained from 2D, 3D and G-QSAR studies was utilized in optimizing N-phenyl-2, 2-dichloroacetamide pharmacophore and to design potent anticancer chemical entities Substitution pattern around pharmacophore, shown in **Fig. 10** was used to design designed chemical entites using Combilib tool of VLife MDS software. Designed compounds were subjected to Lipinski's screen [38] to ensure their drug like pharmacokinetic profile in order to ensure their drug like bioavailability. The following parameters were used as Lipinski's filters (values in parenthesis indicate ideal requirements):

1. Number of Hydrogen Bond Acceptor (A) (<10)

2. Number of Hydrogen Bond donor (D) (<5)

3. Number of Rotatable Bond (R) (<10)

4. XlogP (X) (<5)

5. Molecular weight (W) (<500 g/mol)

6. Polar surface area (S) is (<140 Á)

The fig. 10 (pg. no.33) indicates the structural substitution pattern required around N-phenyl-2, 2-dichloroacetamide based on the result of 2D and 3D QSAR studies.

2.7 Molecular Docking studies

All the designed compounds that showed good predicted activity by all QSAR studies and followed Lipinski's rule as well as reported series of molecules for comparison purpose were subjected to molecular docking for studying the binding mode of designed compounds and were further screened to sort out the best compounds having good binding affinity compared with binding mode of standard DCA. The main molecular docking tool used was GLIDE (Maestro; Schrödinger Inc., USA) for protein-ligand docking studies in to the receptor Pyruvate Dehydrogenase Kinase (PDK) enzyme binding pocket [43]. The crystal structures of PDKs were obtained from Protein Data Bank. (PDK II-PDB Code: 2BU8; PDK I-PDB Code: 2Q8H) [42]. All structures were prepared for docking using 'Protein preparation wizard' and 'Ligand preparation wizard' in Maestro wizard of Schrödinger 10.2. In the refinement component, a restrained impact minimization of the co-crystallized complex was carried out. This helps in reorientation of side chain hydroxyl groups. It uses the OPLS-AA force field for this purpose. The co-crystallized ligand was removed from active site and the grids were defined by centering them on the ligand in the crystal structure. Then our structures (ligands) were imported in the project table, built using maestro structure builder panel and prepared by Ligprep module which produces the low energy conformers of ligands using MMFF94 force field. The lower energy conformations of the ligands were selected and docked into the grid generated from protein structure using extra precision (XP) docking mode. In this docking method, the ligands are flexible and receptor is rigid, except the active binding site of protein which has slight flexibility. The final evaluation is done with glide score (docking score) and single best pose is generated as the output for particular ligand.

G-score = a*vdw + b*coul + Lipo + H-bond + Metal + BuryP + RotB + Site ……equation 1

where, vdW, Van der Waals energy; Coul_Coulombic energy; Lipo-Lipophilic contact term; HBond, Hydrogen-bonding term; Metal_Metal-binding term; BuryP-Penalty for buried polar groups; RotB, Penalty for freezing rotatable bonds; Site, Polar interactions at the active site. The coefficients of vdW and Coul are: a = 0.065, b = 0.130 respectively.

The accurate prediction of protein-ligand interaction geometries is essential for the success of virtual screening approach in structure-based drug design. The docking results were evaluated based on Glide Score (G-Score), Hydrogen bonds (H-bond) and Vander Waals (vdW) interactions between ligand and receptor.

2.7.1 Molecular docking studies using other software tools (VLife MDS 3.5 & Molegro Virtual Docker, 2007)

After consideration of docking results obtained using glide (Schrödinger), same compounds were again subjected to molecular docking studies by using VLife MDS 3.5 & MVD 2007 to only check out whether the obtained docking score for all the compounds are in similar order with that of the

obtained by glide or not. The results of docking studies using glide, Schrödinger are shown in Table no.10 and Figure no.11 to 15.

2.7.1.1 VLife MDS 3.5

In VLife MDS 3.5, the compounds were subjected to grid-based batch docking with keeping fitness function as dock score and by setting all the parameters to their default values.

Grid batch docking parameters are:

Default grid interval size = 1 O

Rotation angle for ligand rotation = 100

No. of bumps allowed = 4

Specify cavity = 1

Log = Enable logging

VLife MDS uses following fitness functions for calculating docking score:

E (docking score) = InterEq + InterEvdW + IntraEq + IntravdW + IntraEtor.... equation 2

Where, InterEq = Intermolecular electrostatic energy of complex

InterEvdW = Intermolecular vdW energy of complex

IntraEq = Intramolecular electrostatic energy of ligand

IntraEvdW = Intramolecular vdW energy of ligand

IntraEtor = Intramolecular torsion energy of ligand

All the energy components were calculated using MMFF force field.

2.7.1.2 Molegro Virtual Docker, 2007

In MVD 2007, initially the cavity detection was carried out followed by which the compounds in. mol format were imported in the same workshop and docking wizard was set according to the binding site cavity selection and volume size of the axis adjustment. Customize search algorithm parameters were set at their default values. After checking again for any warning with the whole setup, the docking wizard was started as a separate process.

Customize search algorithm parameters are:

Algorithm = MolDock optimizer

No. of runs = 10

Population size = 50

Max iterations = 2000

Scaling factor = 0.50

Crossover rate = 0.90

Maximum No. of multiple poses of each run = 5

MVD uses following equation for calculating MolDock score:

$$E_{score} = E_{inter} + E_{intra}$$.... equation 3

Where, E_{inter} = Ligand-Protein interaction energy

E_{intra} = Internal energy of the ligand

2.8 **Prediction of ADMET properties**

Sometimes compounds that show very high activity in vitro however are proved later to have no in vivo activity, or to be highly toxic in in-vivo models. Lack of in vivo activity may be attributed to undesirable pharmacokinetic properties and the toxicity may result from the formation of reactive metabolites. The failure of NMEs at latter stages of drug discovery process due to lack of drug like pharmacokinetic profile has forced us to set filters of ADMET properties. Thus, we have ensured that only drug-like NMEs would be selected for experimental validation. All designed compounds which showed good binding affinity were filtered by predicting their Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) properties by means of QikProp tool of Schrödinger 10.2 [44] as well as Discovery Studio (DS), Accelrys [45]. Prediction of ADMET properties was used as the last screen to sort out those compounds which already followed Lipinski's rule, showed good predicted activity as well as good binding affinity with PDK enzymes.

2.8.1 ADMET prediction by QikProp, Schrödinger 10.2

In order to ensure druglike pharmacokinetic profile, ADMET properties were predicted. QikProp tool of Schrödinger predicts both physicochemical significant descriptors and pharmacokinetically relevant properties. It also evaluates the acceptability of analogues based on Lipinski's rule of 5, which is essential to ensure drug like pharmacokinetic profile while using rational drug design. All the analogues were neutralized before being used by QikProp. This program is designed using the BOSS program and the OPLS-AA force field. It uses Monte Carlo statistical mechanics simulations on organic solutes in periodic boxes of explicit water molecules to perform all predictions. This process resulted in configurationally averages for a no. of descriptors, including H-bond counts and solvent-accessible surface area (SASA). Correlations of these descriptors to determine properties experimentally were obtained and then algorithms that mimic the full Monte Carlo simulations and produce comparable results were developed by the QikProp tool.

2.8.2 ADMET prediction by Discovery Studio, Accelrys

DS provides the methods for assessing the disposition and potential toxicity of a ligand within an organism as the ADMET protocols contain published models which are used to compute and analyze ADMET properties. Thus, ADMET properties specify the rules to remove ligands which are not drug-like, unsuitable leads etc. based on the presence or absence and frequency of certain chemical groups.

3. Results And Discussion

3.1 2D QSAR models

Using MLR, two sets of 5 meaningful descriptors having two common descriptors i.e., XlogP and T_N_N_3 were obtained. So, total 8 descriptors were considered for final results, out of which XlogP showed up to 55% contribution for anticancer activity.

$pIC_{50} = 0.6875(\pm 0.0244)\ XlogP - 3.1211$

$r^2 = 0.71$, $q^2 = 0.64$, F-test = 25.21, $Pred\_r^2 = 0.56$.

From the results, it is known that XlogP alone satisfies all evaluation parameters. It shows highest correlation with activity (as shown in correlation matrix) and also shows proper distribution of data points [Fig. 2.3.3.2 (a)]. To increase the predictive power, different combinations of descriptors were made by keeping XlogP and T_N_N_3 (second highest & negatively contributing) as constant descriptors as mentioned above. Two models from training set-A, B and C were generated in both the descriptor sets.

Table 4
Statistical parameters of developed QSAR models for descriptor set-I by forward variable selection method and MLR

Statistical Parameters	Training Set A		Training Set B		Training Set C	
	Test Set a_1	Test Set a_2	Test Set b_1	Test Set b_2	Test Set c_1	Test Set c_2
N	20	20	20	20	26	24
df (n-k-1)	14	14	14	14	20	18
r^2	0.8586	0.8586	0.8975	0.8975	0.8684	0.7945
r^2 se	0.1342	0.2541	0.0753	0.1723	0.1253	0.2016
q^2	0.7757	0.7757	0.8011	0.8011	0.7874	0.7507
q^2 se	0.1352	0.2541	0.0462	0.1963	0.2653	0.2121
F-Test	18.75	24.51	16.99	19.03	26.76	20.63
Pred_r^2	0.8415	0.9163	0.9287	0.8944	0.8021	0.8822
Pred_r^2 se	0.2544	0.2109	0.1287	0.1977	0.2290	0.1987
Best_Rand_ r^2	0.3963	0.4212	0.4258	0.3713	0.4188	0.3576
Best_Rand_ q^2	-0.0184	-0.0321	-0.001	-0.0032	-0.01744	-0.0011
Z score_ r^2	7.6394	8.3323	7.4714	6.7863	8.3864	6.5592
Z score_ q^2	6.4562	6.1276	6.3782	6.5591	6.1943	6.1133
α_Rand_ r^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
α_Rand_ q^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(+) VELY CONTRIBUTING			(-) VELY CONTRIBUTING		
DESCRIPTORS (Set-I; Test Set b_1)	XlogP (0.6710) SssSE_index (0.1672) T_N_Cl_4 (0.1439) T_O_O_2 (0.3485)			T_N_N_3 (-0.6221)		

pIC_{50} = 0.6710 XlogP − 0.6221 T_N_N_3 + 0.3485 T_O_O_2 + 0.1672 SssSE_index

+ 0.1439 T_N_Cl_4−3.3148 …**equation 4**

Table 5
Statistical parameters of developed QSAR models for descriptor set-II by forward variable selection method and MLR

Statistical Parameters	Training Set A		Training Set B		Training Set C	
	Test Set a_1	Test Set a_2	Test Set b_1	Test Set b_2	Test Set c_1	Test Set c_2
N	20	20	20	20	26	24
df (n-k-1)	14	14	14	14	20	18
r^2	0.8421	0.8421	0.9087	0.9087	0.7765	0.8256
r^2 se	0.1276	0.2145	0.0863	0.0998	0.1563	0.1762
q^2	0.7701	0.7701	0.8111	0.8111	0.6751	0.7134
q^2 se	0.2145	0.0653	0.0432	0.1782	0.2763	0.0853
F-Test	14.94	18.56	16.75	18.42	19.92	18.77
Pred_r^2	0.7950	0.8807	0.9371	0.9128	0.7279	0.7966
Pred_r^2 se	0.1139	0.2190	0.1134	0.1241	0.1124	0.2090
Best_Rand_ r^2	0.4282	0.4840	0.4473	0.4367	0.4189	0.4355
Best_Rand_ q^2	0.0954	0.1233	0.0539	0.0964	0.1762	0.0674
Z score_ $r^{26.8769}$	8.0403	6.9164	6.9244	7.5972	6.6381	5.6692
Z score_ $q^{25.4454}$	5.892	5.3771	4.7853	6.0021	5.6862	5.2289
α_Rand_ r^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
α_Rand_ q^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(+) VELY CONTRIBUTING			(-) VELY CONTRIBUTING		
DESCRIPTORS (Set-II; Test Set b_1)	Xlog (0.6785) SdssS(sulphate)count (0.3773) T_O_Cl_6 (0.1699) T_2_S_0 (0.1199)			T_N_N_3 (-0.6202)		

$pIC_{50} = 0.6785\ XlogP - 0.6202\ T\_N\_N\_3 + 0.3773\ SdssS(sulphate)count + 0.1699\ T\_O\_Cl\_6 + 0.1199\ T\_2\_S\_0 - 3.3371$ **...equation 5**

3.1.1 Accuracy of model

The value of residuals is a key factor in validating the accuracy of the model. As the value of residual is near to zero, the model is considered as more accurate since it shows minimum (≈ 0) difference in actual and predicted activity.

Residuals = Actual Biological Activity (pIC_{50}) – Predicted Activity

Table 6
Test set b_1 & b_2 and Training set A + C along with biological activity, predicted activity and residuals data (set-I)

Sr. no.	Compound	Biological Activity (pIC_{50})	Predicted Activity	Residuals
		Training Set A + C		
1	3e	-0.643	-0.731	0.088
2	5b	-0.678	-0.849	0.171
3	3q	-0.77	-0.61	-0.16
4	3d	-0.815	-0.989	0.174
5	4i	-0.892	-0.943	0.051
6	4h	-0.93	-0.989	0.059
7	3g	-1.05	-1.053	0.003
8	5d	-1.086	-1.115	0.029
9	4e	-1.098	-1.049	-0.049
10	4d	-1.148	-1.148	0.0
11	4c	-1.177	-1.079	-0.098
12	3i	-1.411	-1.419	0.008
13	4f	-1.568	-1.543	-0.025
14	2d	-1.613	-1.574	-0.039
15	3a	-1.823	-1.935	0.112
16	4g	-1.908	-1.779	-0.129
17	2b	-1.967	-1.957	-0.010
18	1	-2.253	-2.185	-0.068
19	2f	-2.321	-2.3	-0.021
20	DCA	-2.53	-2.465	-0.064
		Test Set b_1		
1	5c	-0.701	-0.845	0.144
2	3m	-1.023	-1.138	0.115
3	3j	-1.356	-1.33	-0.026
4	3l	-1.6	-1.55	-0.05
5	3c	-1.99	-1.981	-0.009
6	2c	-2.255	-2.19	-0.065
		Test Set b_2		
1	3k	-0.678	-0.801	0.123
2	3h	-1.098	-1.103	0.005
3	4a	-1.138	-1.087	-0.051
4	3f	-1.568	-1.565	-0.003
5	4b	-1.613	-1.592	-0.021
6	2e	-2.114	-2.059	-0.055

Similar calculations were carried out for Test sets a_1 & a_2 and c_1 & c_2 of Set-I as well as for Set-II. In both the sets, models b_1 & b_2 (evaluated on training set A + C) were found to be the best models.

