Journal of Enzyme Inhibition and Medicinal Chemistry, 2011; 26(3): 319–331

ISSN 1475-6366 print/ISSN 1475-6374 online
DOI: 10.3109/14756366.2010.506437

informa
healthcare

RESEARCH ARTICLE

Use of Quantitative Structure–Activity Relationship (QSAR) and ADMET prediction studies as screening methods for design of benzyl urea derivatives for anti-cancer activity

Deepak Lokwani[1], Shashikant Bhandari[1], Radha Pujari[2], Padma Shastri[2], Ganesh shelke[2], and Vidya Pawar[1]

[1]Department of Pharmaceutical Chemistry, AISSMS College of Pharmacy, Pune, India, and [2]National Centre for Cell Science (NCCS), Ganeshkhind, Pune, India

Abstract

2D and 3D quantitative structure–activity relationship studies have been carried out for establishing a correlation between the structural properties of benzyl urea derivatives and their anti-tumour activities. From this correlation, the new chemical entities were designed, and their activity and absorption, distribution, metabolism, excretion, and toxicity properties were also predicted. Finally, the most promising compounds from these screening were synthesized and biologically evaluated for their anti-cancer properties. Compound 1-(2, 4-dimethylphenyl)-3, 3-dimethyl-1-(2-nitrobenzyl) urea **(7d)** showed significant anti-proliferative activity (at 100 µg/mL) in human cancer cell lines-T-cell leukemia (Jurkat J6), myelogenous leukemia (K562), and breast cancer (MCF-7) compared to reference standard 5-flurouracil.

Keywords: QSAR, k-nearest neighbour–molecular field analysis (kNN–MFA), ADMET, Anti-cancer

Introduction

Cancer drug discovery has undergone a remarkable series of changes over the last decade. The first generation of anti-cancer drugs such as deoxyribonucleic acid (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 anti-cancer drugs, a new class of anti-cancer drugs known as molecularly targeted agents was being developed that work by targeting a biochemical pathway or protein that is unique to or upregulated 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 tyrosine kinases (PTKs)[4].

PTKs protein over expression or gene amplification, mutation, or rearrangement have been demonstrated in multiple malignancies including cancers of the head and neck, ovary, cervix, bladder, prostate, esophagus, stomach, brain, breast, endometrium, colon, and lung[5,6]. Therefore design of inhibitors toward PTKs is an attractive approach for development of new therapeutic agents[7,8]. Most of the tumours initially respond to PTKs but majority of them become resistant to the drug treatment[9,10]. The mechanism underlying acquired drug resistances are not well understood[11].

Our interest have been focused on a special group of compounds containing a urea moiety in their structures, as urea derivatives were synthesized largely in recent years and have become of particular interest to chemists and biologists because of their wide range of biological

Address for Correspondence: Shashikant Bhandari, Department of Pharmaceutical Chemistry, AISSMS College of Pharmacy, Near RTO, Kennedy Road, Pune-411001, Maharashtra, India. E-mail: drugdesign1@gmail.com

(Received 07 March 2010; revised 01 July 2010; accepted 02 July 2010)

activities such as anti-convulsant activity[12], colchicine-binding antagonist[13] and chemokine receptor $CXCR_3$ antagonist[14]. Some of urea derivatives exhibit cell growth inhibition on numerous cancer cell lines and are reported to be potent inhibitors of the PTKs activity of a number of transmembrane growth factor receptors such as 1, 3-disubstituted urea derivatives[15,16], primaquine urea derivatives[17], alkyl[3-(2-chloroethyl)ureido]benzene[18], 4,5-disubstituted thiazolyl urea derivatives[19]. Therefore, search for lead compound is still a continuous quest for the researchers working in this area.

Computational models are one of the powerful tools to design highly active molecules[20,21] that are able to predict structures and the biological activities of anti-cancer compounds. Many QSAR studies[22–26] were developed and published as screening methods for design of new chemical entities (NCE's). As shown in Figure 1, to further explore the structural requirements of selected series of benzyl urea derivatives for their anti-proliferative activity, two methods of QSAR, 2D-QSAR and 3D-QSAR were carried out and NCEs were designed using results of QSAR model. 2D-QSAR and 3D-QSAR studies were performed using multiple linear regression (MLR) analysis[27] and k-nearest neighbour–molecular field analysis (kNN–MFA), respectively. The kNN–MFA methodology relies upon a simple distance learning approach. In this method an unknown member is classified according to the majority of its kNNs in the training set. The nearness is measured by an appropriate distance metrics (e.g. a molecular similarity measure calculated using field interactions of molecular structures). The standard kNN–MFA method[28] was implemented simply as follows: (i) the distances between an unknown object (u) and all other objects in the training set were calculated; (ii) the *k* objects were selected from the training set most similar to object u, according to the calculated distances; and (iii) the object u was classified with the group to which the majority of the *k* objects belong. An optimal *k* value is selected by optimization through the classification of a test set of samples or by leave-one-out (LOO) cross-validation. The variables and optimal *k* values were chosen using different variable selection methods. Here we have used simulated annealing (SA)[29] as variable selection method.

Designed compounds were screened by two types of screening methods for finding of new compounds with anti-cancer activity: (i) Lipinski's rule and prediction of activity using regression equation generated by 2D-QSAR studies and (ii) prediction of absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties. Seven compounds were selected from results of molecular modeling studies and were synthesized. The compounds were screened for examining their anti-cancer effect on human T-cell leukemia cell lines—Molt-4 and Jurkat J6, myelogenous leukemia cell line—K562 and breast cancer cell line—MCF 7.

Experimental part

QSAR studies

All QSAR studies were performed using V-Life Molecular Design Suite Software, version 3.5[30]. Biological activity expressed in terms of IC_{50} was converted in to pIC_{50} ($pIC_{50} = \log 1/IC_{50}$). Both 2D- and 3D-QSAR models were generated using a training set of 20 molecules with varied chemical and biological activities. Test set of four molecules with distributed biological activity was used to assess the predictive power of generated QSAR models. The sphere exclusion method[31] was used for the selection of molecules in training and test sets.

Uni-Column statistics for training set and test set were generated to check correctness of selection criteria for trainings and test set molecules (Table 1). Selection of molecules in the training set and test is a key and important feature of any QSAR model, therefore due care was taken in such a way that biological activities of all compounds in test set lie within the maximum and minimum value range of biological activities of training set of compounds (Table 1).

The maximum and minimum value in training and test set were compared in a way that

1. the maximum value of pIC_{50} of test set should be less than or equal to maximum value of pIC_{50} of training set.
2. the minimum value of pIC_{50} of test set should be higher than or equal to minimum value of pIC_{50} of training set.

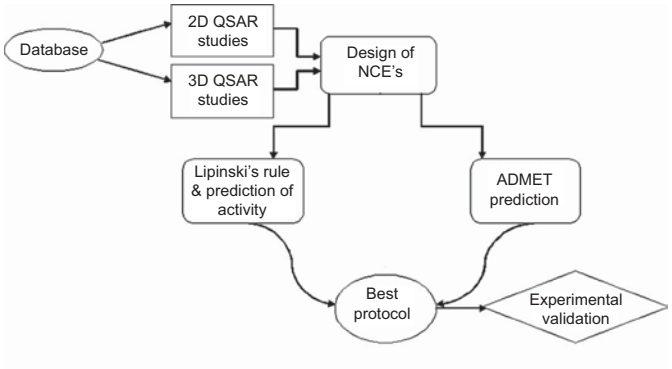


Figure 1. Flowchart of screening methods for new lead compounds.

Table 1. Uni-Column statistics for training and test set of compounds.

	Column name	Average (mean)	Max*	Min*	Std Dev	Sum
Training set	pIC_{50}	−1.1457	0.5690	−1.8760	0.6964	−22.9140
Test set	pIC_{50}	−0.8125	−0.0790	−1.5490	0.6269	−3.2500

*Higher the value of pIC_{50} (Greater +ve value), higher is the potency.

This observation showed that test set was interpolative and derived within the minimum—maximum range of training set.

The mean and standard deviation pIC_{50} values of sets of training and test provide insights to relative difference of mean and point density distribution of two sets.

1. Mean in test set were found to be higher in test set than mean in training set indicating that presence of relatively more active molecules as compared to inactive ones.
2. Higher standard deviation in training set indicated that training set has widely distributed activity of molecules as compared to test set molecules.

The most widely used MLR analysis was used to correlate biological activities with physicochemical properties of the selected series of compounds. Molecules were optimized by Merck Molecular Force Field (MMFF) energy minimization method. A total of 287 2D descriptors were computed using V-Life Molecular Design Suite Software, which include various physicochemical descriptors; Baumann alignment-independent topological descriptors[32] and MMFF atom type count descriptor. After computation of all descriptors, invariable descriptors that have no variation in their values were removed. 2D-QSAR equations were generated by MLR analysis using SA variable selection method.