Fig no. 8a- Test set a_1 & a_2 (Training Set B + C) Fig no. 8b- Test set c_1 & c_2 (Training Set A + B)

Fig no. 8c- Test set b_1 & b_2 (Training Set A + C) [Best model]\

Figure 8. Plot **of actual versus predicted activity (Set-I)**

3.2 Interpretation of 2D QSAR

The present QSAR models reveal that XogP as well as Baumann's alignment independent (AI) descriptors have major contribution in explaining variation in activity. Descriptors T_X_Y_Z can be defined as total count of fragments formed with atom types X and Y separated by topological distance of Z bonds [33, 34]. Interpretation of descriptors which contributed significantly for QSAR models are given below, the value given in parenthesis are percentile contribution of descriptor for the activity:

1. **XlogP [55%, 56%]**: It is the ratio of solute concentration in octanol & water and generally termed as Octanol: Water Partition Coefficient. Highest (+) ve contribution signifies that highly hydrophobic environment is necessary for increasing anticancer activity.
2. **T_N_N_3 [-21%, -19%]**: This is an AI descriptor signifying the count of no. of Nitrogen atoms (single, double or triple bonded) separated from any other -N atom (single, double or triple bonded) by 3 bond distance in a molecule. Since it is contributing negatively, it shows that 3 bond distance between any two -N atoms are detrimental for anticancer activity.
3. **SssSE_index [8%], SddssS(sulphate)count [10%] & T_2_S_0 [9%]**: First is electrotopological state indices for no. of -S atom connected with two single bonds, second descriptor defines the total no. of 'S' group connected with two single and two double bonds and third shows count of no. of double bonded –S atom present in the molecular structure. Thus, (+) ve contribution of all the descriptors ultimately indicates that the presence of –S atom in molecule is beneficial for the activity.
4. **T_N_Cl_4 [5%]**: This is the count of No. of -N atoms (single, double or triple bonded) separated from any -Cl atom by 4 bond distance in a molecule. Hence presence of -Cl at

$-R_2$ or $-R_4$ (-Meta position) on the ring is valuable for improving anticancer activity.

1. **T_O_O_2 [11%]**: Count of No. of -O atoms (single or double bonded) separated from any other -O atom by 2 bond distance in a molecule is helpful for activity. E.g., O = S = O group.
2. **T_O_Cl_6 [6%]**: It shows that 6 bond distance between -O atom (single or double bonded) and -Cl atom in the molecule is preferred for the activity.

Careful observation of descriptors in models suggests that XlogP is an indicator variable which positively contributes for QSAR Eq. (55–56%) and this signifies the importance of hydrophobicity of molecule is prerequisite for significant anticancer activity. Other descriptors like T_N_N_3 which is inversely proportional to activity shows the presence of -N on $-R_1$ (-ortho position) on ring is detrimental for biological activity. Whereas the other descriptors are contributing positively in more or less percentage revealing the importance of respective atoms/groups at different respective position on the ring for potential anticancer molecule design.

3.3 3D QSAR models

Using SA-kNN-MFA, 6 descriptors were finalized which were satisfying all statistical parameters in the generated models. After calculating residuals for each model, models b_1 & b_2 were found to be the best models.

Table 7
Comparison of the various statistical results of 3D QSAR generated by SA-kNN-MFA method

Statistical Parameters	Training Set A		Training Set B		Training Set C	
	Test Set a1	Test Set a2	Test Set b1	Test Set b2	Test Set c1	Test Set c2
N	20	20	20	20	26	24
kNN	2	2	2	2	2	2
q^2	0.7728	0.7728	0.8861	0.8861	0.6132	0.8445
q^2 se	0.2908	0.2908	0.1997	0.1997	0.0928	0.1988
Pred_r^2	0.8112	0.9163	0.8078	0.9002	0.7855	0.9190
Pred_r^2 se	0.1120	0.1935	0.2588	0.2045	0.1497	0.2040
	(+) VELY CONTRIBUTING			(-) VELY CONTRIBUTING		
DESCRIPTOS (Test Set b_1)	S_599 (0.0669, 0.2990)			E_534 (-0.3180, -0.2277)		
	H_382 (0.6625, 0.7061)					
	H_396 (0.6555, 0.8002)					
	H_459 (0.7698, 0.8690)					
	H_529 (0.6396, 0.6881)					

1. **(b) Fig. showing rotated view to see closeness of grid points to the nucleus.**
2. **(c) Magnified view of Fig no. 9a.**

3.3.1 Interpretation of 3D QSAR

3D QSAR was used to optimize the electrostatic, steric and hydrophobic requirements around N-Phenyl-2, 2-dichloroacetamide pharmacophore. The property values for the generated data points helped us for the design of potent NMEs. The ranges of data point values were based on the variation of the field values at the chosen points using the most active molecule and its nearest neighbour set. Points generated in SA-kNN-MFA 3D QSAR model are E_534 (-0.3180, -0.2277), S_599 (0.0669, 0.2990), H_382 (0.6625, 0.7061), H_396 (0.6555, 0.8002), H_459 (0.7698, 0.8690) and H_529 (0.6396, 0.6881) i.e., electronic, steric and hydrophobic data points at lattice points of 534, 599, 382, 396, 459 and 529 respectively.

- Negative value in electrostatic data points indicated the requirement of electronegative substituents (e.g. -Cl, -Br, -OH) for enhancing biological activity.
- Low range of positive steric value indicated that moderate bulky groups (i.e.-C_6H_5,

-$CH_2C_6H_5$) are required to increase activity.

- Highly positive values in hydrophobic field indicated the necessity of highly hydrophobic substituents for significant anticancer activity (-CH_3, -CH_2CH_3, -C_6H_5, -$CH_2C_6H_5$) of N-Phenyl-2, 2-dichloroacetamide derivatives.

3.4 G-QSAR models

Following the application of several QSAR approaches, 6 descriptors were finalized, and MLR models that satisfied all statistical parameters were evaluated. Models b1 and b2 were judged to be the best models in 2D and 3D QSAR after calculating residuals for each model.

Table 8
Resultant G-QSAR descriptors & models with comparison of their statistical parameters generated using forward variable selection method and MLR

Statistical Parameters	Training Set A		Training Set B		Training Set C	
	Test Set a_1	Test Set a_2	Test Set b_1	Test Set b_2	Test Set c_1	Test Set c_2
N	20	20	20	20	26	24
df (n-k-1)	14	14	14	14	20	18
r^2	0.8125	0.8125	0.8779	0.8779	0.7936	0.8066
r^2 se	0.1155	0.1752	0.0082	0.1432	0.1752	0.2002
q^2	0.7698	0.7698	0.7895	0.7895	0.6984	0.6818
q^2 se	0.1762	0.1382	0.0877	0.042	0.1832	0.1488
F- test	17.3305	17.6735	18.8969	18.6284	15.3785	16.6818
Pred_r^2	0.8689	0.9189	0.8955	0.8245	0.9154	0.8041
Pred_r^2 se	0.1364	0.1332	0.2214	0.1985	0.1432	0.2385
Best_Rand_ r^2	0.4531	0.4490	0.4878	0.4129	0.4088	0.44
Best_Rand_ q^2	0.0884	0.1121	0.0632	0.0921	0.1544	0.0664
Z score_ r^2	8.1233	7.021	7.2191	6.5411	7.8731	6.7681
Z score_ q^2	5.9124	5.3288	5.3162	6.1	5.7129	5.3731
α_Rand_ r^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
α_Rand_ q^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(+) VELY CONTRIBUTING			(-) VELY CONTRIBUTING		
CONTRIBUTING DESCRIPTORS (Test Set b_1)	R_1-Xlog$_{10}$P (0.4374) R_3-Xlog$_{10}$P (0.6773) R_4-Xlog$_{10}$P (0.3495) R_1-Hydrogens Count (0.5871)			R_1-2Path Count (-0.2262) R_1-Nitrogens Count (-0.4236)		

pIC_{50} = 0.6773 R_3-XlogP + 0.5871 R_1-Hydrogens Count + 0.4374 R_1-XlogP – 0.4236 R_1-Nitrogen Count + 0.3495 R_4-XlogP – 0.2262 R_1-2 Path Count – 0.2821 …**equation 6**

1) R_1-XlogP (+ ve)	**Hydrophobicity at -R_1 position is beneficial.**
2) R_3-XlogP (+ ve)	Highest contribution shows hydrophobicity at -R_3 position is extremely essential.
3) R_4-XlogP (+ ve)	Hydrophobicity at -R_4 position is also favourable.
4) R_1-H Count(+ ve)	No. of -H atoms at -R_1, also meaning no substituent at -R_1 is advantageous.
5) R_1-2Path Count (-ve)	No. Of total fragments of two bond path at -R_1 is detrimental for activity.
6) R_1-N Count (-ve)	-N atom at -R_1 is destructive for anticancer activity.

3.4.1 Interpretation of all QSAR model

Based on results of all QSAR models we got total of 6 descriptors signifying the requirement of particular category of substituent at specific position (site) on the pharmacophore.

3.5 Design of New Molecular Entities (NMEs) containing N-phenyl-2, 2-dichloroacetamide pharmacophore

A total of 160 NMEs were designed by applying interpreted 2D, 3D and G-QSAR results. Out of 160, only 14 NMEs exhibited higher predicted activity than the most potent compound of the original series by applying all three QSAR predictions along with Lipinski's screen score 6 and hence were selected for further in-silico studies.

Sr. No	Com-pds.	$-R_1$	$-R_2$	$-R_3$	$-R_4$	S.S.	Screen Result	Predicted Activity			
2D QSAR		3D QSAR	G QSAR								
Set-I	Set-II										
1	D1	-H	-Cl	-C ≡ CH	-Benzyl	6	ADRXWS	1.323	0.955	3.476	4.879
2	D2	-H	-C ≡ CH	$-CH_3$	$-C_6H_5$	6	ADRXWS	0.952	0.881	3.376	4.612
3	D3	-H	$-CH_3$	$-CH_3$	$-C_6H_5$	6	ADRXWS	0.713	0.501	3.031	4.349
4	D4	-H	$-CH_3$	-H	-Benzyl	6	ADRXWS	0.601	0.723	2.955	4.119
5	D5	-H	-H	$-CH_3$	$-C_6H_5$	6	ADRXWS	0.616	0.198	2.903	3.313
6	D6	-H	-C ≡ CH	-SO3H	-H	6	ADRXWS	0.526	0.091	2.811	3.016
7	D7	-H	-I	-Br	-H	6	ADRXWS	0.377	-0.001	2.791	3.003
8	D8	-H	-Cl	-Benzyl	-H	6	ADRXWS	0.290	-0.233	2.508	2.638
9	D9	-H	-H	-H	-Benzyl	6	ADRXWS	0.174	-0.200	2.440	2.524
10	D10	-H	-H	$-C_6H_5$	-Pyr	6	ADRXWS	0.117	-0.387	1.226	2.463
11	D11	-H	-Cl	$-SO_3H$	-H	6	ADRXWS	0.031	-0.455	1.195	2.188
12	D12	-H	-H	-H	-Ph	6	ADRXWS	-0.019	-0.459	0.612	1.995
13	D13	-H	-H	-I	-Br	6	ADRXWS	-0.132	-0.519	0.491	1.731
14	D14	-H	-Cl	-Br	-H	6	ADRXWS	-0.481	-0.598	0.155	0.817
S.S. = Lipinski's screen score											

Table 9 **Structures of designed NMEs along with predicted activity obtained by 2D (both descriptor sets), 3D and G-QSAR equations (arranged in descending order)**

Table 9
Structures of designed NCE's along with predicted activity obtained by 2D (both descriptor sets), 3D and G-QSAR equations (arranged in descending order)

Sr. No	Com-pds.	$-R_1$	$-R_2$	$-R_3$	$-R_4$	S.S.	Screen Result	Predicted Activity			
2D QSAR		3D QSAR	G QSAR								
Set-I	Set-II										
1	D1	-H	-Cl	-C ≡ CH	-Benzyl	6	ADRXWS	1.323	0.955	3.476	4.879
2	D2	-H	-C ≡ CH	-CH3	$-C_6H_5$	6	ADRXWS	0.952	0.881	3.376	4.612
3	D3	-H	-CH3	-CH3	$-C_6H_5$	6	ADRXWS	0.713	0.501	3.031	4.349
4	D4	-H	-CH3	-H	-Benzyl	6	ADRXWS	0.601	0.723	2.955	4.119
5	D5	-H	-H	-CH3	$-C_6H_5$	6	ADRXWS	0.616	0.198	2.903	3.313
6	D6	-H	-C ≡ CH	-SO3H	-H	6	ADRXWS	0.526	0.091	2.811	3.016
7	D7	-H	-I	-Br	-H	6	ADRXWS	0.377	-0.001	2.791	3.003
8	D8	-H	-Cl	-Benzyl	-H	6	ADRXWS	0.290	-0.233	2.508	2.638
9	D9	-H	-H	-H	-Benzyl	6	ADRXWS	0.174	-0.200	2.440	2.524
10	D10	-H	-H	$-C_6H_5$	-Pyr	6	ADRXWS	0.117	-0.387	1.226	2.463
11	D11	-H	-Cl	-SO3H	-H	6	ADRXWS	0.031	-0.455	1.195	2.188
12	D12	-H	-H	-H	-Ph	6	ADRXWS	-0.019	-0.459	0.612	1.995
13	D13	-H	-H	-I	-Br	6	ADRXWS	-0.132	-0.519	0.491	1.731
14	D14	-H	-Cl	-Br	-H	6	ADRXWS	-0.481	-0.598	0.155	0.817
S.S. = Lipinski's screen score											

3.6 Results of molecular docking studies

In molecular docking results (Maestro, Schrödinger 10.2), it was found that N-phenyl-2,2-dichloroacetamide analogues mimic DCA and binds to the DCA binding region of PDK II & PDK I active site.

3.6.1 Evaluation of molecular docking results with enzyme PDK II (PDB code: 2BU8)

3.6.1.1 G-score

The scoring function of GLIDE docking program is presented in G-score form. G-score indicates the binding affinity of the designed compound to the receptor or enzyme. G-score of standard compound DCA was found to be -5.055; whereas out of 14 designed chemical entities, only 8 showed better G-score than the standard. The G-scores of the designed NMEs D5, D12, D8, D11, D6, D4, D9 and D14 were found to be -9.921, -8.423, -7.615, -7.492, -6.893, -6.671, -6.368, -6.399 respectively. More negative is the value of G-scores, higher is the binding affinity of that compound. The close analysis of these results suggests that the designed NMEs have more binding affinity with enzyme than standard.

3.6.1.2. H-Bond interactions

H-bond is one of the most widely used parameter for the evaluation of the docking results, as it is an influential parameter in the activity of drug compound. The numbers of H-bond interactions as well as their length in standard were compared with that of the designed chemical entites. DCA itself involves two important H-bonding interactions with the key binding amino acids Arg-154 and Tyr-80 of the protein backbone as well as two other amino acid pocket region of Ile 157 and His 115 are also formed as reported in the standard ligand plot. Here also, out of 14 NMEs, only 8 compounds showed better results than standard as well the most potent compound of the series selected for QSAR. Their key interactions are –OH of compound with –NH of Arg-154 (O—H) and = O (O_2) of compound with –OH of Tyr 80 which are the same amino acids involved in standard. Compound D11 came out with an additional H-bond between –OH of SO_3H group of compound with = O of Asn-150 which shows that D11 binds with much greater affinity with protein rather than the compounds having two H-bonds. The H-bond length is also an important parameter in molecular docking studies as if it is found lesser than the H-bond length of standard, then it means that our compounds bind with higher affinity with the respective amino acid. Here we found all the 8 compounds have shorter length of both the **H-bonds** compared to standard. Compound D5 showed the highest binding affinity with the key binding amino acids in the binding pocket of PDK II.

3.6.1.3. Contacts

The contacts are represented in the form of Vender Waals (vdW) interactions.

- Good vdW interactions
- Bad vdW interactions
- Ugly vdW interactions

It was found that same 8 NMEs have more No. of good vdW interactions, less number of bad vdW and no ugly contacts when compared with DCA.

3.6.2 Evaluation of molecular docking studies results with enzyme PDK I (PDB code: 2Q8H)

3.6.2.1 G-score

G-score of DCA with PDK I was found to be -4.035, whereas above said 8 designed chemical entities showed better G-score than the standard. The G-score of the designed chemical entities D5, D12, D8, D11, D6, D4, D9 and D14 was found to be -9.881, -7.022, -6.235, -5.995, -5.983, -5.971, -5.668 and – 5.389 respectively. Thus, designed chemical entities bind with PDK I with more affinity than the standard compound.

3.6.2.2. H-Bond interactions

DCA involves two important H-bonding interactions with key binding amino acids i.e., Arg 188 and Tyr 114 along with Ile 191 and Leu 87 as a binding pocket region. All 8 designed chemical entities which showed better results with PDK II, also showed promising results with PDK I. They formed H-bonds of shorter length compared to DCA with Arg 188 (= O of compound with –NH) and Tyr 114 (-NH of compound with –OH) which is the standard binding pocket region involving Ile191 and Leu 87 amino acids in the pocket as reported in literature. Here also, Compound D11 showed an additional H-bond between –OH of SO_3H group of D11 with = O of Asn-140 showing higher affinity of D11 for PDK I compared to compounds having two H-bonds.

3.6.2.3. Contacts

All 8 designed chemical entities showed more no. of good vdW interactions, less No. of bad vdW and no ugly contacts compared to DCA.

Thus, out of 14, these 8 NMEs showed higher G-score, H-bonding with key binding amino acids having shorter length compared to DCA, higher No. of good vdW bonds and very less No. of bad or ugly vdW interactions with both the PDK enzymes respectively.

3.6.3 Results of docking scores by using VLife MDS and MVD 2007

We aimed to compare the order of docking score (G-score) of NMEs with both the PDKs with the order of docking scores obtained by using other two softwares (keeping same compounds and protein backbones). We got the similar order of docking scores i.e., the compound showed highest G-score with both the PDKs also showed highest docking score by using VLife MDS (VLife dock score) as well as by using MVD (MolDock score) for both the PDKs. Hence the obtained results were found reliable and satisfactory.

Therefore, we can conclude that 8 designed chemical entities came out with excellent results with PDK II & PDK I by using GLIDE (Maestro, Schrödinger 10.2) as well as VLife MDS 3.5 & MVD 2007 (for comparison purpose) and will bind with the same amino acids pocket region of active binding site of both the PDK receptors with much higher affinities involving more No. of H-bonds, higher no. of vdW bonds and lesser no. of bad vdW interactions. Thus, these 8 NMEs were subjected to further ADMET predictions and the other 6 NMEs were removed from in-silico studies consideration.

Sr. no.	Compd.	G-score	No. of H-bond	H-bond interactions			Good VDW	Bad VDW	Ugly VDW	VLife Dock Score	MolDock Score
		2BU8		Arg-154	Tyr-80	Asn-140					
1	**D5**	-9.921	2	1.191	1.295	-	123	0	0	-8.1472	-89.465
2	**D12**	-8.423	3	1.677	1.296	1.154	104	0	0	-7.8527	-83.825
3	**D8**	-7.615	3	1.778	2.098	1.282	89	0	0	-7.8292	-76.821
4	**D11**	-7.492	3	1.872	2.350	1.592	77	1	0	-7.3782	-73.614
5	**D6**	-6.893	2	1.875	2.412	-	73	1	0	-6.7721	-69.333
6	D4	-6.671	2	1.875	2.427	-	69	1	0	-6.3134	-62.187
7	D9	-6.368	2	1.877	2.497	-	65	1	0	-6.1144	-59.162
8	D14	-6.399	2	2.162	3.008	-	52	1	0	-6.0072	-55.392
9	5c	-5.110	2	2.821	3.010	-	39	2	1	-5.8274	-53.175
10	**Std.**	-5.055	2	3.220	3.140	-	32	4	1	-5.5645	-53.878
5c = Most potent compound of the original series											

Table 10
Results of molecular docking studies performed using extra precision mode of Glide (Maestro, Schrödinger) (arranged in descending order) with PDK II (PDB Code: 2BU8)

Sr. no.	Compd.	G-score	No. of H-bond	H-bond interactions			Good VDW	Bad VDW	Ugly VDW	VLife Dock Score	MolDock Score
		2Q8H		Arg-188	Tyr-114	Asn-150					
1	**D5**	-9.881	2	1.591	2.255	-	105	0	0	-5.9176	-69.875
2	**D12**	-7.022	3	1.679	2.286	1.169	99	0	0	-5.8527	-67.465
3	**D8**	-6.235	3	1.778	2.292	1.432	82	0	0	-5.8282	-66.901
4	D11	-5.995	3	1.872	2.350	1.592	77	1	0	-5.7982	-63.714
5	D6	-5.983	2	1.873	2.412	-	64	1	0	-5.7721	-59.765
6	D4	-5.971	2	1.875	2.427	-	66	1	0	-5.6134	-57.449
7	D9	-5.668	2	1.877	2.497	-	60	1	0	-5.5544	-50.332
8	D14	-5.389	2	2.162	3.008	-	52	1	0	-5.1092	-48.675
9	5c	-4.211	2	2.821	3.010	-	33	2	1	-3.9834	-44.987
10	**Std.**	-4.035	2	2.80	3.150	-	24	4	1	-4.7945	-47.578
5c = Most potent compound of the original series											

Table 11 **Results of molecular docking studies performed using extra precision mode of Glide (Maestro, Schrödinger) (arranged in descending order) with PDK I (PDB Code: 2Q8H)**

Table 11
Results of molecular docking studies performed using extra precision mode of Glide (Maestro, Schrödinger) (arranged in descending order) with PDK I (PDB Code: 2Q8H)

Sr. no.	Com-pd.	G-score	No. of H-bond	H-bond interactions			Good VDW	Bad VDW	Ugly VDW	VLife Dock Score	MolDock Score
		2Q8H		Arg-188	Tyr-114	Asn-150					
1	D5	-9.881	2	1.591	2.255	-	105	0	0	-5.9176	-69.875
2	D12	-7.022	3	1.679	2.286	1.169	99	0	0	-5.8527	-67.465
3	D8	-6.235	3	1.778	2.292	1.432	82	0	0	-5.8282	-66.901
4	D11	-5.995	3	1.872	2.350	1.592	77	1	0	-5.7982	-63.714
5	D6	-5.983	2	1.873	2.412	-	64	1	0	-5.7721	-59.765
6	D4	-5.971	2	1.875	2.427	-	66	1	0	-5.6134	-57.449
7	D9	-5.668	2	1.877	2.497	-	60	1	0	-5.5544	-50.332
8	D14	-5.389	2	2.162	3.008	-	52	1	0	-5.1092	-48.675
9	5c	-4.211	2	2.821	3.010	-	33	2	1	-3.9834	-44.987
10	Std.	-4.035	2	2.80	3.150	-	24	4	1	-4.7945	-47.578
5c = Most potent compound of the original series											

The results of this virtual screening did support the postulation that the active NMEs act on the same enzyme target where PDK inhibitor DCA acts.

3.7 ADMET predictions

NMEs were generated using CombiLib were analyzed by Lipinski's rule to ensure their drug-like pharmacokinetic profile while designing new molecules. In addition to ensure drug like pharmacokinetic properties their other ADMET properties were also predicted and compared with their ideal ranges for a chemical entity to act as a drug. Out of 8 NMEs, 7 designed entities showed satisfactory results within the ideal ranges whereas only D6 compound didn't fit in the ideal ranges of various pharmacokinetic properties to act as a drug and hence was removed from the final consideration of new molecules to be subjected to experimental validation.

3.7.1 ADMET prediction by QikProp, Schrödinger

Numbers of properties of designed analogues were predicted by QikProp tool, Schrödinger 9.0 which was used as last screening tool to select the final NMEs. Here we have reported only descriptors which contributed significantly for predicting drug like properties of the molecule. These properties are as follow: (figures in parenthesis indicate ideal values in order the test compounds to have drug like pharmacokinetic properties)

1. Lipinski's Rule: Compounds that satisfy this rule were expected to have drug like pharmacokinetic profile.
2. Brain/blood partition coefficient (CNS) (-2 to 2)
3. Percent Human Oral absorption (> 80% is high, < 25% is poor)
4. Number of possible metabolites (should range from 1–8)

Ideal range of Lipinski's rule properties have been satisfied by the finally considered 7 NMEs. Hence, it concludes that these NMEs can act as drug and have drug like bioavailability.

CNS parameter is related with absorption of entity through Blood brain barrier; standard limit for CNS is -2 to + 2, where – 2 show inactive CNS penetration and + 2 shows active CNS penetration. Since neurotoxicity has been reported as the major toxic effect of DCA in literature, all the designed compounds must have CNS parameter value as -2 (inactive BBB penetration), which has been satisfied by all 7 finally considered new molecules. Therefore, all are considered as nontoxic compounds.

Percent human oral absorption parameter is related with extent of oral absorption of drug, indicating suitable route of administration and exhibit the extent of oral bioavailability profile. Out of 7, 6 NMEs showed 100% human oral absorption whereas D14 showed 95%, hence we can say that all finally considered designed NMEs can be orally absorbed and will exhibit high bioavailability profile.

All compounds showed No. of metabolites in ideal range as well as relatively more No. of metabolites as compared to standard so that we can say that they can be metabolized and excreted easily (mostly via urine) and will not cause any side effect or will not produce any toxic metabolite.

So, after detailed analysis of ADMET predictions, it will worth saying that all the 7 selected compounds can exhibit drug like pharmacokinetic profile.

3.7.2 ADMET prediction by Discovery Studio, Accelrys

DS defined the prediction level for all ADMET properties. Mainly four properties have been considered:

1. ADMET absorption in which there are four prediction level: 0-Good, 1-Moderate, 2-Poor and 3-Very Poor
2. ADMET Plasma Protein Binding (PPB) level: 0-binding is < 90%, 1-binding is > 90% and 2-binding is > 95%
3. ADMET Cytochrome P4502D6 (CYP2D6) Probability level: < 0.5-Unlikely to inhibit CYP2D6 enzyme (Non-inhibitors of CYP2D6) and > 0.5-Likely to inhibit CYP2D6 enzyme (Inhibitors of CYP2D6)
4. ADMET probability level: < 0.5-Unlikely to cause dose-dependent liver injuries (Nontoxic) and > 0.5- Likely to cause dose-dependent liver injuries (Toxic).
5. All 7 designed compounds showed good absorption, PPB level and found non inhibitors of CYP2D6 enzyme. So, these results are better than standard. From the results of ADMET hepatotoxicity probability, they were found to be nontoxic to liver whereas standard showed probability of hepatotoxicity. (Table 11)

Sr. no.	Comp.	Mol. Wt. (g/mol)	DonorHB	AccptHB	QPlog Po/w	% Human Oral Abs	CNS	No. of possible metabolites	ADMET absorption level	ADMET PPB level	ADMET CYP2D6 probability	ADMET Hepatotoxicity probability
1	D4	308.202	3	4	2.631	100	-2	5	0	2	0.297	0.072
2	D5	294.175	2	4	3.145	100	-2	5	0	2	0.366	0.251
3	D8	328.620	3	3	2.684	100	-2	4	0	2	0.405	0.317
4	D9	294.175	2	4	3.492	100	-2	4	0	2	0.455	0.357
5	D11	318.561	2	5	0.144	100	-2	4	0	2	0.415	0.360
6	D12	280.149	3	2	2.597	100	-2	4	0	2	0.455	0.455
7	D14	317.389	2	2	0.077	95	-2	3	0	1	0.485	0.43
8	5c	272.938	2	2	2.812	95	-1	2	0	1	0.468	0.413
9	Std	128.931	1	2	1.594	86.88	2	2	1	0	0.514	0.529
5c = Most potent compound of the original series												

Table 12 Results of ADMET properties

From the results of molecular modeling studies, seven compounds (D4, D5, D8, D9, D11, D12 and D14) have been selected for further experimental validation.

4. Conclusion

Table 13
Outcome of the research work

OBJECTIVES SET FORTH	RESULTS OF IN-SILICO STUDIES				
	Data	Selected series compounds		Designed NME's	
Design of potent NMEs (QSAR studies)	Most potent pIC_{50} (μM)	-0.64		0.9–4.8	
	Least potent pIC_{50} (μM)	-3.0048		(-0.5)-0.8	
Design of selective PDK inhibitors (Molecular Docking studies)	Enzyme Isoforms	PDK II (2BU8)	PDK I (2Q8H)	PDK II (2BU8)	PDK I (2Q8H)
	Highest G-score	-5.11	-4.211	-9.921	-9.881
	Lowest G-score	-2.49	-2.12	-6.39	-5.389
Final selection of drug like molecules (ADMET Predictions)	Lipinski's screen score	5		6	
	% Human Oral Absorption	95%		100%	
	CNS Parameter	-1		-2	
	No. of possible metabolites	2		5	

The results of 2D, 3D and G-QSAR studies using different interpretation approaches have yielded detailed insights of proper working and ways of thinking in this research area. It will also motivate new users to use molecular modeling viz. 2D, 3D & G-QSAR studies, molecular docking studies as a

tool ADMET prediction for design of selective, potent new molecular entities which can suffice the ADMET properties essential for drug like pharmacokinetic profile.

The results of activity prediction show significant improvement of activity. Ultimately QSAR studies are useful in understanding the structural requirement to design novel potent molecules. The binding affinity for design NMEs were checked by subjecting them to docking studies in active binding site of protein receptor. A screening approach has thus facilitated the identification of suitable compounds from designed library for anticancer activity.