The 3D-QSAR studies were performed by kNN–MFA using SA variable selection method. kNN–MFA method requires suitable alignment of given set of molecules after optimization, alignment was carried out by template based alignment method (Figure 2). Molecular alignment was used to visualize the structural diversity in the given set of molecules. This was followed by generation of common rectangular grid around the molecules. The steric and electrostatic interaction energies were computed at the lattice points of the grid using a methyl probe of charge +1. The resulting set of aligned molecules was then used to build 3D-QSAR models and information's generated were used to predict activity of those designed molecules that have the similar template or set of atoms.

All generated QSAR models were evaluated and best model was selected using following statistical measures: n, number of molecules (>20 molecules); k, number of descriptors in a model (statistically $n/5$ descriptors in a model); df, degree of freedom ($n - k - 1$) (higher is better); r^2, square of regression (>0.7); q^2, cross-validated r^2 (>0.5); pred_r^2 – r^2 for external test set (>0.5); SEE, standard error of estimate (smaller is better); F-test, F-test for statistical significance of the model (higher is better, for same set of descriptors and compounds); F_prob. Alpha—Error probability (smaller is better); Z score, —Z score calculated by the randomization test (higher is better); best_ran_q^2 highest q^2 value in the randomization test (as low as compared to q^2); best_ran_r^2 highest r^2 value in the randomization test (as low as compared to r^2); α-statistical significance parameter by randomization test (<0.01).

Validation of QSAR studies

Models generated by 2D and 3D-QSAR studies were cross-validated using standard LOO procedure[28]. The cross-validated r^2 (q^2) value was calculated using equation 1, where y_i and $\hat{y}_i$ are the actual and predicted activities of the ith molecule, respectively, and y_{mean} is the average activity of all molecules in the training set. Both summations are over all molecules in the training set. Because the calculation of the pair-wise molecular similarities, and hence the predictions were based upon the current trial solution, the q^2 obtained is indicative of the predictive power of the current kNN–MFA model.

$$q^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - y_{mean})^2} \quad (1)$$

All cross-validation studies were performed by considering the fact that a value of q^2 is above 0.5 indicating high predictive power of generated QSAR models.

External validation[29] of generated models was carried out by predicting the activity of test set of compounds. The predicted r^2 (pred_r^2) value was calculated using equation 2, where y_i and $\hat{y}_i$ are the actual and predicted activities of the ith molecule in test set, respectively, and y_{mean} is the average activity of all molecules in the training set. Both summations are over all molecules in the test set. The pred_r^2 value is indicative of the predictive power of the current kNN–MFA model for external test set.

$$\text{pred\_}r^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - y_{mean})^2} \quad (2)$$

The statistical significance of the QSAR model for an actual data set was further evaluated by randomization test. The robustness of the QSAR models for experimental training sets was examined by comparing generated model with those derived for random data sets. Random sets were generated by rearranging biological activities of the training set molecules. The significance of the models hence obtained was derived based on calculated Z-score[28,29].

ADMET prediction

ADMET properties were calculated using Discovery Studio (DS) 2.1, Accelrys software[33]. DS provides methods

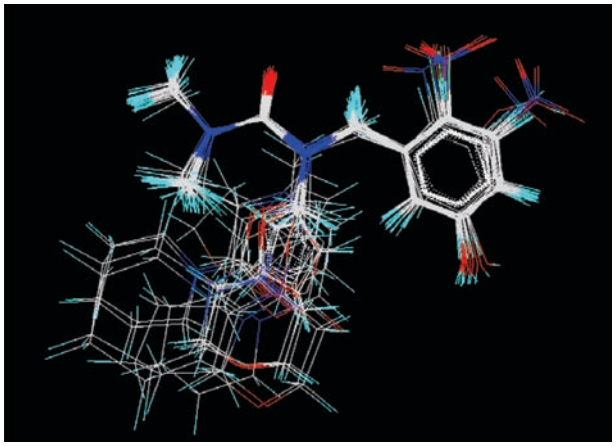


Figure 2. Alignment of substituted benzyl urea derivatives using template based alignment method.

for assessing the disposition and potential toxicity of a ligand within an organism. The ADMET protocols in DS contain published models that are used to compute and analyze ADMET properties. In addition, DS apply specific rules to remove ligands that are not likely drug like, unsuitable leads, etc. based on the presence or absence and frequency of certain chemical groups.

Synthetic studies

All the reactions were carried out with dry, freshly distilled solvents under anhydrous conditions, unless otherwise noted. Fourier transform infrared (FTIR) spectra of the compounds were recorded using KBr on a Varian 640 FTIR spectrophotometer and are reported in cm^{-1}. ^1H NMR spectra were recorded on a Varian Mercury YH300 (300 MHz FT NMR) spectrophotometer using tetramethylsilane as an internal reference (chemical shift represented in δ ppm). Mass spectra were recorded on GC-MS QP5050A System (benchtop quadrupole mass spectrophotometer). Purity of the compounds was checked on thin layer chromatography plates using silica gel G as stationary phase and was visualized in UV light at 254 nm.

Synthesis of secondary amines[15,16] (4a-4g)

Aldehydes **(1a-1g)** (0.1 mol) were dissolved in 25 mL of ethanol and amines **(2a-2g)** (0.1 mol) were added to the solution. The reaction mixture was refluxed for 1–2 h. $NaBH_4$ (0.05 mmol) was added to the reaction mixture solution slowly and stirred under 50°C for 2–3 h. The mixture was evaporated under vacuum and dissolved in ethylacetoacetate (30 mL). The solution was washed with 20 mL water twice, dried over anhydrous sodium sulfate and evaporated.

Synthesis of benzyl urea derivatives[15,16] (7a–7g)

The mixture of CH_2Cl_2 (15 mL), dry DMF **(5)** (3 mL, 40 mmol) and $SOCl_2$ **(6)** (7 mL, 0.10 mol) was stirred at 70°C for 4 h and cooled for 24 h. The solvents and excess $SOCl_2$ were then removed under reduced pressure. The residue dissolved in CH_2Cl_2 (15 mL). Dry pyridine (4 mL) and various amines **(4a-4g)** (40 mmol) were added to the reaction mixture and stirred at 50–60°C for 5–6 h. The reaction mixture was then added to 20 mL ice water, organic layer was separated, and the aqueous layer was extracted with ethyl acetate (2 × 10 mL). The organic layer was combined and washed with saturated $NaHCO_3$, dried with anhydrous Na_2SO_4 for 0.5 h and concentrated under vacuum.

1-benzyl-1-(2-chlorobenzyl)-3,3-dimethylurea (7a)

Yield: 59.10% (Liquid), bp 113–115°C. FTIR (KBr) cm^{-1}: 3095 (Ar C–H), 2976 (Aliphatic C–H), 1712 (C=O), 1314 (ter. C-N); 1081 (C-Cl). ^1H NMR (300 MHz, DMSO, δ ppm): 2.82 (s, 6H, CH_3), 4.58 (s, 4H, CH_2), 7.26–7.72 (m, 9H, Aromatic). Mass: m/z 302 (M^+), 303 (M+1) $^+$. ^{13}C NMR (100 MHz, DMSO, δ ppm): 37.2 (-N(CH_3)$_2$), 50.7 (-CH_2-N-CH_2), 127.2 (Ph C-4), 128.8 (Ph C-3,5), 129.8 (Ph C-2,6), 130.2 (Cl-Ph C-3,4,5), 131.3 (Cl-Ph C-6), 132.8 (Cl-Ph C-2), 134.4 (Ph C-1), 137.3 (Cl-ph C-1), 168.1 (C=O). Elemental Analysis calculated for $C_{17}H_{19}ClN_2O$: C, 67.54; H, 6.29; N, 9.27. Found: C, 67.46; H, 6.38; N, 9.35.

1-benzyl-3,3-dimethyl-1-(2-nitrobenzyl)urea (7b)

Yield: 90.13% (Liquid), bp 107–109°C. FTIR (KBr) cm^{-1}: 3041 (Ar C–H), 2815 (Aliphatic C–H), 1716 (C=O), 1533 (N-O), 1345 (ter C-N), 863 (Ar C-N). ^1H NMR (300 MHz, DMSO, δ ppm): 2.91 (s, 6H, CH_3), 4.63 (s, 4H, CH_2), 7.21–782 (m, 9H, Aromatic). Mass: m/z 313 (M^+), 314 (M+1) $^+$. ^{13}C NMR (100 MHz, DMSO, δ ppm): 37.3 (-N(CH_3)$_2$), 50.9 (-CH_2-N-CH_2), 126.7 (NO_2-Ph C-3,4), 127.3 (Ph C-4), 128.8 (Ph C-3,5), 129.9 (Ph C-2,6), 130.3 (NO_2-Ph C-6), 132.5 (NO_2-Ph C-5), 134.6 (Ph C-1), 137.4 (NO_2-Ph C-1), 144.1 (NO_2-ph C-2), 168.7 (C=O). Elemental Analysis calculated for $C_{17}H_{19}N_3O_3$: C, 65.15; H, 6.07; N, 13.41. Found: C, 65.09; H, 5.98; N, 13.53.