5. Experimental Protocols

5.1 Chemistry

All the reactions were carried out with dry, freshly distilled solvents under anhydrous conditions, unless otherwise noted. Melting points were determined by Veego VMP-D Digital melting point apparatus and were uncorrected. FTIR spectra of the powdered compounds were recorded using KBr on a Varian-160 FTIR spectrophotometer using Diffuse Reflectance Attachment and are reported in cm^{-1}. ^1H NMR spectra were recorded on a Varian Mercury YH300 (300 MHz FT NMR) spectrophotometer using TMS as an internal standard (Chemical shift represented in (delta ppm). Purity of the compounds was checked on TLC plates using silica gel G as stationary phase and iodine vapours as visualizing agent.

5.1.1 Synthesis of N-phenyl-2, 2-dichloroacetamide derivatives

A mixture of substituted aniline (3 mmol) and dichloroacetyl chloride (3.9 mmol, 1.3 equivalent) in dry toluene (20 ml) was stirred under reflux in a round-bottom flask. The progress of the reaction was monitored by TLC (petroleum ether/ethyl acetate 4:1). After 1–5 h, the solvent and excess dichloroacetyl chloride was allowed to evaporate under vacuum. The residual powder was N-phenyl-2, 2-dichloroacetamide derivative, with an 80–90% yield.

5.1.1.1 N- (4-benzyl-3-methylphenyl)-2, 2-dichloroacetamide (D4)

Percent Yield: 71.23% (Solid). M.P.140ºC (uncorrected). FTIR (KBr): 3235(Aromatic C-H Stretch); 1426(Aromatic C-O-C stretch); 1225(C-N Stretch); ^1H NMR (300 MHz, $CDCl_3$, δ ppm): 2.1(s, 3H, -CH); 3.8 (S, 6H, $-OCH_3$); 7.2 pyrazole -CH (s, H); 7.7 (d, 2H, -CH); 7.8(d, 2H, -CH); 7.6(d, 3H, -CH).; ESIMS m/z 340.16 [M + 1].

5.1.1.2 N- (3-methyl-4-biphenyl)-2, 2-dichloroacetamide (D5)

Percent Yield: 61.53% (Solid). M.P.128ºC (uncorrected). FTIR (KBr): 2927(Aromatic C-H Stretch); 1507(Aromatic C-O-C stretch); 1596(Aromatic C = C Stretch); 1295(C-N Stretch);812(C-Cl); ^{1} H NMR (300 MHz, $CDCl_3$, δ ppm): 2.2(s, 3H, -CH); 3.8 (S, 3H, $-OCH_3$); 7.2 pyrazole -CH (s, H); 7.5 (d, 2H, -CH); 7.7(d, 2H, -CH); 7.8(d, 2H, -CH); ESIMS m/z 394.16 [M + 1]

5.1.1.3 (N-(3-benzyl-5-chlorophenyl)-2, 2-dichloroacetamide)

Percent Yield: 71% (Solid). M.P. 136ºC (uncorrected). FTIR (KBr): 2931(Aromatic C-H Stretch); 1033 (Aromatic C-O-C stretch); 1600(Aromatic C = C Stretch); 1245(C-N Stretch);

1646(C = O Stretch); ^{1} H NMR (300 MHz, $CDCl_3$, δ ppm): 2.3(s, 3H, -CH); 3.8(S, 3H, $-OCH_3$); 7.5 pyrazole -CH (s, H); 7.7 (d, 2H, -CH); 7.9(d, 2H, -CH); 8.1 (d, 2H, -CH); ESIMS m/z 369.42 [M + 1]

5.1.1.4 (N-(4-benzylphenyl)-2, 2-dichloroacetamide) (D9)

Percent Yield: 66.47% (Solid). M.P.152ºC (uncorrected). FTIR (KBr): 2962(Aromatic C-H Stretch); 1110(Aromatic C-O-C stretch); 1551(Aromatic C = C Stretch);1285 (C-N Stretch); 1685 (C = O Stretch); ^{1} H NMR (300 MHz, $CDCl_3$, δ ppm): 2.4(s, 3H, -CH); 3.8 (S, 6H, $-OCH_3$); 6.4(s,2H, -CH)7.5 pyrazole -CH (s, H); 7.7 (d, 2H, -CH); 7.9(d, 2H, -CH); 8.1 (d, 2H, -CH); ESIMS m/z 399.16 [M + 1]

5.1.1.5 3- (2, 2-dichloroacetamido)-5-chloro benzene sulfonic acid (D11)

Percent Yield: 76% (Solid). M.P. 142ºC (uncorrected). FTIR (KBr): 2927(Aromatic C-H Stretch); 1025(Aromatic C-O-C stretch); 1611(Aromatic C = C Stretch); 1253(C-N Stretch); 1654(C = O Stretch); ^{1} H NMR (300 MHz, $CDCl_3$, δ ppm): 3.8 (S, 3H, $-OCH_3$); 6.4(s,2H, -CH);7.5 pyrazole -CH (s, H); 7.7 (d, 2H, -CH); 7.9(d, 2H, -CH); 8.1 (d, 2H, -CH); ESIMS m/z 423.28 [M + 1]

5.1.1.6 N- (4-biphenyl)-2, 2-dichloroacetamide (D12)

Percent Yield: 73% (Solid). M.P. 130–132C (uncorrected). FTIR (KBr): 3434(Amide N-H Stretch); 3069(Aromatic C-H Stretch); 1672(Amide C = O Stretch); 1282(C-N stretch); 846(Aromatic C-H bend); 742(C-Cl stretch)

5.1.1.7 N-(3-bromo-5-chlorophenyl)-2,2-dichloroacetamide (D14)

Percent Yield: 78% (Solid). M.P. 156–158C (uncorrected). FTIR (KBr): 3480(Amide N-H Stretch); 3094(Aromatic C-H Stretch); 1633(Amide C = O Stretch); 1262(C-N stretch); 625(C-Br stretch); 720(C-Cl stretch)

5.2 Anticancer screening using Microculture Tetrazolium Test (MTT) Assay: -

For screening anticancer screening experiment, the cells from a particular cell line when in log phase of growth were trypsinized, counted in a hemocytometer and adjusted to appropriate density in a suitable medium. The cells were seeded into 96-well plates in 100µl of respective medium containing 5% FBS, at plating density of 10,000 cells/well and incubated at 37ºC, the conditions maintained include 5% CO_2, 95% air and 100% relative humidity was maintained for 24 h prior to addition of synthesized compounds. The synthesized compounds were solubilized in dimethylsulfoxide and diluted in serum free medium. After 24 h, 100 µl of the medium containing the synthesized compounds at various concentration (e.g.; 6.25, 12.5, 25, 50 µM etc.) was added and incubated at 37ºC, 5% CO_2, 95% air and 100% relative humidity for 48h. Triplicate was maintained and the medium containing without synthesized compounds were served as control.

After 48h, 15µl of MTT (5mg/ml) in phosphate buffered saline (PBS) was added to each well and incubated at 37ºC for 4h. The medium with MTT was then flicked off and the formed dark blue colored formazan crystals were solubilized in 100µl of DMSO and then measured the absorbance at 570 nm using micro plate reader. The percent cell inhibition was determined using

the following formula and graph was plotted between % Cell inhibition and concentration and from this IC_{50} was calculated using Graph Pad Prism software.

The percentage growth inhibition was calculated using following formula,

% Cell Inhibition = Absorbance of compounds / Absorbance of control…equation 7

Abs (drug) = Absorbance value of drug

Abs (control) = Absorbance value of control

Table 14
Data of antiproliferative activity of all seven synthesized compounds on A549 cancer cell line.1

Compound	Conc. (μM)	Absorbance	Percent Growth Inhibition	IC_{50} (μM)	pIC_{50}	Predicted Activity (pIC_{50})			
						2D Set-I	2D Set-I	3D QSAR	G-QSAR
D4	0.0625	0.3059	31.69	0.167	0.775	0.469	0.453	0.505	0.756
	0.125	0.2641	41.03						
	0.25	0.1810	59.58						
	0.5	0.1297	71.04						
	1	0.055	87.65						
D5	0.0625	0.3202	28.51	0.191	0.727	0.446	0.408	0.493	0.701
	0.125	0.2691	39.92						
	0.25	0.1911	57.34						
	0.5	0.1393	68.89						
	1	0.0666	85.11						
D8	0.0625	0.3269	27.01	0.218	0.660	0.309	0.302	0.343	0.652
	0.125	0.2842	36.53						
	0.25	0.2002	55.30						
	0.5	0.1712	61.77						
	1	0.0779	82.60						
D9	0.0625	0.3333	25.58	0.238	0.624	0.221	0.217	0.320	0.601
	0.125	0.2912	34.98						
	0.25	0.2091	53.32						
	0.5	0.1745	61.05						
	1	0.0891	80.01						
D11	0.0625	0.3462	22.69	0.290	0.537	0.153	0.145	0.295	0.511
	0.125	0.2998	33.06						
	0.25	0.2200	50.88						
	0.5	0.1997	55.41						
	1	0.1145	74.43						
D12	0.0625	0.3597	19.75	0.409	0.388	0.131	0.109	0.246	0.353
	0.125	0.3189	28.78						
	0.25	0.2588	42.21						
	0.5	0.2139	52.23						
	1	0.1562	65.12						
D14	0.0625	0.3788	15.41	0.635	0.197	-0.23	-0.59	0.125	0.168
	0.125	0.3418	23.69						
	0.25	0.2906	35.13						
	0.5	0.2523	43.66						
	1	0.1816	59.46						

Declarations

Acknowledgement:

The authors are thankful to Dr. Ashwini R. Madgulkar, Principal, AISSMS College of Pharmacy, for continuous motivation, support and providing necessary infrastructure to carry out this work. Acknowledgement for Research Project Funding-Authors acknowledge and are thankful to ALL INDIA COUNCIL FOR TECHNICAL EDUCATION NEW DELHI FOR Research project funding in the form Research Project Scheme (RPS grant) awarded to Dr. Shashikant V. Bhandari (The RPS Project Sanction Letter Ref. File No. 8-RPS (Policy-1/ I2019-20 Dated 14th Aug. 2020.)

References

1. Whitehouse S, Cooper RH (1974) Mechanism of activation of pyruvate dehydrogenase by dichloroacetate and other halogenated carboxylic acids. Biochem J 141:761–774
2. Stacpoole PW (1989) The pharmacology of dichloroacetate. Metabolism 38:1124–1144
3. Kato M, Li J, .Chuang JL, Chuang DT (2007) Distinct structural mechanisms for inhibition of pyruvate dehydrogenase kinase isoforms by AZD7545, dichloroacetate, and radicicol. Structure 15:992–1004
4. Aicher TD, Anderson RC, Gao J, Shetty SS, Coppola GM (2000) Secondary amides of (R)-3,3,3-trifluoro-2-hydroxy-2-methylpropionic acid as inhibitors of pyruvate dehydrogenase kinase. J Med Chem 43:236–249
5. K.M.Popov, Regulation of mammalian pyruvate dehydrogenase kinase,FEBS Lett.419197–200
6. Aicher TD, Anderson RC, Bebernitz GR, Coppola GM (1999) (R)-3, 3,3-trifluoro-2-hydroxy-2-methylpropionamides are orally active inhibitors of pyruvate dehydrogenase kinase. J Med Chem 42:2741–2746
7. Bebernitz GR, Aicher TD, Stanton JL, Gao J, Shetty SS (2000) Anilides of (R)-trifluoro-2-hydroxy-2-methylpropionic acid as inhibitors of pyruvate dehydrogenase kinase. J Med Chem 43:2248–2257
8. Whitehouse S (1973) Activation of pyruvate dehydrogenase in perfused rat heart by dichloroacetate. Biochem J 134:651–653
9. Henderson GN, Curry SH, Derendrof H, Wright EC (1997) Pharmacokinetics of dichloroacetate in adult patients with lactic acidosis. J Clin Pharmacol 37:416–425
10. Stacpoole PW, Kerr DS, Barnes C, Bunch ST (2006) .Carney, Controlled clinical trial of dichloroacetate for treatment of congenital lactic acidosis in children. Pediatrics 117:1519–1531
11. Stacpoole PW, Nagaraja NV, Huston AD (2003) Efficacy of dichloroacetate as a lactate lowering drug. J Clin Pharmacol 43:683–691
12. Stacpoole PW, Henderson GN, Cornett ZYanR (1998) .James, Pharmacokinetics, metabolism and toxicology of dichloroacetate. Metab Rev 30:499–539
13. Stacpoole PW, Gilbert LR, Neiberger RE, Carney PR, Valenstein E, .Theriaque DDW (2008) .Shuster, Evaluation of long-term treatment of children with congenital lactic acidosis with dichloroacetate. Pediatrics 121:1223–1228
14. Blackshear PJ, Holloway PA (1974) .Alberti, The metabolic effects of sodium dichloroacetate in the starved rat. Biochem J 142:279–286
15. Stacpoole PW, Moore GW, Kornhauser DM (1978) Metabolic effects of dichloroacetate in patients with diabetes mellitus and hyperlipoproteinemia. N Engl J Med 298:526–530
16. Wells PG, Moore GW, Rabin D, Wilkinson GR, .Oates JA (1980) .Stacpoole, Metabolic effects and pharamcokinetics of intravenously administered dichloroaceate in humans. Diabetologia 19:109–113
17. Bersin RM, Stacpoole PW (1997) Dichloroacetae as metabolic therapy for myocardial ischemia and failure. Am HeartJ 134:841–855
18. Preiser JC, Moulart D (1990) .Vincent, Dichloroacetatae administration in the treatment of endotoxin shock. Circ Shock 30:221–228
19. Johnson GA (1977) .Alberti, The metabolic effects of sodium dichloroacetate in experimental hepatitis in rat. Biochem Soc Trans 5:1387–1388
20. Bonnet S, Archer SL, -Turner JAllalunis, Harmony A, Michelakis ED (2007) A mitochondria-K^+ channel axis is suppressed in cancer and its normalization promotes apoptosis and inhibits cancer growth. Cancer Cell 11:37–51
21. Michelakis ED, Webster L (2008) .Mercky, Dichloroacetate (DCA) as a potential metabolic targeting therapy for cancer. Br J Cancer 99:989–994
22. Y.Wong JY, Huggins GS, Debidda M, .Munshi NC, Vivo ID (2008) Dichloroacetate induces apoptosis in endometrial cancer cells. Gynecol Oncol 109:394–402
23. Sun RC, Fadia M, Dahlstrom JE, Parish CR, Board PG, Blackburn AC (2010) Reversal of the glycolytic phenotype by dichloroacetate inhibits metastatic breast cancer cell growth in vitro and in vivo. Breast Cancer Res Treat 120:253–260
24. W.Cao S, Yacoub KT, Shiverick K, Namiki Y, Sakai S, Porvasnik C, Urbanek CJ, Rosser (2008) Dichloroacetate (DCA) sensitizes both wild-type and over expressing Bcl-2 prostate cancer cells in vitro to radiation. Prostate 68:1223–1231
25. Yang Y, Cheng PShangC, Wang D, Yang P, Zhang F, Li T, Lu A, Zhao Y (2010) Novel N-phenyl dichloroacetamide derivatves as anticancer reagents: Design, synthesis and biological evaluation. Eur J Med Chem 45:4300–4306

26. Dudek AZ (2006) Computational methods in developing quantitative structure-activity relationships (QSAR): A review. Comb Chem High T Scr 9:213–228
27. Golbraikh A (2002) Beware of q^2! J Mol Graphics Model 20:269–276
28. L.Eriksson J, Jaworska AP, Dowell PGramatica (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environ Health Perspect 111:1361–1375
29. S.R.Johnson (2008) The trouble with QSAR (or how I learned to stop worrying and embrace fallacy. J Chem Inf Model 48:25–26
30. A.Tropsha, Best practices for QSAR model development,validation and exploitation,Mol.Inf.29478–488
31. VLife M Design Suite; version 3.5 & 4.0 (2012) VLife Sciences Technologies Pvt.Ltd., Pune, India. (www.vlifesciences.com)
32. Halgren TA (1996) Merck molecular force field. II. MMFF94 Vander Waals and electrostatic parameters for intermolecular interactions. J Comp Chem 17:520–552
33. K.Baumann, An alignment-independent versatile structure descriptor for QSAR and QSPR based on the distribution of molecular features,J.Chem.Inf.Comput.Sci.4226–32
34. Hall LH, Kier LB (1995) Electrotopological state indices for atom types: A novel combination of electronic, topological and valence state information. J Chem Inf Comput Sci 35:1039–1045
35. C.Rucker G, Rucker MMeringer (2007) y-Randomization and its variants in QSPR/QSAR. J Chem Inf Model 47:2345–2357
36. Zheng W (2000) Novel variable selection quantitative structure-property relationship approach based on the k-nearest-neighbor principle. J Chem Inf Comput Sci 40:185–194
37. Shen M, Xiao Y, Golbraikh A, Gombar VK, Tropsha (2003) A Development and validation of k-nearest-neighbor, QSPR models of metabolic stability of drug candidates. J Med Chem 46:3013–3020
38. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development setting. Adv Drug Deliv Rev 46:3–26
39. Topliss TG (1979) Chance factors in studies of quantitative structure-activity relationships. J Med Chem 22:1238–1244
40. Ajmani S, Jadhav K, Kulkarni SA (2006) Three-dimensional QSAR using the k-nearest neighbour method and its interpretation. J Chem Inf Model 46:24–31
41. I.Papandreov T (2011) Anticancer drugs that target metabolism: Is dichloroacetate the new paradign? Int J Cancer 128:1001–1008
42. www.rscb.org
43. Glide Molecular docking tool of Schrödinger Inc., version 10.2,LLC, New York, USA
44. QikProp Schrödinger Inc., version 9.0,LLC, New York, USA
45. Discovery, Studio version 2.1, Accelrys Inc.: San Diego, CA, USA
46. Cancer (who.int)

Figures

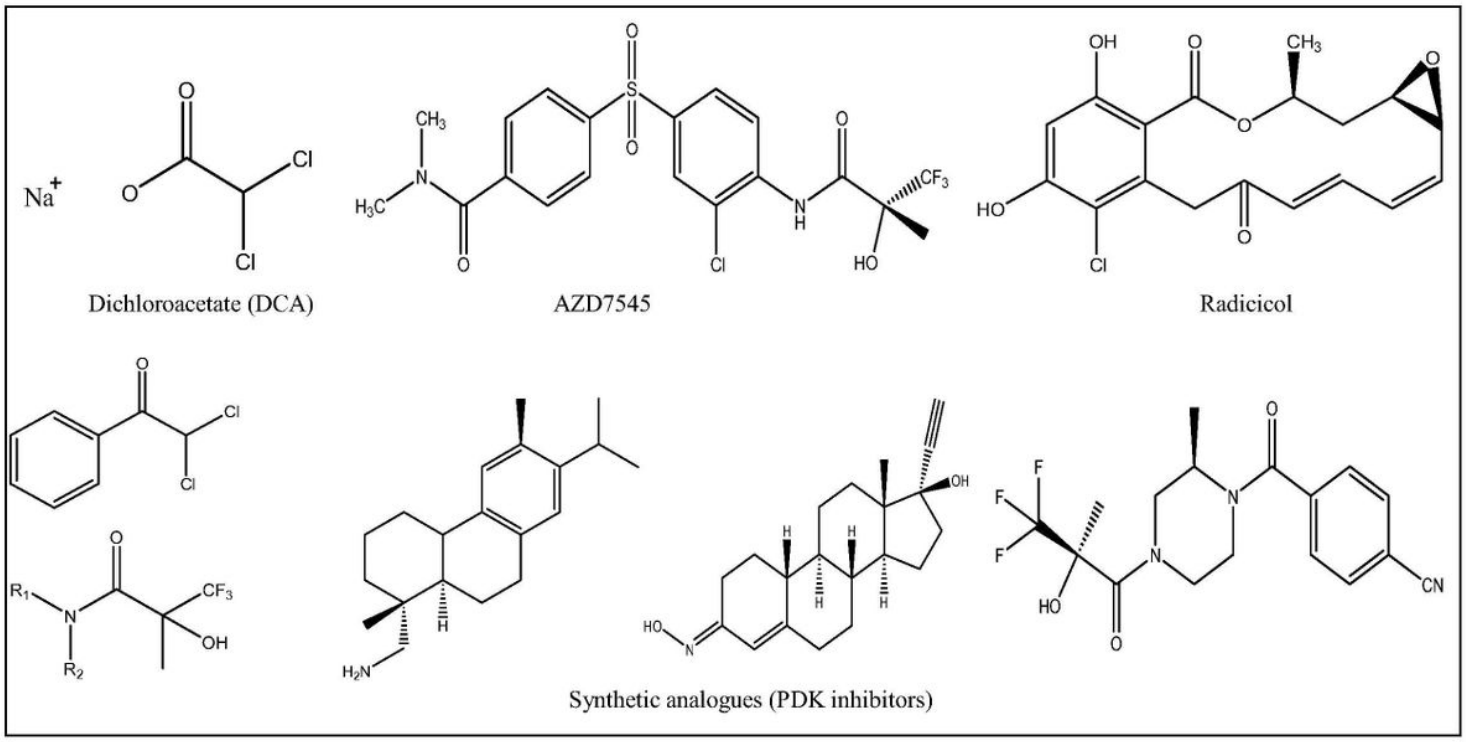


Figure 1

Several reported orally effective PDK inhibitors

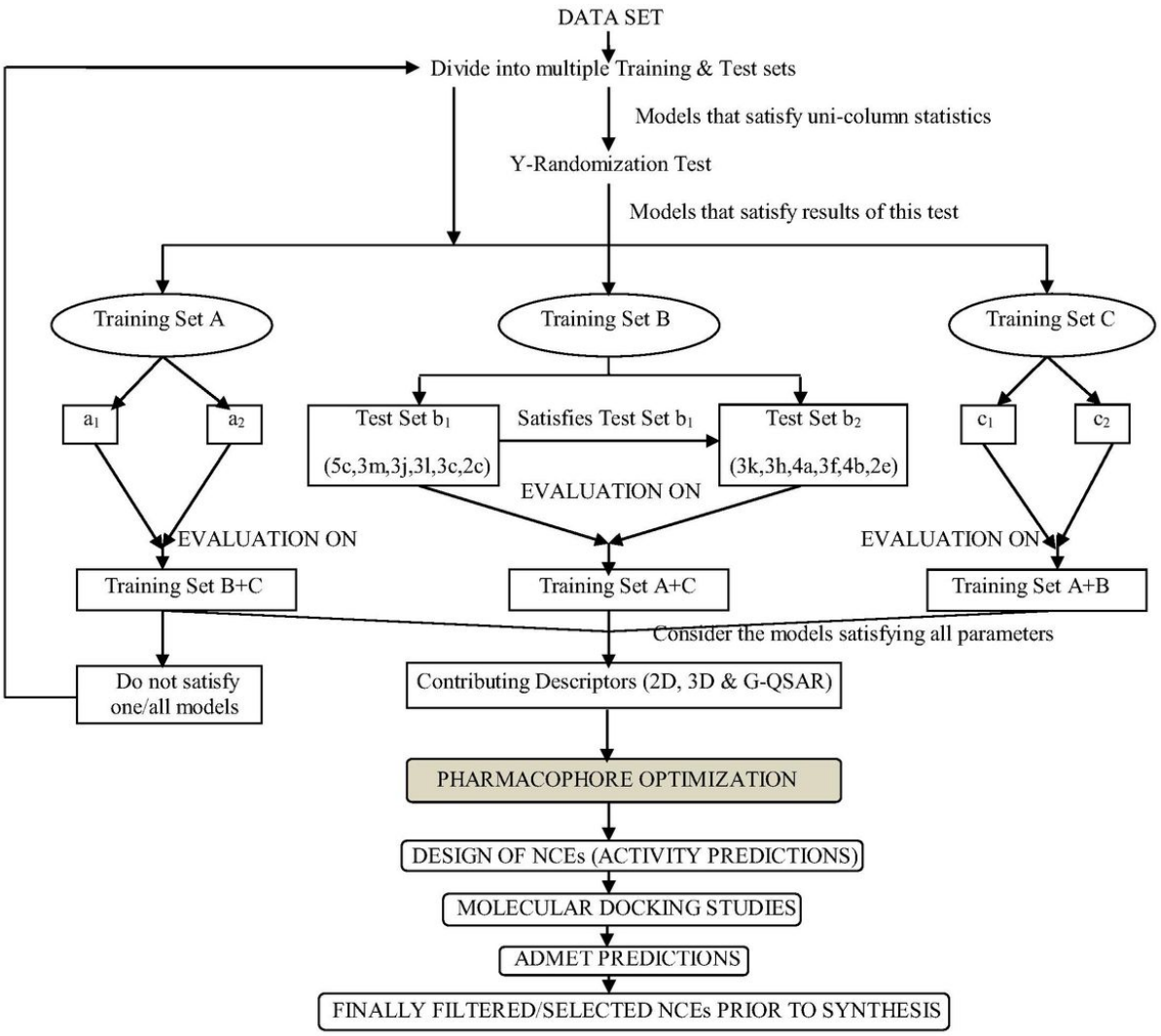


Figure 2

Experimental design

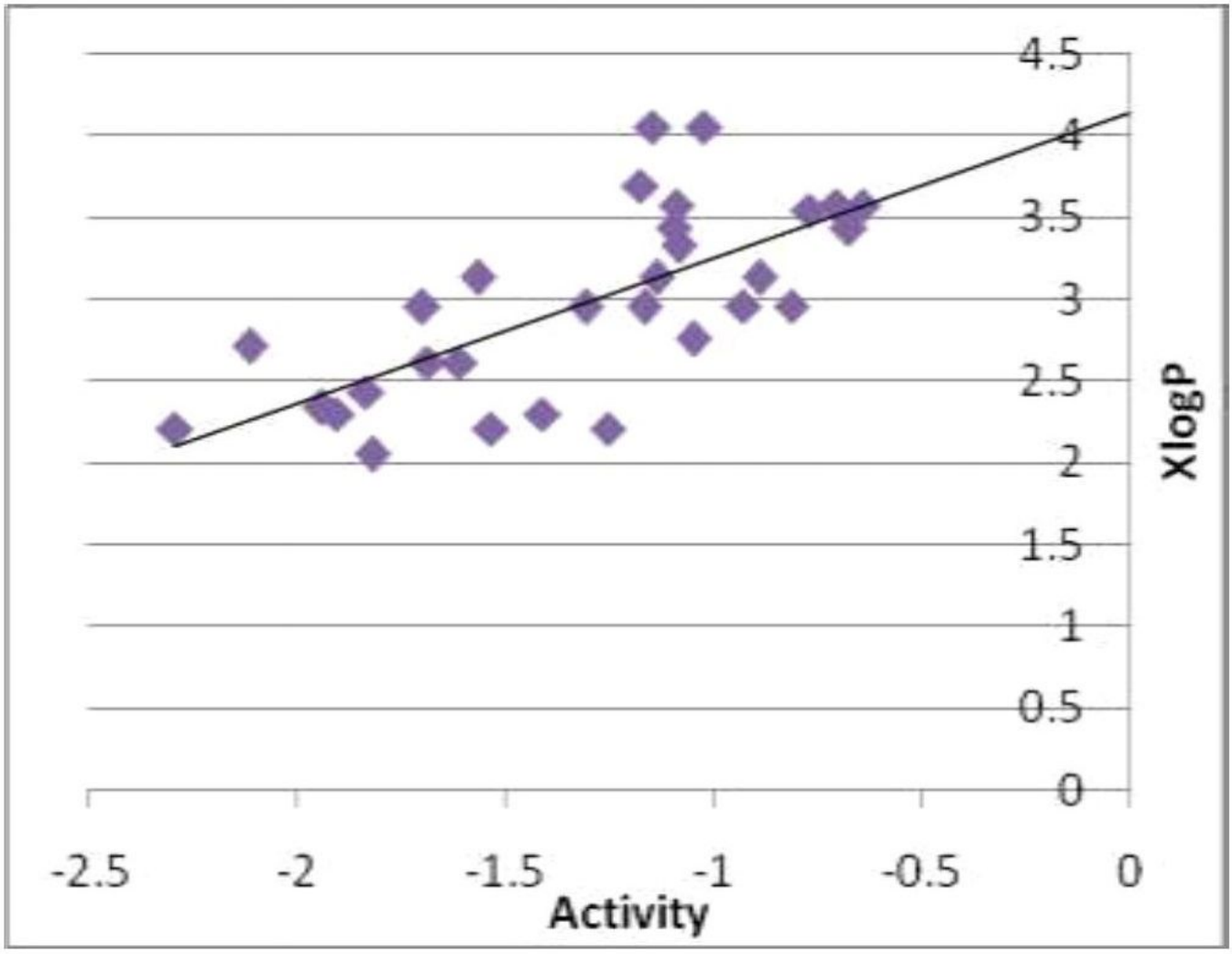


Figure 3

Fitness plot for descriptor $Xlog_{10}P$ (highest contributing descriptor)