1-(2-chlorobenzyl)-3,3-dimethyl-1-(naphthalen-1-yl) urea (7c)

Yield: 41.53% (Liquid), bp 118–120°C. FTIR (KBr) cm^{-1}: 3029 (Ar C–H), 2923 (Aliphatic C–H), 1722 (C=O), 1319 (ter. C-N); 1085 (C-Cl). ^1H NMR (300 MHz, DMSO, δ ppm): 2.81 (s, 6H, CH_3), 4.63 (s, 4H, CH_2), 7.16–7.29 (m, 4H, Aromatic), 7.51–7.65 (m, 7H, Naphthyl). Mass: m/z 338 (M^+), 339 (M+1) $^+$. ^{13}C NMR (100 MHz, DMSO, δ ppm): 37.6 (-N(CH_3)$_2$), 49.9 (-N-CH_2), 108.1 (Naphthyl C-2), 120.3 (Naphthyl C-4,9), 125.5 (Naphthyl C-3,6,7,8), 128.1 (Ph C-3,4,5,6), 131.7 (Ph C-2), 133.5 (Naphthyl C-5,10), 136.1 (Naphthyl C-1), 138.1 (Ph C-1), 168.5 (C=O). Elemental Analysis calculated for $C_{20}H_{19}ClN_2O$: C, 71.00; H, 5.62; N, 8.28. Found: C, 72.02; H, 5.53; N, 8.37.

1-(2,3-dimethylphenyl)-3,3-dimethyl-1-(2-nitrobenzyl) urea (7d)

Yield: 41.89% (Liquid), bp 114–115°C. FTIR (KBr) cm^{-1}: 3012 (Ar C–H), 2881 (Aliphatic C–H), 1721 (C=O), 1523 (N-O); 1329 (ter C-N), 859 (Ar C-N). ^1H NMR (300 MHz, DMSO, δ ppm): 2.20 (s, 3H, CH_3), 2.33 (s, 3H, CH_3), 2.80 (s, 6H, CH_3), 4.82 (s, 2H, CH_2), 7.41–7.52 (m, 3H, Aro-

matic), 7.57–7.73 (m, 3H, Aromatic). Mass: m/z 327 (M^+), 328 (M+1) $^+$. ^{13}C NMR (100 MHz, DMSO, δ ppm): 18.5 (2,4 dimethyl), 37.1 (-N(CH_3)$_2$), 50.7(-N-CH_2), 127.5 (dimethyl Ph C-3,5,6), 128.7 (Ph C-3,4,5,6), 130.5 (dimethyl Ph C-2,4), 132.1 (Ph C-1), 136.6 (dimethyl Ph C-1), 144.3 (Ph C-2), 169.5 (C=O). Elemental Analysis calculated for $C_{18}H_{21}N_3O_3$: C, 66.04; H, 6.47; N, 12.84. Found: C, 66.10; H, 6.34; N, 12.76.

1-(2-ethylbenzyl)-3,3-dimethyl-1-(naphthalen-1-yl) urea (7e)

Yield: 83.13% (Liquid), bp 122–124°C. FTIR (KBr) cm^{-1}: 3025 (Ar C–H), 2934 (Aliphatic C–H), 1717 (C=O), 1359 (ter C-N). ^{1}H NMR (300 MHz, DMSO, δ ppm): 1.19–1.35 (t, 3H, CH_3, J = 8.1 Hz), 2.58–2.74 (q, 2H, CH_2, J = 8.1 Hz), 2.93 (s, 6H, CH_3), 4.92 (s, 2H, CH_2), 7.21–7.36 (m, 4H, Aromatic), 7.47–7.76 (m, 7H, Naphthyl). Mass: m/z 332 (M^+), 333 (M+1) $^+$. ^{13}C NMR (100 MHz, DMSO, δ ppm): 14.3 (-CH_2CH_3), 26.3 (-CH_2CH_3), 37.4 (-N(CH_3)$_2$), 49.8 (-N-CH_2), 108.3 (Naphthyl C-2), 120.5 (Naphthyl C-4,9), 124.3 (Ph C-3,4,5,6), 125.8 (Naphthyl C-3,6,7,8), 128.2 (Ph C-2), 132.9 137.1 (Naphthyl C-1), (Naphthyl C-5,10), 138.3 (Ph C-1), 168.4 (C=O). Elemental Analysis calculated for $C_{22}H_{24}N_2O$: C, 79.51; H, 7.22; N, 8.43. Found: C, 79.44; H, 7.10; N, 8.55.

1-(3-ethoxy-4-hydroxybenzyl)-3,3-dimethyl-1-(naphthalen-1-yl)urea (7f)

Yield: 62.50% (Liquid), bp 103–105°C. FTIR (KBr) cm^{-1}: 3589 (Ar OH), 3098 (Ar C–H), 2859 (Aliphatic C–H), 1716 (C=O), 1391 (ter C-N); 1329 (ter C-N), 1256 (C-O-C). ^{1}H NMR (300 MHz, DMSO, δ ppm): 1.24–1.34 (t, 3H, CH_3, J = 8.1 Hz), 2.76 (s, 6H, CH_3), 4.03–4.17 (q, 2H, CH_2, J = 8.2 Hz), 4.77 (s, 2H, CH_2), 5.59 (s, 1H, OH), 7.16–7.23 (m, 3H, Aromatic), 7.37–7.52 (m, 7H, Naphthyl). Mass: m/z 364 (M^+), 365 (M+1) $^+$. ^{13}C NMR (100 MHz, DMSO, δ ppm): 15.1 (-CH_2CH_3), 37.2 (-N(CH_3)$_2$), 50.4 (-N-CH_2), 66.2 (-CH_2CH_3), 108.4 (Naphthyl C-2), 117.1(Ph C-2,5), 120.1 (Naphthyl C-4,9), 127.3 (Ph C-3,6,7,8), 128.1 (Ph C-1), 132.4 (Naphthyl C-5,10), 133.5 (Ph C-1), 136.8 (Naphthyl C-1), 147.5 (Ph C-3,4), 169.2 (C=O). Elemental Analysis calculated for $C_{22}H_{24}N_2O_3$: C, 72.52; H, 6.59; N, 7.69. Found: C, 72.63; H, 6.49; N, 7.62.

1-(2,3-dimethylphenyl)-1-(2-ethylbenzyl)-3,3-dimethylurea (7g)

Yield: 73.73% (Liquid), bp 124–125°C. FTIR (KBr) cm^{-1}: 2998 (Ar C–H), 2917 (Aliphatic C–H), 1693 (C=O), 1320 (ter C-N). ^{1}H NMR (300 MHz, DMSO, δ ppm): 1.14–1.26 (t, 3H, CH_3), 2.02 (s, 3H, CH_3, J = 7.9 Hz), 2.21 (s, 3H, CH_3), 2.62-2.78 (q, 3H, CH_3, J = 7.9 Hz), 2.29 (s, 6H, CH_3), 4.84 (S, 2H, CH_2), 7.11–7.43 (m, 6H, Aromatic). Mass: m/z 310 (M^+), 311 (M+1) $^+$. ^{13}C NMR (100 MHz, DMSO, δ ppm): 13.9 (-CH_2CH_3), 17.6 (2,4 dimethyl), 26.3 (-CH_2CH_3), 38.2 (-N(CH_3)$_2$), 49.8 (-N-CH_2), 126.2 (Ph C-2,4,5,6), 128.1 (dimethyl Ph C-3,4,5), 133.2 (dimethyl Ph C-2,4), 136.1 (Ph C-1,3), 137.2 (dimethyl Ph C-1), 167.4 (C=O). Elemental Analysis calculated for $C_{20}H_{26}N_2O$: C, 77.38; H, 8.44; N, 9.02. Found: C, 77.23; H, 8.35; N, 9.09.

Anti-proliferative activity by MTT assay

The human T-cell leukemia cell lines (Molt-4 and Jurkat J6), myelogenous leukemia cell line (K562) and breast cancer cell line (MCF-7) were procured from ATCC (American type culture collection). Molt-4, Jurkat J6, and K562 were maintained in RPMI 1640 medium and MCF-7 cell line in Eagle's minimum essential medium supplemented with 10% fetal calf serum and 100 U of penicillin and streptomycin. The cell lines were cultured and maintained in a humidified atmosphere of 5% CO_2 at 37°C.