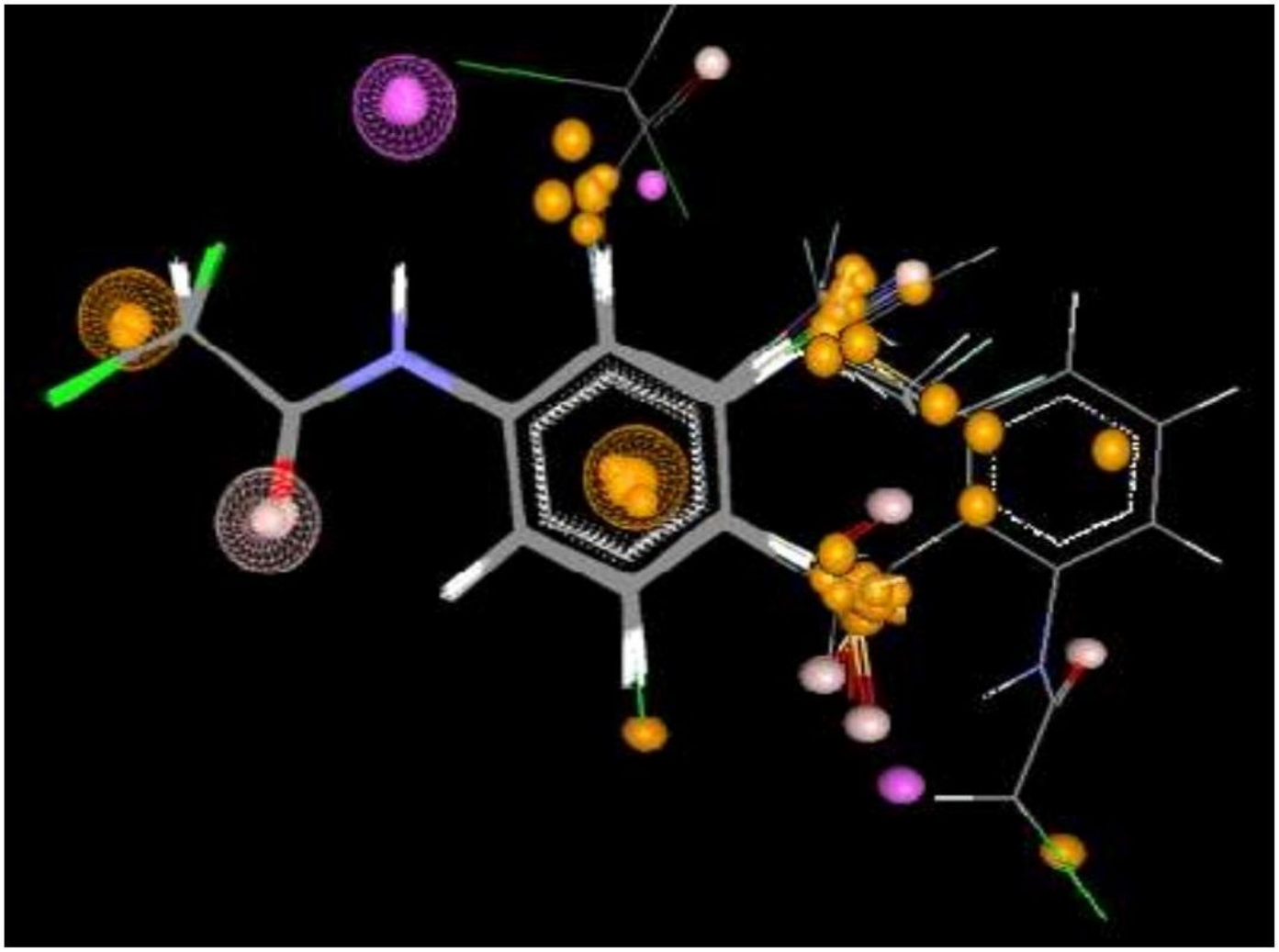


Figure 4

Identified pharmacophore features and alignment of molecules by MolSign

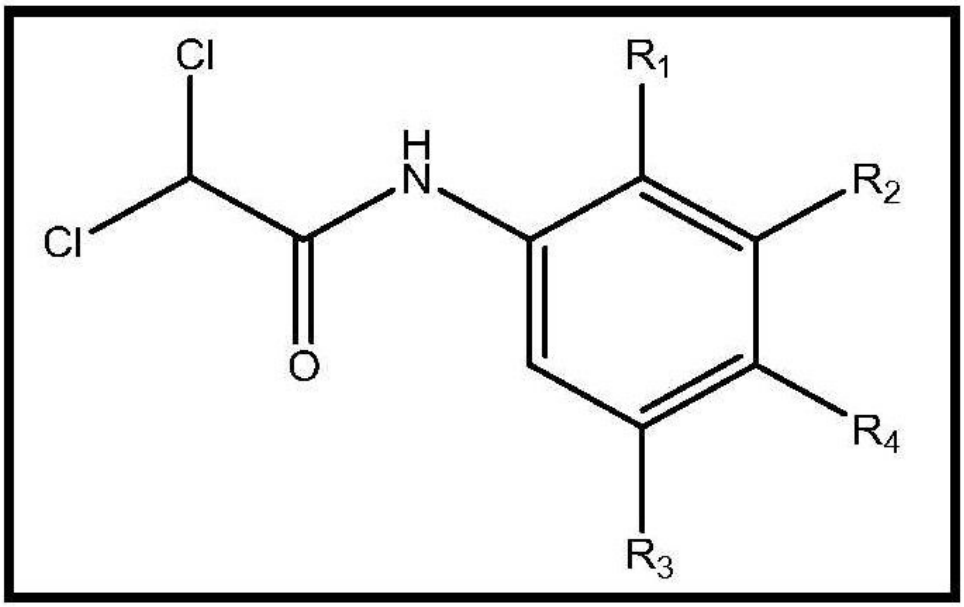


Figure 5

Template used for G-QSAR

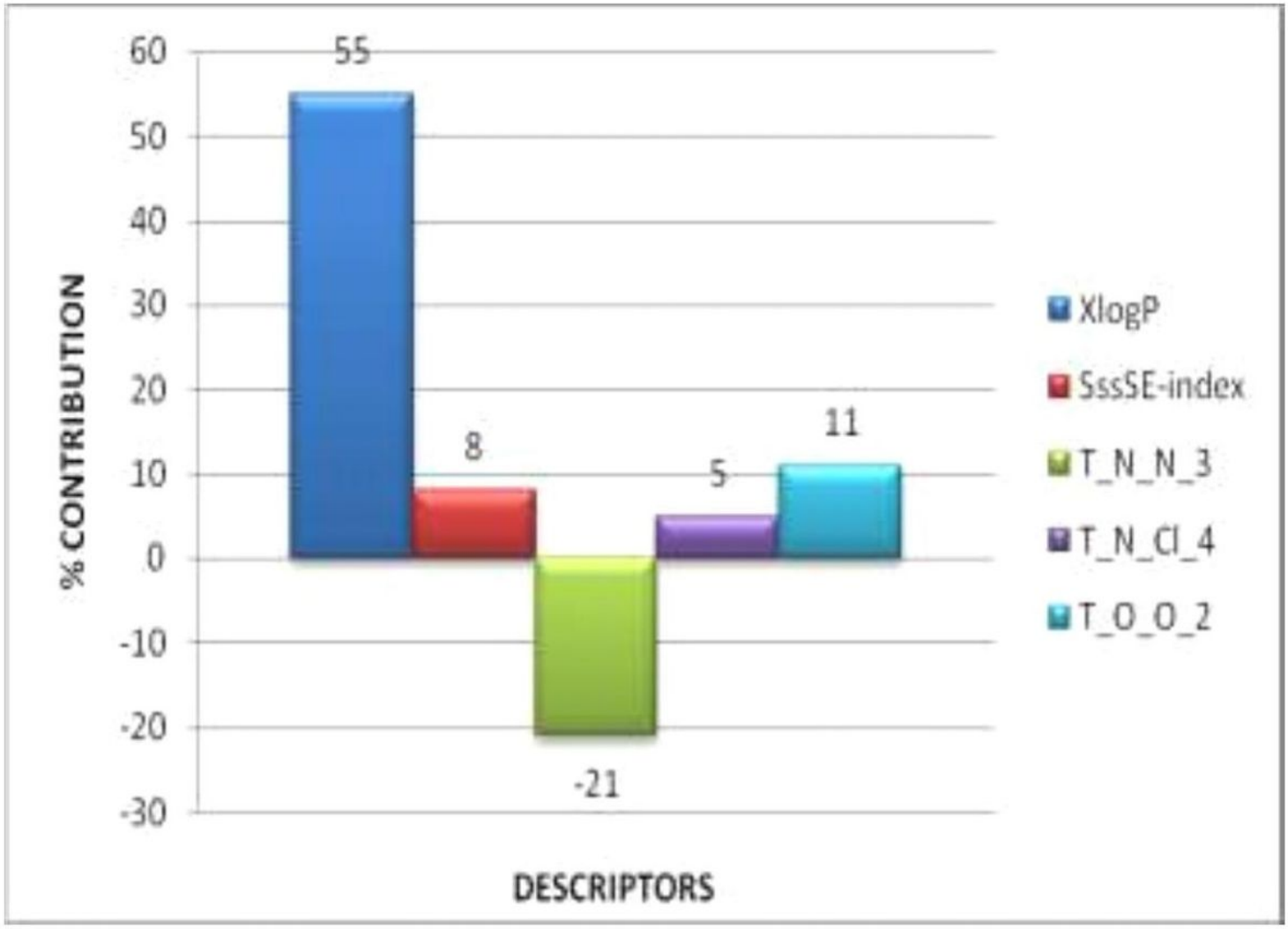


Figure 6

Contribution plot of selected descriptors of set-I

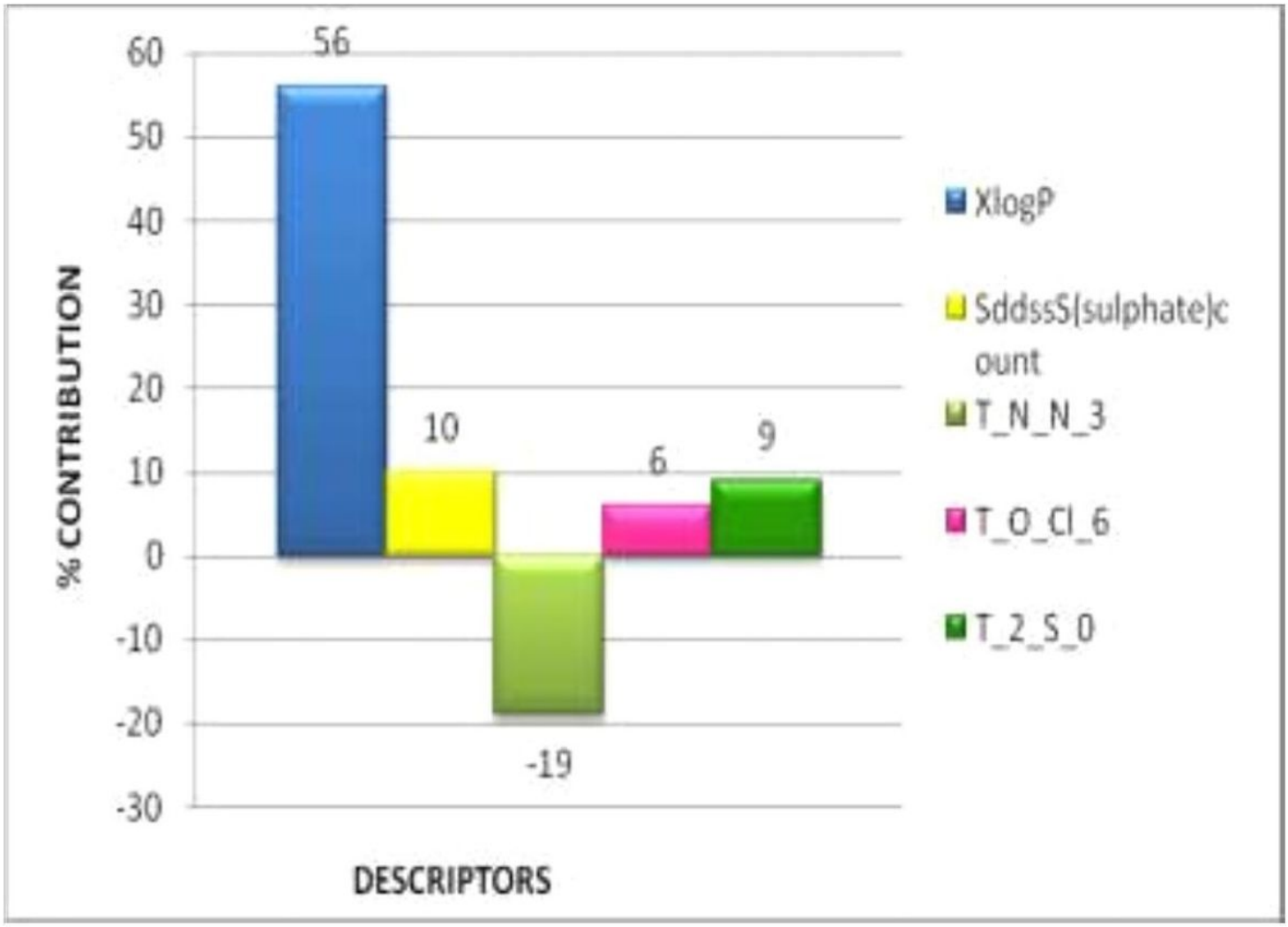
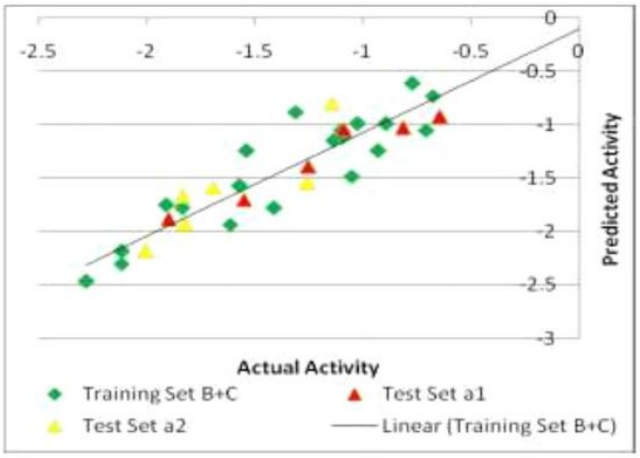
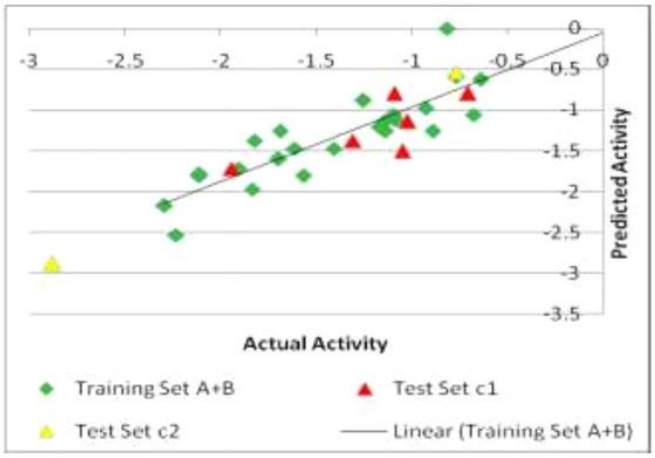


Figure 7

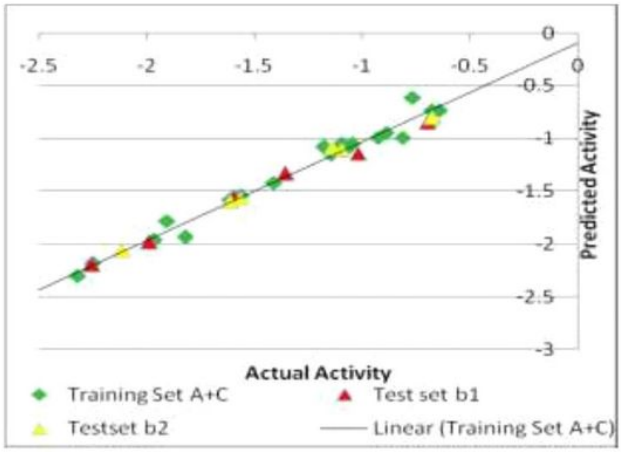
Contribution plot of selected descriptors of set-II



Test set a_1 & a_2 (Training Set B+C)



Test set c_1 & c_2 (Training Set A+B)



Test set b_1 & b_2 (Training Set A+C) [Best model]

Figure 8

Plot of actual versus predicted activity (Set-I)

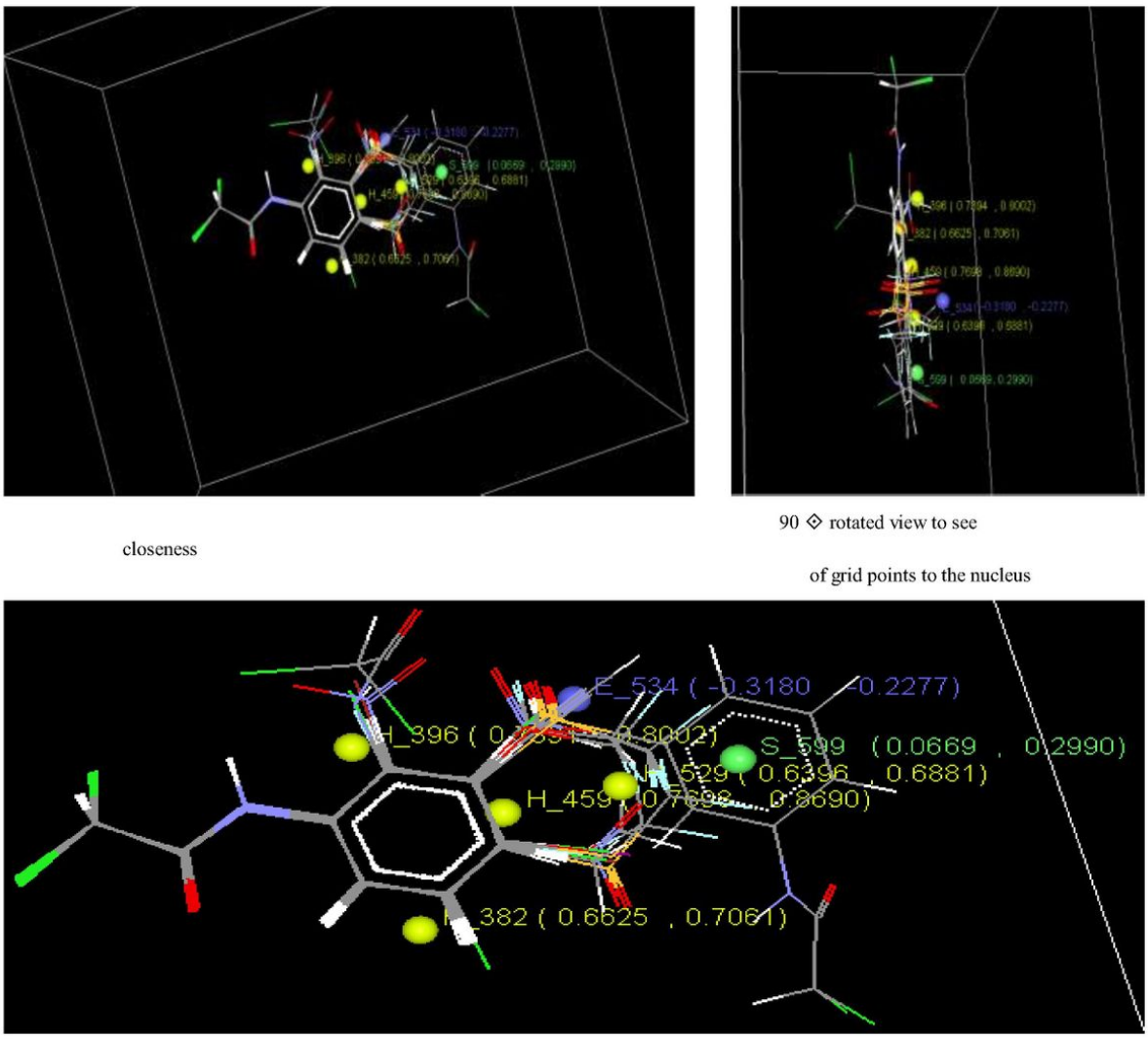


Figure 9

Grid points generated by SA-kNN–MFA method in 3D rectangular grid showing contributions of electrostatic, steric and hydrophobic functional groups for significant anticancer activity (test set b_1)

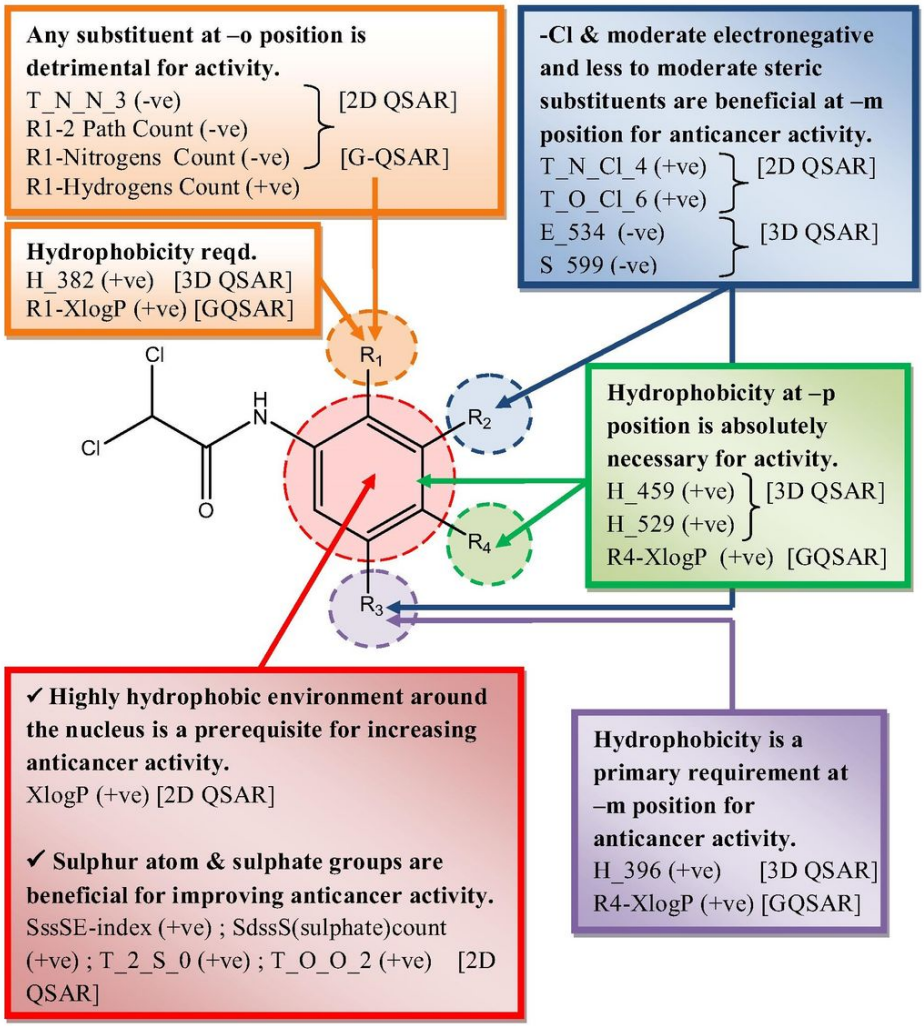


Figure 10

Pharmacophoric requirements around N-phenyl-2, 2-dichloroacetamide nucleus from 2D, 3D & G-QSAR

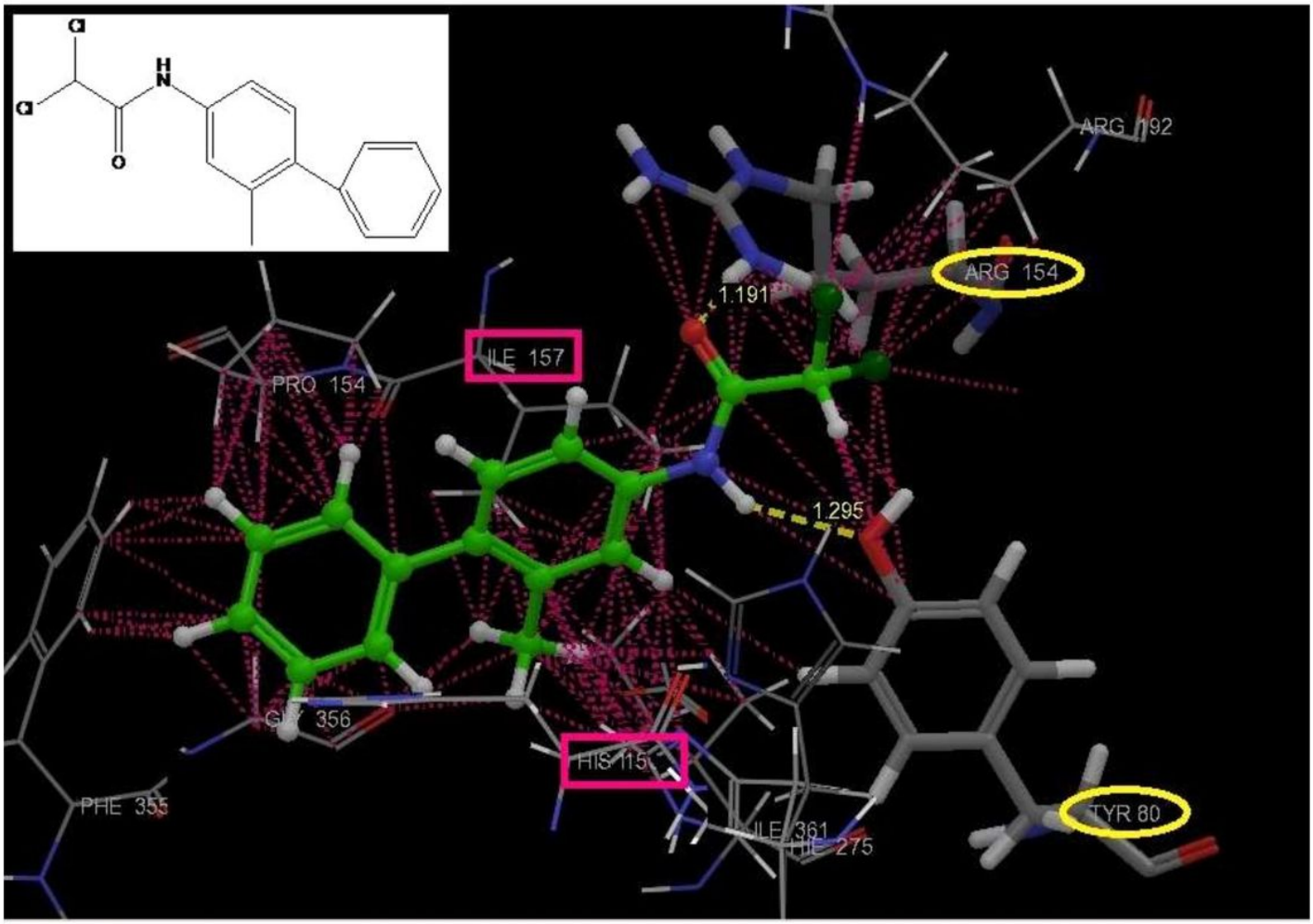


Figure 11

Binding pose of compound D5 in receptor binding pocket of PDK II (PDB Code:2BUB)

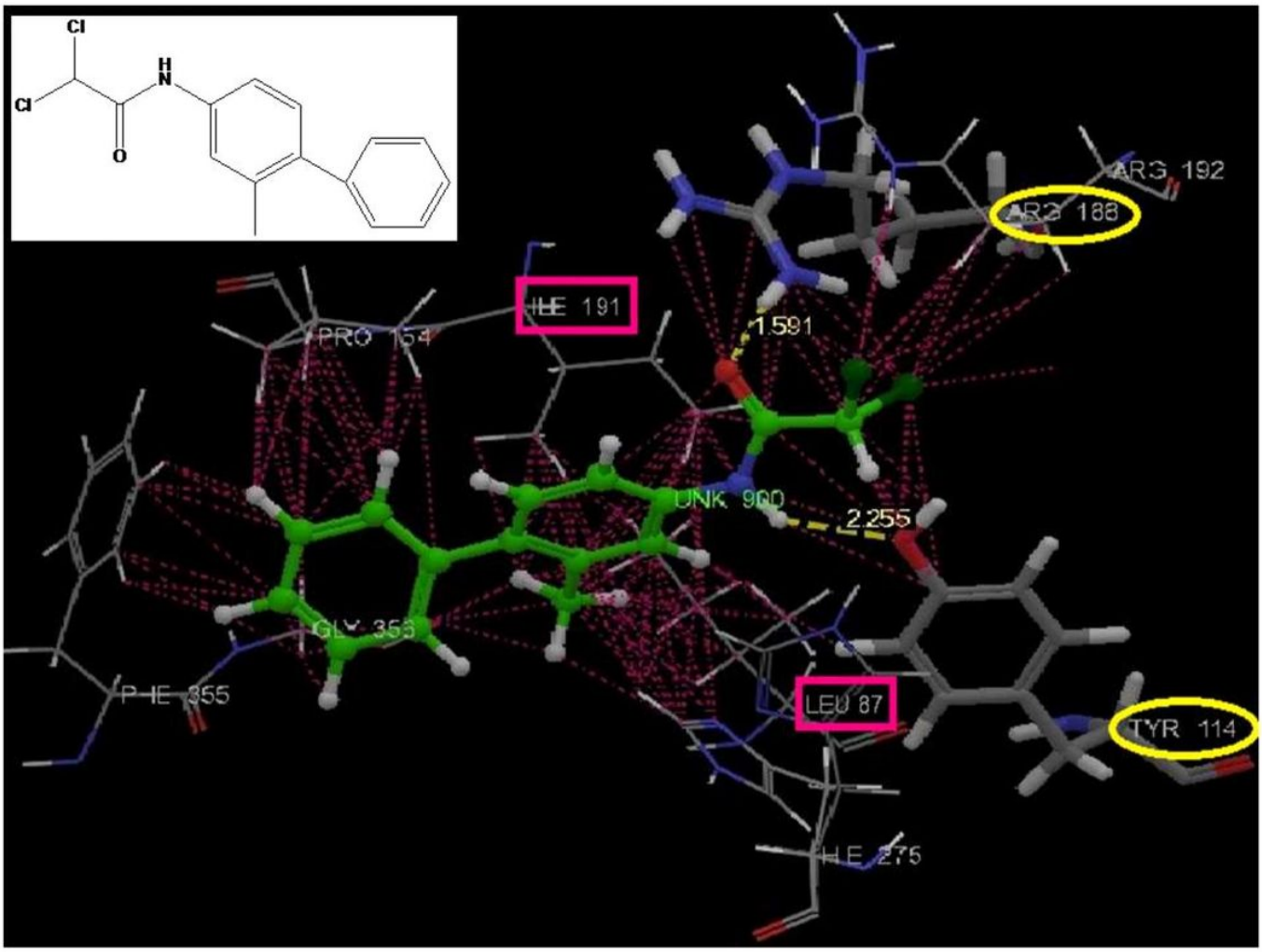


Figure 12

Binding pose of compound D5 in receptor binding pocket of PDK I (PDB Code: 2Q8H)