The anti-proliferative activity of the compounds was determined by MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay, using 5-flurouracil as reference drug. Cell lines were seeded at a density of 25,000 cells/well/100 µL in a 96 well tissue culture plate for suspension cultures and 5000 cells /well in a 96 well tissue culture plate for adherent cultures. Suspension cells were treated with compounds **7a–7g** within 1 h of seeding whereas adherent cells were allowed to adhere for 12 h and medium was replaced. The compounds at concentration of 25, 50, 100, and 200 µg/mL were used in the volume of 10 µL/well. After the incubation of 48 h, the cell survival was determined by addition of MTT solution (10 µL/well of 5 mg/mL MTT in phosphate-buffered saline) and incubated for 4 h at 37°C in 5% CO_2. For suspension cells, the formazon crystals were solubilized by adding 10% sodium dodecyl sulfate in 0.1N HCL, incubated overnight and optical absorbance was measured at 570–650 nm. For adherent cells, the medium containing MTT was aspirated; DMSO (100 µL/well) was added and optical absorbance was measured after 10 min at 570 nm.

Results and discussion

QSAR studies

A series of total 24 compounds for which absolute IC_{50} values reported[15] was used for correlating chemical composition (structure) with their anti-proliferative activity. Several 2D-QSAR and 3D-QSAR models were generated for training set of 20 compounds using MLR and SA kNN–MFA (SA kNN–MFA) method, respectively. The best QSAR model was selected on the basis of value of Statistical parameters like r^2 (square of correlation coefficient for training set of compounds), q^2 (cross-validated r^2), and pred_r^2 (predictive r^2 for the test set of compounds). All QSAR models were validated and tested for its predictability using an external test set of four compounds. Statistical results generated by both 2D and 3D-QSAR analysis showed that both QSAR model have good internal as well as external predictability (Table 2).

The generated QSAR models were evolved by repeating the MLR and kNN–MFA methods to check the accuracy and precision of both the methods. The frequency of use of a particular descriptor in the population of equations indicated the relevant contributions of the descriptors. The best 2D-QSAR model had five contributing descriptors including constant (equation 3)

Table 2. Statistical results of 2D-QSAR equation generated by MLR method and 3D-QSAR model generated by SA kNN–MFA for benzyl urea derivatives.

Sr. No.	Statistical parameter	Results 2D-QSAR	3D-QSAR
1.	r^2	0.8787	—
2.	r^2SE	0.2411	—
3.	q^2	0.7215	0.6276
4.	q^2SE	0.2378	0.2843
5.	Pred_r^2	0.5728	0.9106
6.	Pred_r^2SE	0.2162	0.2199
7.	F-Test	19.914	—
8.	α q^2	0.0100	—
9.	Best-Rand q^2	0.2458	—
10.	Z-score q^2	2.6730	—
11	N	20	20
12	Nearest neighbour	—	4
11	Contributing descriptors	1. chiv1 (+ 0.6694) 2. MMFF_6 (−0.8569) 3. T_C_O_7 (−0.1222) 4.T_N_O_2 (+1.1713)	1. S_715(0.3304, 6.0782) 2. E_232 (-0.0255, 0.1432) 3. E_1064 (0.1346, 0.1345)

$$\text{pIC50} = +0.6694\,\text{chiv1} - 0.8569\,\text{MMFF\_6} - \text{O.1222}\,\text{T\_C\_O\_7} + 1.1713\,\text{T\_N\_O\_2} - 4.7880 \quad (3)$$

Results of 3D-QSAR indicated the requirement of two electropositive groups and one steric group around benzyl urea pharmacophore for anti-cancer activity (Table 2). The kNN–MFA QSAR method explores formally the active analogue approach, which implies that compounds display similar profiles of pharmacological activities. In this method, the activity of each compound is predicted as average activity of *k* most chemically similar compounds from that data set. The generated 3D-QSAR model showed a good correlation between the experimental obtained and computationally predicted activity (Figure 3). Residual values obtained by subtraction of predicted activities from experimental biological activities were found to near to zero indicating the good predictability of selected QSAR Model (Table 3). The plots of observed vs. predicted activity for the optimal cross-validated kNN–QSAR model are depicted in Figure 3.

Interpretation of QSAR studies

It is simple to interpret 2D-QSAR MLR equation where each descriptor's contribution can be seen by the magnitude and sign of its regression coefficient. A descriptor's coefficient magnitude shows its relative contribution with respect to other descriptors and sign indicates whether it is directly (+) or inversely (–) proportional to the activity. The developed MLR equation 2 indicated that the descriptor T_N_O_2 played most significantly important role (contribution ~41% for biological activity) for anti-proliferative activity of benzyl urea derivatives (Figure 4A). T_N_O_2 is an alignment-independent descriptor that specifies the count of number of nitrogen atoms (single, double, or triple bonded) separated from any oxygen atom (single or double bonded) by 2 bond distances in a molecule. This descriptor suggested the requirement of urea molecule in benzyl urea pharmacophore for biological activity. The next influential descriptor was MMFF_6 (contribution ~30% for biological activity) that was inversely proportional to the activity. This is an atom type count descriptor based on MMFF atom types and their count in each molecule and '6' indicated number of times that atom type has been found in a given molecule.

The descriptor chiv1 (contribution ~24% for biological activity) was found to be directly proportional to the activity and showed the role of atomic valence connectivity index (order 1) that is calculated as the sum of $1/\sqrt{v_i v_j}$ over all bonds between heavy atoms i and j where $i < j$. Finally, the descriptor governing variation in the activity was T_C_O_7 (contribution approximately 5% for biological activity) and was found to be inversely proportional to the activity, which indicated that the presence of substituents with direct attachment of oxygen on aromatic ring (e.g. OH) may be unfavorable or detrimental for the anti-cancer activity.

3D-QSAR was used to optimize the electrostatic and steric requirements around benzyl urea pharmacophore. 3D data points were generated that contribute to SA kNN–MFA 3D-QSAR model, are shown in Figure 4B. The range of property values for the generated data points helped for the design of potent NCEs. The range was based on the variation of the field values at the chosen points using the most active molecule and its nearest neighbour set. The points generated in SA kNN–MFA 3D-QSAR model are E_232 (-0.0255, 0.1432), S_715 (0.3304, 6.0782), and E_1064 (0.1346, 0.1345) i.e. electrostatic and steric interaction field at lattice points 232, 715, and 1064, respectively. These points were suggested the significance and requirement of electrostatic and steric properties as mentioned in the ranges in parenthesis for structure-activity relationship and maximum biological activity of benzyl urea derivatives.

Negative and positive values in electrostatic field descriptors indicated the requirement of negative and positive electrostatic potential respectively for enhancing the biological activity of benzyl urea derivatives. Therefore electronegative and electropositive substituents were preferred at the position of generated data points E_232 (-0.0255, 0.1432) and E_1064 (0.1346, 0.1345) respectively around benzyl urea pharmacophore.

Similarly the positive values of steric descriptors suggested the requirement of sterically bulky groups at the position of generated data point S_715 (0.3304, 6.0782) around benzyl urea pharmacophore for maximum activity. Thus the KNN–MFA models leads to identification of various local interacting molecular features responsible for activity variation and hence provide direction for design of new molecules in a convenient way.

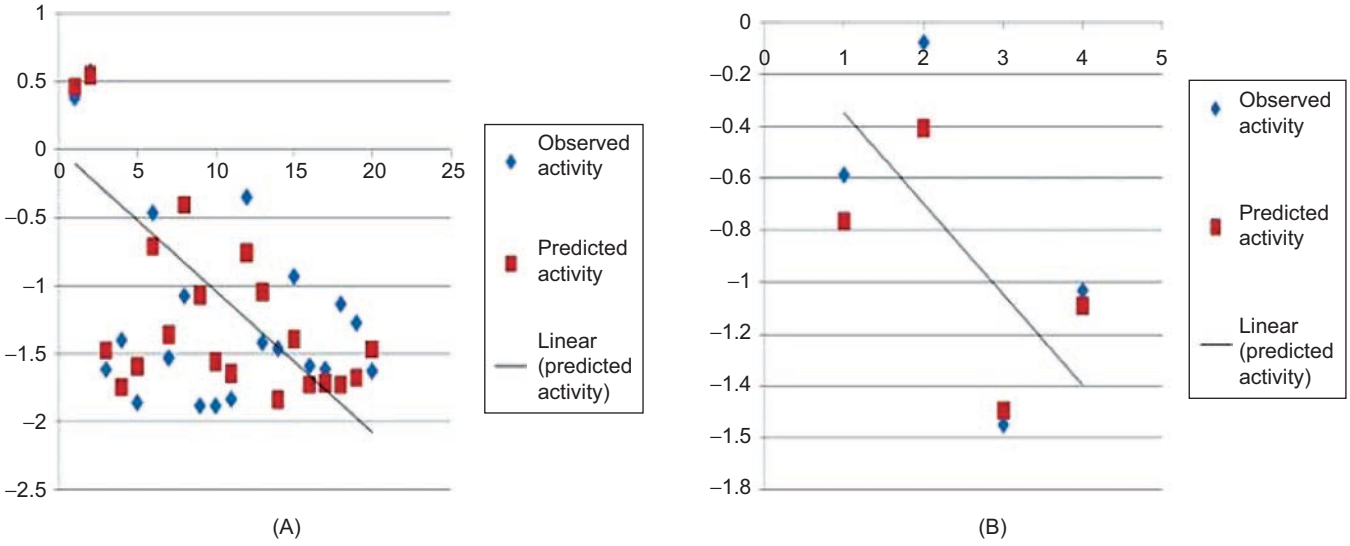
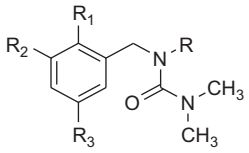


Figure 3. Comparison of observed activity vs predicted activity for training set (A) and test set (B) of compounds.

Table 3. Structure of training and test sets of compounds along with observed and predicted activity.



					Observed activity		SA kNN–MFA	
Mol. No.	R	R_1	R_2	R_3	IC_{50} (μM)	PIC_{50}	Predicted activity	Residual
1	-4-propylmorpholine	NO_2	H	H	0.42	0.3768	0.4596	−0.0828
2	-4-ethylmorpholine	NO_2	H	H	0.27	0.5686	0.5438	0.0248
3	-2-methylfuran	NO_2	H	H	40.7	−1.6095	−1.4756	−0.1339
4	-ethylpiperidine-1-carboxylate	NO_2	H	H	24.9	−1.3961	-1.7448	0.3487
5	-2-methylfuran	Cl	H	OH	71.3	−1.8530	−1.5958	−0.2572
6*	-4-ethylmorpholine	Cl	H	OH	3.88	−0.5888	−0.7637	0.1754
7	-4-propylmorpholine	Cl	H	OH	2.92	−0.4653	−0.7129	0.2476
8	-ethylpiperidine-1-carboxylate	Cl	H	OH	33.5	−1.5250	−1.3619	−0.1631
9	-4-ethylmorpholine	H	OH	H	11.8	−1.0718	−0.409	−0.662
10	-4-propylmorpholine	H	OH	H	7.5	−0.8750	−1.074	0.199
11	-2-methylfuran	H	OH	H	75.1	−1.8756	−1.5603	−0.3153
12	-ethylpiperidine-1-carboxylate	H	OH	H	67.2	−1.8273	−1.6452	−0.1821
13	-4-ethylmorpholine	H	NO_2	H	2.24	−0.3504	−0.7603	0.4099
14*	-4-propylmorpholine	H	NO_2	H	1.20	−0.0791	−0.4080	0.3289
15*	-2-methylfuran	H	NO_2	H	35.4	−1.5490	−1.4951	−0.0539
16	-ethylpiperidine-1-carboxylate	H	NO_2	H	26.0	−1.4149	−1.048	−0.366
17	-2-methylfuran	NO_2	H	OH	28.6	−1.4563	−1.8394	0.3831
18*	-4-ethylmorpholine	NO_2	H	OH	10.8	−1.0334	−1.0911	0.0577
19	-4-propylmorpholine	NO_2	H	OH	8.5	−0.9294	−1.3932	0.4638
20	-ethylpiperidine-1-carboxylate	NO_2	H	OH	38.4	−1.5843	−1.7281	0.1438
21	-2-methylfuran	Br	H	OH	40.5	−1.6074	−1.7168	0.1094
22	-4-ethylmorpholine	Br	H	OH	13.5	−1.1303	−1.7308	0.6005
23	-4-propylmorpholine	Br	H	OH	18.6	−1.2695	−1.6762	0.4067
24	-ethylpiperidine-1-carboxylate	Br	H	OH	41.8	−1.6211	−1.4703	−0.1514

*Compounds in test set.

Design and screening of NCEs

The information obtained from 2D and 3D-QSAR studies were used to optimize the electrostatic and steric requirements around the benzyl urea nucleus for enhancing the anti-cancer activity (Figure 4C).

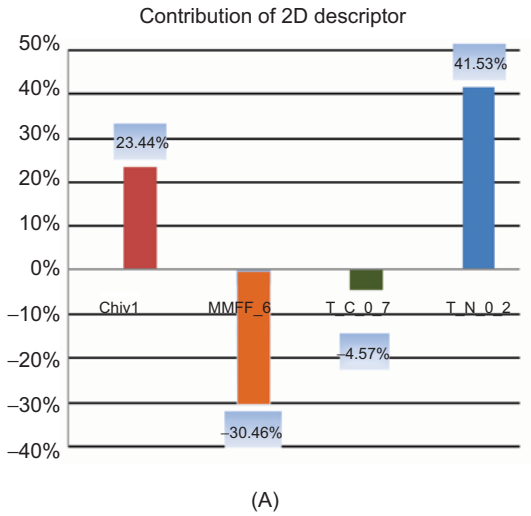
Descriptors generated in 2D-QSAR equation signified the importance of urea group for anti-cancer activity of compounds. Similarly electrostatic and steric points generated around common template or pharmacophore in 3D-QSAR suggested substitution of sterically bulky group at N-R_2 position around urea and electropositive group at $\mathbf{R_1}$ position around phenyl ring.

The NCEs were designed by CombiLib tool of V-Life Molecular Design Suite Software using pharmacophore shown in Figure 4C. CombiLib tool passed the designed compounds through Lipinski's screen to ensure drug-like pharmacokinetic profile of the designed compounds. The Lipinski's screen parameters used as filters are:

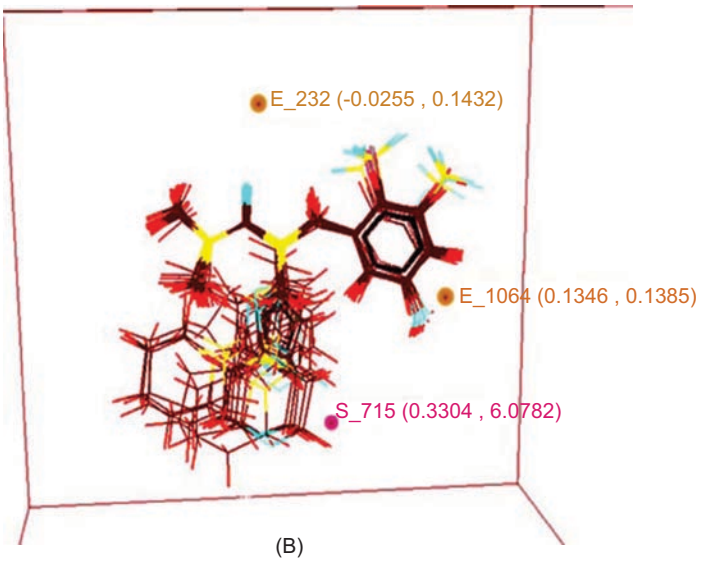
1. number of Hydrogen Bond Acceptor (A) (<10)
2. number of Hydrogen Bond donor (B) (<5)
3. number of Rotatable Bond (R) (<10)
4. XlogP (X) (<5)
5. molecular weight (W) (<500 g/mol)
6. polar surface area (S) is (<140 A^0)

More than 100 molecules were generated using CombiLib tool that follows the Lipinski's rule, but we have selected only 30 most active molecules on the basis of their activity, predicted using MLR equation (Table 4).

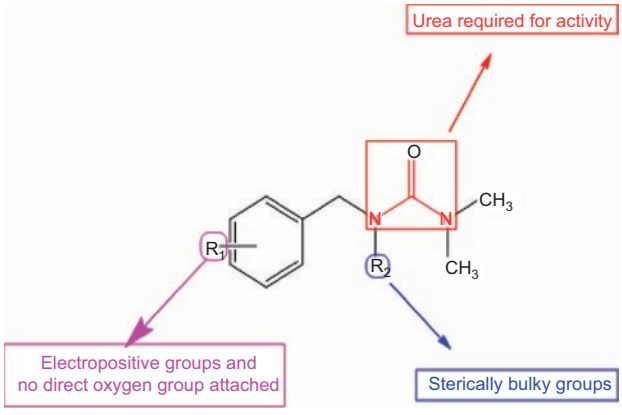
In Table 4, Compounds qualifying all required parameters set for Lipinski's screen/filter are indicated by ADRXWS strings. The columns containing the Lipinski's screen score and strings of alphabets, ADRXWS indicated that all six Lipinski's screen parameters are satisfied by corresponding compound and depending on number of requirement satisfied, screen score varied in between 1 and 6. It was concluded from predicted activity of designed compound that presence of electropostive group at $\mathbf{R_1}$ position mainly at ortho and para position of phenyl



(A)



(B)



(C)

Figure 4. Results of QSAR studies. (A) Contributions of 2D descriptors for biological activity developed using MLR equation (2D-QSAR). (B) Contributions of electrostatic and steric 3D data points towards biological activity developed using kNN–MFA method (3D-QSAR). (C) Pharmacophore requirement around benzyl urea.