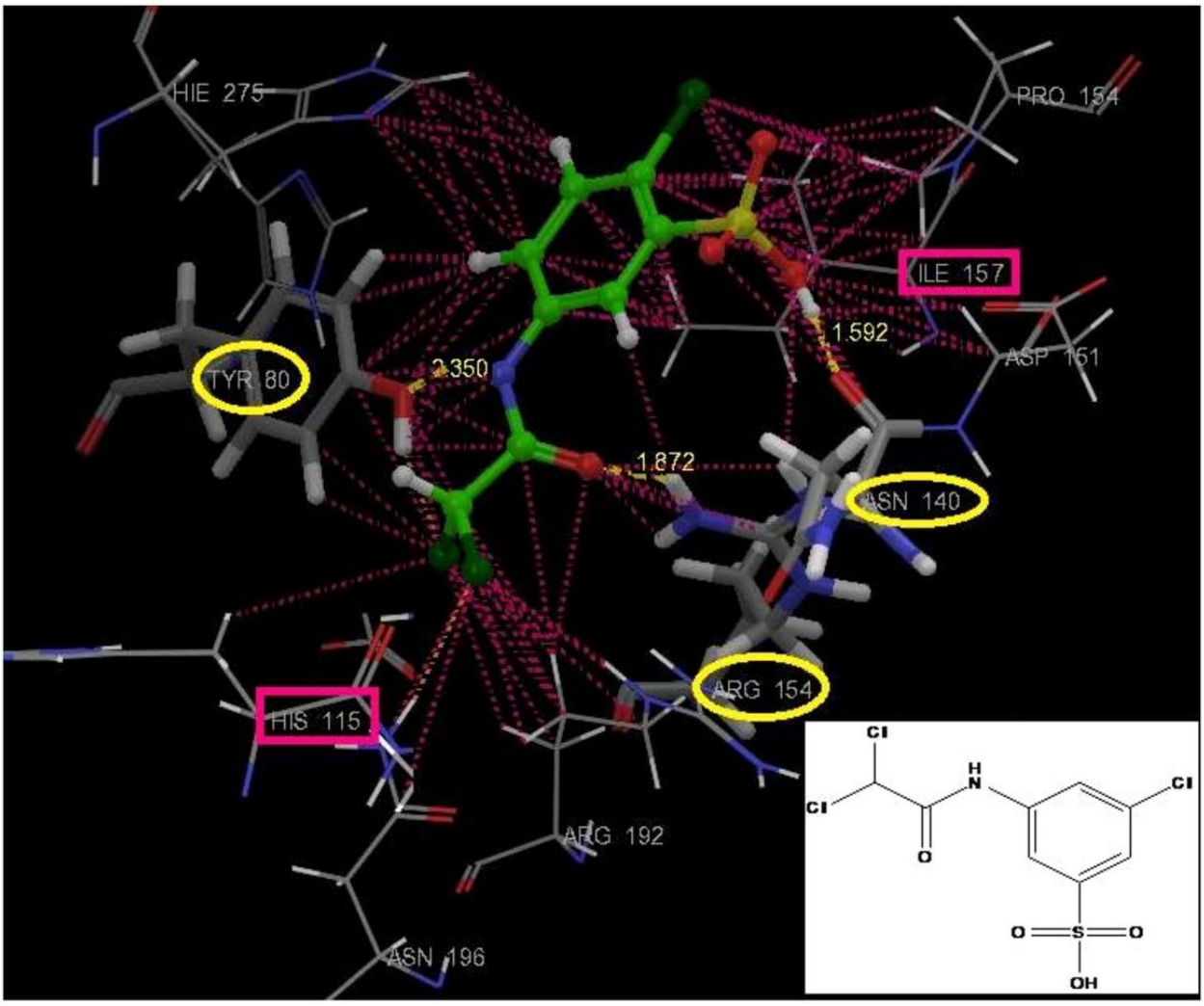


Figure 13

Binding pose of compound D11 in receptor binding pocket of PDK II (PDB Code: 2BU8)

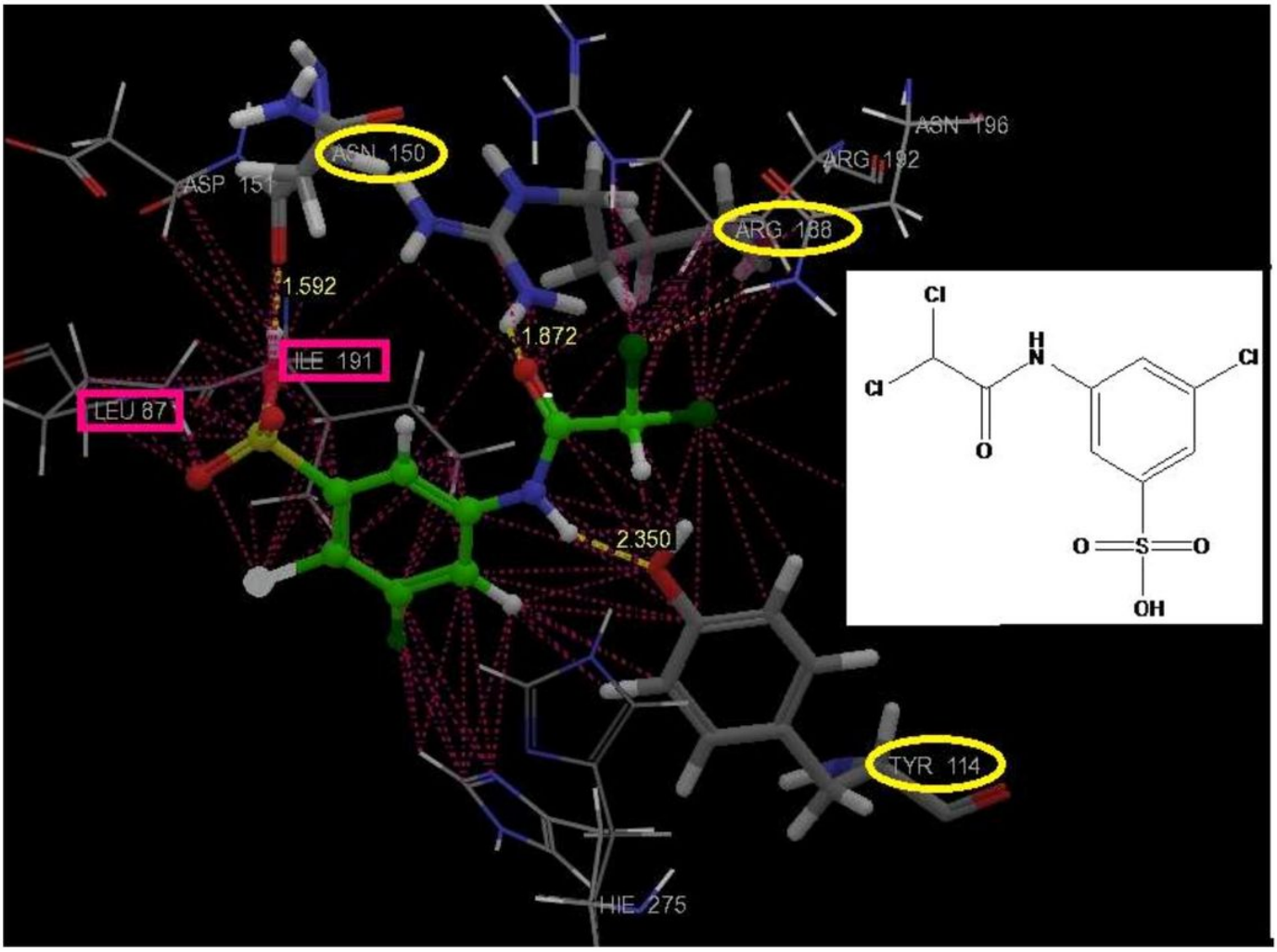


Figure 14

Binding pose of compound D11 in receptor binding pocket of PDK I (PDB Code: 2Q8H)

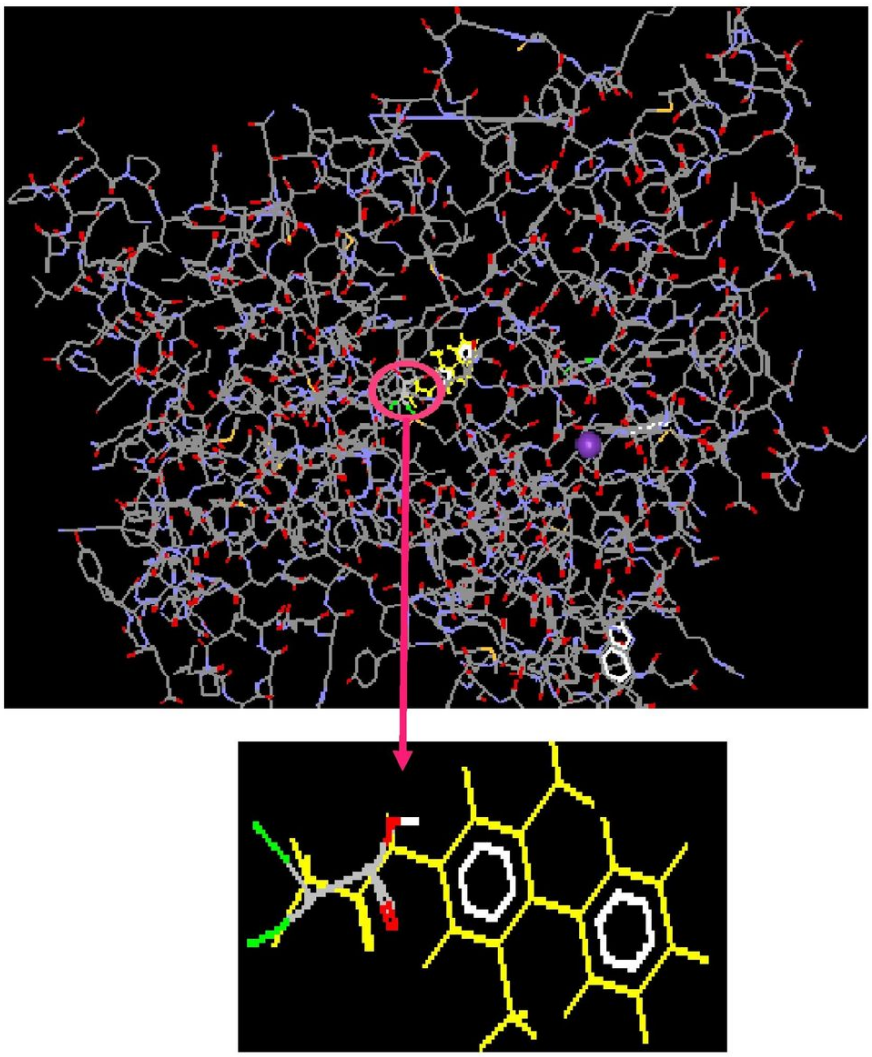


Figure 15

Superimposition of compound D11 on standard (DCA) by VLife MDS 3.5 (PDB Code: 2BU8) Showing the same binding region (pocket) of NCE

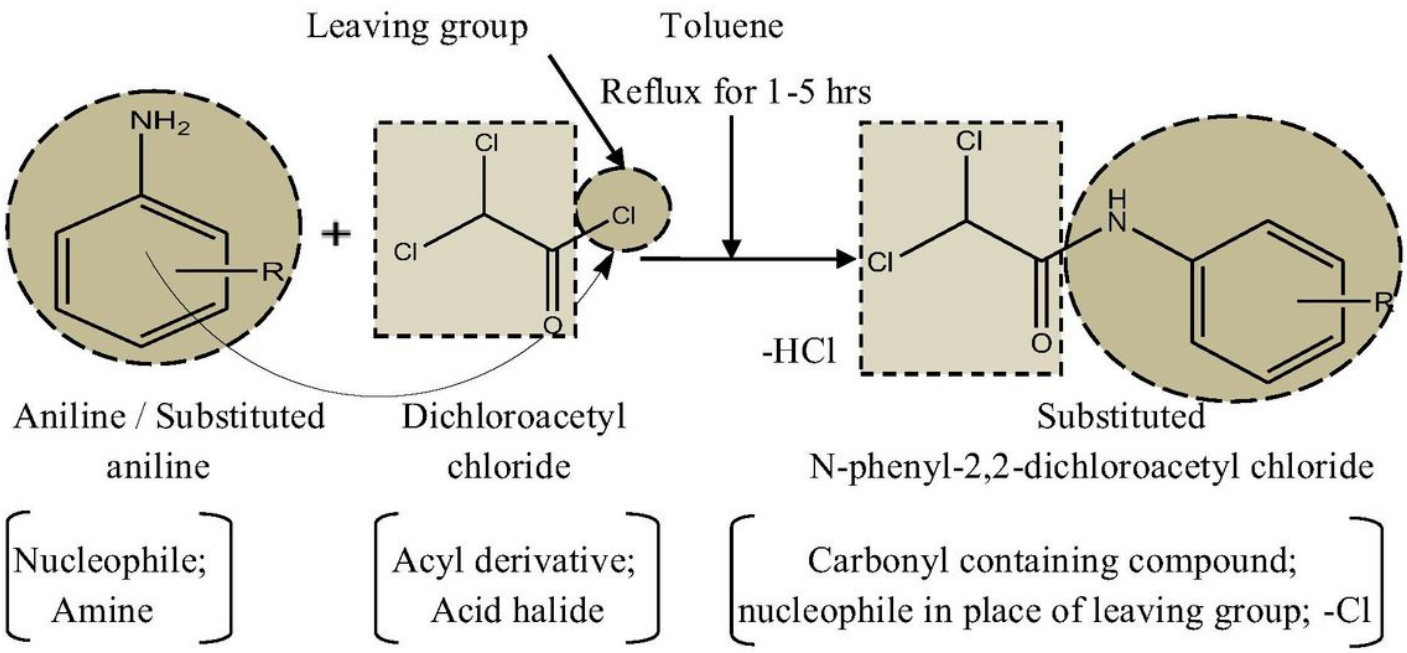


Figure 16

Synthetic route of reaction