Table 4. Structure of designed NCEs along with predicted activity obtained by MLR equation generated by 2D-QSAR.

Sr. No.	Compound code	R_1	R_2	Screen result	Screen score	Predicted activity
1	E5	2-NO_2	Benzyl	ADRXWS	6	0.413
2	A1	2-CH_3	Nephthyl	ADRXWS	6	0.356
3	J2	2-C_2H_5	2,4-Dimethyphenyl	ADRXWS	6	0.291
4	E2	2-NO_2	2,4-Dimethyphenyl	ADRXWS	6	0.146
5	A5	2-CH_3	Benzyl	ADRXWS	6	−0.0198
6	J8	2-C_2H_5	1-Morpholine	ADRXWS	6	−0.0921
7	A8	2-CH_3	1-Morpholine	ADRXWS	6	−0.1992
8	C1	3-OCH_3 4-OH	Nephthyl	ADRXWS	6	−0.2112
9	J1	2-C_2H_5	Nephthyl	ADRXWS	6	−0.2214
10	E1	2-NO_2	Nephthyl	ADRXWS	6	−0.2256
11	E8	2-NO_2	1-Morpholine	ADRXWS	6	−0.2318
12	H1	2-Cl	Nephthyl	ADRXWS	6	−0.2377
13	A6	2-CH_3	2-Furan	ADRXWS	6	−0.2412
14	H5	2-Cl	Benzyl	ADRXWS	6	−0.2489
15	J7	2-C_2H_5	2-Pyrrol	ADRXWS	6	−0.2801
16	A2	2-CH_3	2,4-Dimethyphenyl	ADRXWS	6	−0.3125
17	H8	2-Cl	1-Morpholine	ADRXWS	6	−0.3890
18	E7	2-NO_2	2-Pyrrol	ADRXWS	6	−0.4114
19	E6	2-NO_2	2-Furan	ADRXWS	6	−0.4223
20	C2	3-OCH_3 4-OH	2,4-Dimethyphenyl	ADRXWS	6	−0.4569
21	C5	3-OCH_3 4-OH	Benzyl	ADRXWS	6	−0.4987
22	D8	2-OH	1-Morpholine	ADRXWS	6	−0.5239
23	C8	3-OCH_3 4-OH	1-Morpholine	ADRXWS	6	−0.5673
24	D1	2-OH	Nephthyl	ADRXWS	6	−0.6166
25	F1	3-NO_2	Nephthyl	ADRXWS	6	−0.7709
26	D5	2-OH	Benzyl	ADRXWS	6	−0.9876
27	H2	2-Cl	2,4-Dimethyphenyl	ADRXWS	6	−0.9943
28	D6	2-OH	2-Furan	ADRXWS	6	−1.2319
29	F5	3-NO_2	Benzyl	ADRXWS	6	−1.2879
30	J6	2-C_2H_5	2-Furan	ADRXWS	6	−1.1341

ring and presence of sterically bulky group like benzyl, naphthyl at **R_2** position significantly enhanced the activity of compounds.

In next screening step, ADMET properties of all designed compounds were predicted and compared with predicted ADMET properties of standard compounds using DS, Accelrys software (Table 5). Prediction of ADMET properties was used as last screen to sort out those compounds that already followed Lipinski's rule and showed good predicted activity. DS defined the prediction level for all ADMET properties. For ADMET Absorption, there are four prediction level: 0 – Good, 1 – Moderate, 2 – Poor, 3 – Very Poor, for ADMET plasma protein binding (PPB) level: 0 – binding is <90%, 1 – binding is >90%, 2 – binding is >95%, for ADMET Cytochrome P450 2D6 (CYP2D6) Probability level: <0.5 – Unlikely to inhibit CYP2D6 enzyme (Non-inhibitors of CYP2D6), >0.5 – Likely to inhibit CYP2D6 enzyme (Inhibitors of CYP2D6) and for ADMET probability level: < 0.5 – Unlikely to cause dose-dependent liver injuries (Non-toxic), >0.5 – Likely to cause dose-dependent liver injuries (Toxic).

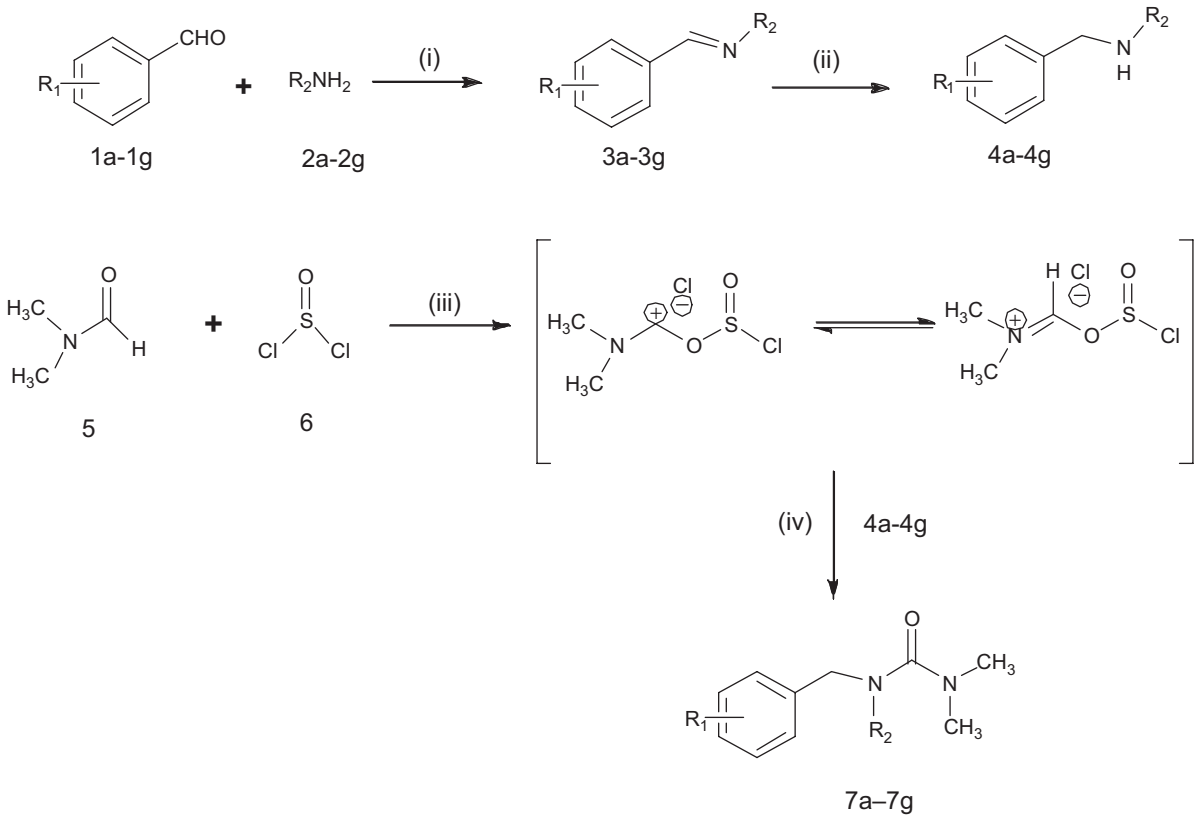
All designed compounds showed good absorption, PPB level and found non-inhibitors of CYP2D6 enzyme. As shown in Table 5, Gefitinib and 5-flurouracil showed probability of heptotoxicity, whereas all designed compounds were found to be non-toxic to liver.

Chemistry

The seven compounds (Table 5) that followed the all screening criteria were selected and synthesized. The synthetic route is outlined in **Scheme 1**. In the first step, condensation reaction of aldehydes **(1a-1g)** with various amines **(2a-2g)** gave the Schiff bases **(3a-3g)**. Reduction of the latter with sodium borohydride afforded the corresponding secondary amines **(4a-4g)**. In second step, Vilsmeier–Haack reaction was carried out, in which dry dimethyl formamide **(5)** in thionyl chloride **(6)** and dichloromethane was refluxed for 4 h and excessive $SOCl_2$ was evaporated. The residue was then dissolved in dichloromethane, and stirred with various secondary amines **(4a-4g)** obtained in the first step dissolved in dry pyridine and CH_2Cl_2. As a result, benzyl urea derivatives **(7a–7g)** were synthesized. The structures of various synthesized compounds were assigned on the basis of different chromatographic, spectral and qualitative and quantitative organic analytical studies. The physical data, FTIR, ^{1}HNMR and mass spectral data for all synthesized compounds are reported in experimental protocols.

Table 5. Prediction of ADMET properties by using discovery studio, accelrys.

Sr. No.	Compound code	ADMET absorption level	ADMET PPB level	ADMET CYP2D6 probability	ADMET hepatotoxicity probability
1	C-1	0	2	0.772	0.655
2	E-5	0	2	0.455	0.456
3	E-2	0	0	0.297	0.529
4	E-7	0	0	0.435	0.357
5	F-1	0	2	0.772	0.596
6	J-2	0	1	0.584	0.589
7	D-6	0	0	0.415	0.357
8	H-5	0	2	0.366	0.317
9	H-1	0	2	0.841	0.609
10	F-5	0	2	0.455	0.43
11	J-1	0	2	0.841	0.642
12	C-5	0	2	0.613	0.602
13	C-2	0	1	0.485	0.609
14	A-8	0	2	0.405	0.072
15	E-1	0	2	0.782	0.682
16	Imatinib	0	1	0.514	0.655
17	Gefitinib	0	2	0.663	0.973
18	5-Flurouracil	1	0	0.019	0.834



Scheme 1. Synthetic route for benzyl urea derivatives. Reagents & condition: (i) ethanol, reflux, 1–2 h (ii) $NaBH_4$, 50°C, 2–3 h (iii) dry CH_2Cl_2, 70°C, 4 h (iv) dry CH_2Cl_2, dry pyridine, 60°C, 5–6 h.

Anti-proliferative activity

All synthesized compounds were evaluated for their effect on proliferation in Molt-4, Jurkat J6, MCF-7, and K562 cell lines by MTT colorimetric assay. Compounds were tested at 25, 50, 100, and 200 g/mL concentrations and their % decrease in cell proliferation were calculated by considering untreated controls as 100%. 5-Flurouracil was used as reference compound. The results are depicted in Table 6 and Figure 5.

Molt-4 cell line did not respond to any of the seven compounds tested but in K562 cell line, all compounds showed anti-proliferative activity. In case of Jurkat J6, and MCF-7, only compound **7d** exhibited significant anti-proliferative effect. Compound **7d** showed 35% and

Table 6. Anti-proliferative activity of compounds **7a–7g** in cancer cell lines.

Compound code	R_1	R_2	Conc (μg/mL)	% Inhibition			
				Molt-4 (%)	Jurkat J6 (%)	MCF-7 (%)	K562 (%)
7a (H-5)	-2-Cl	-Benzyl	25	0	3.9	8.85	—
			50	6.32	7.45	10.07	6.71
			100	10.42	7.78	10.07	9.59
			200	—	—	—	17.80
7b (E-5)	-2-NO_2	-Benzyl	25	4.44	4.82	17.3	—
			50	6.5	4.6	9.22	0
			100	0.52	0.65	4.80	6.19
			200	—	—	—	16.66
7c (H-1)	-2-Cl	-1-Naphthyl	25	0	2.08	7.50	—
			50	4.66	3.61	5.54	14.92
			100	0.30	5.5	2.11	12.91
			200	—	—	—	7.94
7d (E-2)	-2-NO_2	-2,4-dimethylphenyl	25	7.83	9.4	10.81	—
			50	3.91	18	11.91	13.08
			100	7.98	35	17.67	42.05
			200	—	—	—	45.37
7e (J-1)	-2-C_2H_5	-1-Naphthyl	25	0	4.82	0	—
			50	0.30	10.05	0	10.64
			100	3.46	11.73	7.99	13.78
			200	—	—	—	18.15
7f (C-1)	-3-OC_2H_5-4-OH	-1-Naphthyl	25	0	18.09	0.52	—
			50	4.59	10.05	17.79	9.773
			100	2.48	16.55	2.36	12.91
			200	—	—	—	18.67
7g (J-2)	-2-C_2H_5	-2,4-dimethylphenyl	25	6.6	14.36	0	—
			50	7.68	12.39	0	3.40
			100	0	14.3	0	11.43
			200	—	—	—	20.49
5-Flurouracil	—	—	25	65	40	4.93	—
			50	64	41	15.71	13.35
			100	61	37	21.59	7.76
			200	—	—	—	8.55

17.7% decrease in cell proliferation in Molt-4 and Jurkat cell lines respectively at concentration of 100 μg/mL, and was comparable with positive control-5-flurouracil that showed 37%, and 21.59% decrease in cell proliferation, respectively. Compound **7d** exhibited good anti-proliferative activity (42.05%) as compared to that of 5-flurouracil (7.76%) at concentration of 100 μg/mL in K562 cell line. These results suggested that presence of NO_2 group at ortho position of phenyl ring $\mathbf{R_1}$ position and 2,4-dimethylphenyl at $\mathbf{R_2}$ position improved the cytotoxicity in MCF -7, Jurkat J6, and K562 cell line. This also indicated that presence of –Cl, and C_2H_5 at ortho position of phenyl ring at $\mathbf{R_1}$ position and benzyl and naphthyl group at $\mathbf{R_2}$ position showed poor cytotoxic activity. From QSAR results, it was found that substituents at $\mathbf{R_1}$ position with direct oxygen substituents may decrease the cytotoxic activity but these compounds showed good binding affinity at EGFR receptor found by docking results, hence compound **7f** containing OC_2H_5 at meta position and OH at para position of phenyl ring at $\mathbf{R_1}$ position was taken further for experimental validation, it show comparable cytotoxic activity with 5-flurouracil in MCF-7 but poor *in vitro* anti-tumour activity against Jurkat J6. Therefore, it was revealed that presence of oxygen group at phenyl ring at $\mathbf{R_1}$ position may decrease the activity. Also substituents at ortho position of phenyl ring at $\mathbf{R_1}$ position rather than meta and para position may help in increasing the *in vitro* anti-proliferative activity.

Conclusion

In this study, we tested the possibility of developing NCE′s using QSAR and ADMET predicting studies. The proposed 3D and 2D-QSAR models, due to the high internal and external predictive ability, can therefore act as a useful aid to the costly and time consuming experiments for determining electrostatic and steric requirement around benzyl urea pharmacophore that is required to achieve better anti-proliferative activity and thus aided in generation of diverse combinatorial library. Prediction of ADMET properties of designed compound helped in selecting the compounds having drug-like properties. A screening approach has thus facilitated the identification of suitable compounds from designed library for

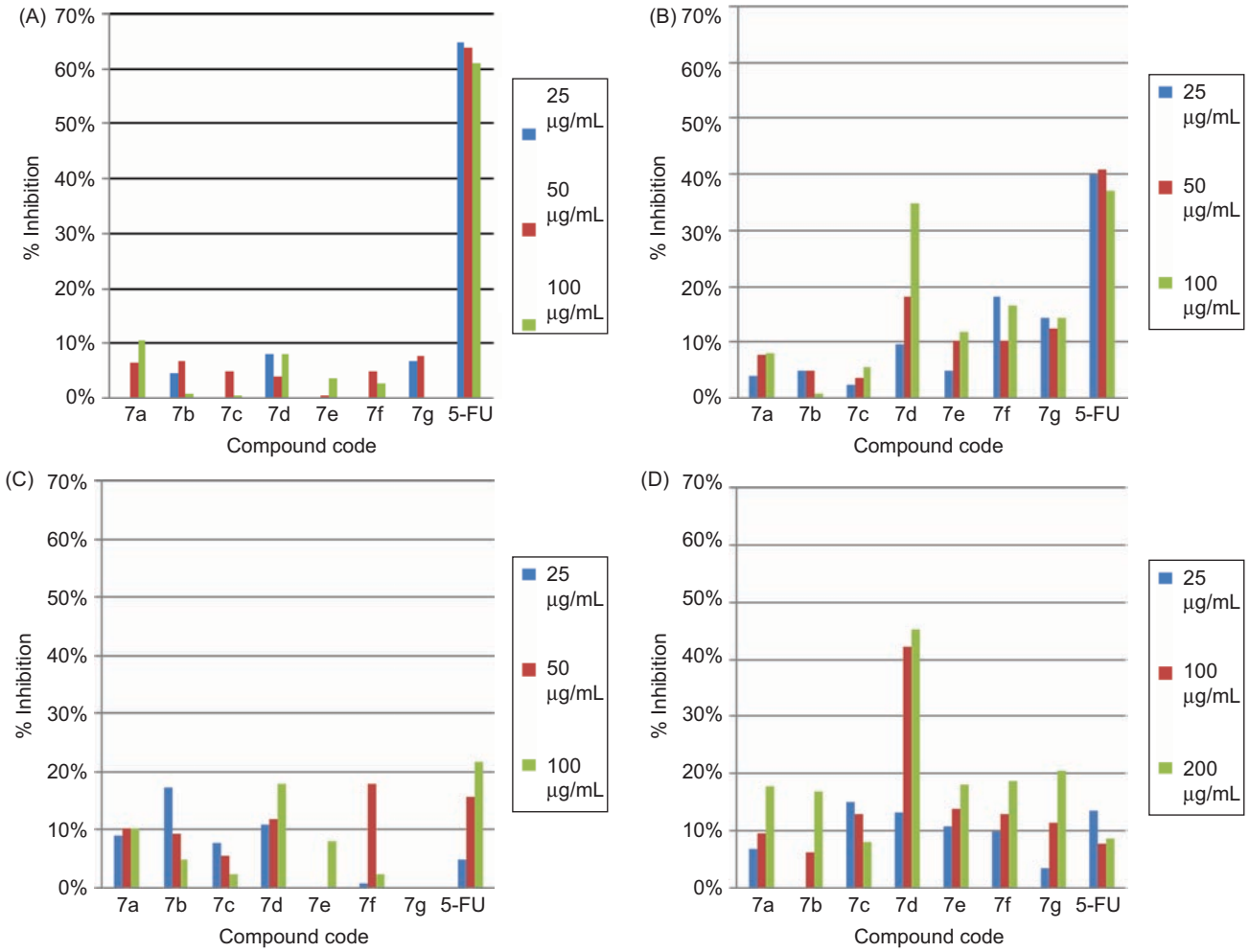


Figure 5. Effect of compounds **7a–7g** on (A) Molt-4, (B) Jurkat J6, (C) MCF-7, and (D) K562 cell lines on cell proliferation determined by MTT assay.

anti-cancer activity. Results of molecular modeling studies were cross verified by testing the cytotoxic activity of designed compounds in four cell line; Molt-4, Jurkat J6, MCF-7, and K562. Compound **7d** was found to be the potent among the compounds tested. Therefore this was concluded that presence of electropositive group like NO_2 at phenyl ring at $\mathbf{R}_1$ position and sterically bulky group at $\mathbf{R}_2$ like 2, 4-dimethylphenyl position can help the benzyl urea pharmacophore for inhibiting the tumour cells.

Declaration of interest

The authors report no conflicts of interest.

References

1. Chabner BA, Roberts TG Jr. Timeline: Chemotherapy and the war on cancer. Nat Rev Cancer 2005; 5:65–72.
2. Workman P. Genomics and the second golden era of cancer drug development. Mol Biosyst 2005; 1:17–26.
3. Druker BJ. Perspectives on the development of a molecularly targeted agent. Cancer Cell 2002; 1:31–36.
4. Morin MJ. From oncogene to drug: development of small molecule tyrosine kinase inhibitors as anti-tumor and anti-angiogenic agents. Oncogene 2000; 19:6574–6583.
5. Krause DS, Van Etten RA. Tyrosine kinases as targets for cancer therapy. N Engl J Med 2005; 353:172–187.
6. Hirsch FR, Varella-Garcia M, Bunn PA Jr, Di Maria MV, Veve R, Bremmes RM, Barón AE, Zeng C, Franklin WA. Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. J Clin Oncol 2003; 21:3798–3807.
7. Grünwald V, Hidalgo M. Developing inhibitors of the epidermal growth factor receptor for cancer treatment. J Natl Cancer Inst 2003; 95:851–867.
8. Bridges AJ. Chemical inhibitors of protein kinases. Chem Rev 2001; 101:2541–2572.
9. Kobayashi S, Boggon TJ, Dayaram T, Jänne PA, Kocher O, Meyerson M, Johnson BE, Eck MJ, Tenen DG, Halmos B. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. N Engl J Med 2005; 352:786–792.
10. Yamamotoa H, Toyookaa S, Mitsudomib T. Impact of EGFR mutation analysis in non-small cell lung cancer. Lung Cancer 2009; 63:315–321.
11. Kwak EL, Sordella R, Bell DW, Godin-Heymann N, Okimoto RA, Brannigan BW, Harris PL, Driscoll DR, Fidias P, Lynch TJ, Rabindran SK, McGinnis JP, Wissner A, Sharma SV, Isselbacher KJ, Settleman J, Haber DA. Irreversible inhibitors of the EGF receptor may circumvent acquired resistance to gefitinib. Proc Natl Acad Sci USA 2005; 102:7665–7670.
12. Shimshoni JA, Bialer M, Wlodarczyk B, Finnell RH, Yagen B. Potent anticonvulsant urea derivatives of constitutional isomers of valproic acid. J Med Chem 2007; 50:6419–6427.

13. Fortin JS, Lacroix J, Desjardins M, Patenaude A, Petitclerc E, C-Gaudreault R. Alkylation potency and protein specificity of aromatic urea derivatives and bioisosteres as potential irreversible antagonists of the colchicine-binding site. Bioorg Med Chem 2007; 15:4456–4469.
14. Watson RJ, Allen DR, Birch HL, Chapman GA, Galvin FC, Jopling LA, Knight RL, Meier D, Oliver K, Meissner JW, Owen DA, Thomas EJ, Tremayne N, Williams SC. Development of CXCR3 antagonists. Part 3: Tropenyl and homotropenyl-piperidine urea derivatives. Bioorg Med Chem Lett 2008; 18:147–151.
15. Li HQ, Zhu TT, Yan T, Luo Y, Zhu HL. Design, synthesis and structure-activity relationships of antiproliferative 1,3-disubstituted urea derivatives. Eur J Med Chem 2009; 44:453–459.
16. Cao P, Huang XF, Ding H, Ge HM, Li HQ, Ruan BF, Zhu HL. Synthesis and cytotoxic evaluation of substituted urea derivatives as inhibitors of human-leukemia K562 cells. Chem Biodivers 2007; 4:881–886.
17. Dzimbeg G, Zorc B, Kralj M, Ester K, Pavelic K, Andrei G, Snoeck R, Balzarini J, De Clercq E, Mintas M. The novel primaquine derivatives of N-alkyl, cycloalkyl or aryl urea: synthesis, cytostatic and antiviral activity evaluations. Eur J Med Chem 2008; 43:1180–1187.
18. Bechard P, Lacroix J, Poyet P, C-Gaudreault R. Synthesis and cytotoxic activity of new alkyl[3-(2-chloroethyl)ureido]benzene derivative. Eur J Med Chem 1994; 29:963–966.
19. Ling S, Xin Z, Zhong J, Jian-xin F. Synthesis, Structure, and Biological Activity of Novel 4,5-Disubstituted Thiazolyl Urea Derivatives. Hetero Chem 2008; 19:2–6.
20. Gonzalez-Díaz H, Saiz-Urra L, Molina R, Santana L, Uriarte E. A model for the recognition of protein kinases based on the entropy of 3D van der Waals interactions. J Proteome Res 2007; 6:904–908.
21. González-Díaz H, Saíz-Urra L, Molina R, González-Díaz Y, Sánchez-González A. Computational chemistry approach to protein kinase recognition using 3D stochastic van der Waals spectral moments. J Comput Chem 2007; 28:1042–1048.
22. Szántai-Kis C, Kövesdi I, Eros D, Bánhegyi P, Ullrich A, Kéri G, Orfi L. Prediction oriented QSAR modelling of EGFR inhibition. Curr Med Chem 2006; 13:277–287.
23. Shi WM, Shen Q, Kong W, Ye BX. QSAR analysis of tyrosine kinase inhibitor using modified ant colony optimization and multiple linear regression. Eur J Med Chem 2007; 42:81–86.
24. Peng T, Pei J, Zhou J. 3D-QSAR and receptor modeling of tyrosine kinase inhibitors with flexible atom receptor model (FLARM). J Chem Inf Comput Sci 2003; 43:298–303.
25. Kamath S, Buolamwini JK. Receptor-guided alignment-based comparative 3D-QSAR studies of benzylidene malonitrile tyrphostins as EGFR and HER-2 kinase inhibitors. J Med Chem 2003; 46:4657–4668.
26. Assefa H, Kamath S, Buolamwini JK. 3D-QSAR and docking studies on 4-anilinoquinazoline and 4-anilinoquinoline epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors. J Comput Aided Mol Des 2003; 17:475–493.
27. Hansch C, Leo A. Exploring QSAR. In: Heller SR. Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington, 1995
28. Zheng W, Tropsha A. Novel variable selection quantitative structure–property relationship approach based on the k-nearest-neighbor principle. J Chem Inf Comput Sci 2000; 40:185–194.
29. Ajmani S, Jadhav K, Kulkarni SA. Three-dimensional QSAR using the k-nearest neighbor method and its interpretation. J Chem Inf Model 2006; 46:24–31.
30. VLifeMDS; Molecular Design Suite version 3.5, V-life Sciences Technologies Pvt. Ltd., Pune, India, 2004. (www.vlifesciences.com).
31. Hudson BD, Rahr RM, Wood J, Osman J. Parameter Based Methods for Compound Selection from Chemical Databases. Quant Struct-Act Relat 1996; 15:285–289.
32. Baumann K. An alignment-independent versatile structure descriptor for QSAR and QSPR based on the distribution of molecular features. J Chem Inf Comput Sci 2002; 42:26–35.
33. Discovery Studio, version 2.1; Accelrys Inc.: San Diego, CA, USA.