

N-Phenyl Ureidobenzenesulfonate derivatives as novel anticancer agents: QSAR and Molecular docking studies

Azar Mostoufi^{1,2}, Fariba Aliyan¹, Masood Fereidoonzhad^{1,*}

¹Department of Medicinal Chemistry, School of Pharmacy, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.

²Research Center of Marine Pharmaceutical Science, Ahvaz Jundishahpour University of Medical Science, Ahvaz, Iran.

Abstract

DNA double strand-breaks (DSBs) are the most deleterious lesions that can affect the genome of living beings and are lethal if not quickly and properly repaired. Recently, N-phenyl ureidobenzenesulfonates (PUB-SOs) as tubulin inhibitors that block cell cycle progression in S-phase and induce DNA DSBs are discovered. Here, a set of PUB-SOs derivatives were applied to quantitative structural activity relationship (QSAR) analysis. A series of chemometric methods like MLR, FA- MLR, PCR and partial least squared included in variable selection genetic algorithm (GA-PLS), were used to relate structural features of these compounds with their anti-proliferative activity against MCF-7 cell line. New potent lead compounds were also designed based on new structural patterns using *in silico*-screening study. Molecular docking studies of these compounds on DNA and tubulin were conducted. The results obtained from validated docking protocols indicate that the main amino acids located in the active site cavity in charge of essential interactions with tubulin are Ala30, Lys B254, Asn B258, Met B259, Asn A101, Glu A183, Thr A179, Leu B255, Ser A178 and Gln B247 and the most important base pairs inside the minor groove of DNA responsible for essential intercalation with DNA are G2, G4, G10, G12, A5, A6, C9 and C11.

Keywords: *in silico*-screening, Molecular Docking, N-phenyl ureidobenzenesulfonates, QSAR

1. Introduction

Spontaneous DNA damage occurs frequently in living cells. It is estimated that the number of DNA lesions including base losses, single- and double-strand breaks (DSBs) can be close to 100,000 lesions per cell per day (1, 2). On one hand, cells exploit a variety of specialized DNA repair mechanisms to restore the integrity of the DNA. These DNA repair mechanisms collectively termed the DNA damage response are responsible to detect DNA damage and arrest cell cycle to repair DNA lesions (1, 3). Among all iden-

tified DNA damages, DSBs are one of the most cytotoxic lesions and the most difficult DNA lesion to repair (4).

DSBs are particularly troublesome because they can lead to cell death if not repaired. And, if not repaired correctly, can cause deletions, translocations, and fusions in the DNA. These consequences are collectively referred to as genomic rearrangements, and are commonly found in cancerous cells (5, 6).

In this context, the development of new anticancer agents inducing DNA DSBs is a promising strategy in cancer drug therapy. Recently a new class of anticancer agents referred to as N-phenyl ureidobenzenesulfonate derivatives (PUB-

Corresponding Author: Masood Fereidoonzhad, Department of Medicinal Chemistry, School of Pharmacy, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.
Email: Fereidoonzhad-m@ajums.ac.ir

SOs) were developed (1, 7). Beside apparent structure similarities between PUB-SOs and phenyl 4-(2-oxoimidazolidin-1-yl)-benzenesulfonates, their mechanisms of action are different and can vary from blocking the cell cycle progression in S-phase and inducing the phosphorylation of histone H2AX (gH2AX), which evidences the induction of DNA DSBs instead of the expected arrest of the cell cycle in the G2/M-phase, to inhibition of tubulin (8, 9).

In this work, for a set of PUB-SOs derivatives being able to inhibit tubulin and induce DNA DSBs, two different drug design methodologies has been applied: QSAR and molecular docking simulations. In a comprehensive study, to describe the physicochemical properties of the molecules, we used a very large descriptor set such as geometrical, functional, constitutional, topological groups, atom-centered fragments, 2D autocorrelation, quantum, electrostatic and chemical. Also, to model the relationship between the structural characteristics and anti-proliferative activity of the

studied compounds against MCF7 cancer cell line, different chemometrics methods were used:

- 1) multiple linear regression (MLR),
- 2) factor analysis-based multiple linear regression (FA-MLR),
- 3) principal component regression (PCR) and

4) partial least squared combined with genetic algorithm for variable selection (GA-PLS). Also, on all compounds of dataset and those designed, a validated molecular docking simulation technique was also applied to achieve their detailed molecular binding site in intercalating with base pairs of DNA and interacting with major amino acids in the active site of tubulin.

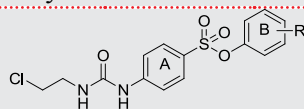
2. Materials and methods

2.1. Data set

A dataset comprised of 54 PUB-SOs derivatives as a set of potent dual inhibitor of tubulin and inducer of DNA DSBs are selected for this study (1). Table 1 includes structural features and

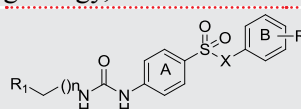
Table 1. Chemical structure of the compounds used in this study and their experimental and cross-validated predicted activity as well as their docking binding energy, on DNA and tubulin.

Name	R	MCF-7			
		Exp.pIC ₅₀	Pred.pIC ₅₀	ΔE^1 for DNA (kcal/mol)	ΔE^1 for Tubulin (kcal/mol)
1a	2-iProp	5.744727	5.710421	-9.31	-9.46
2a	2-OMe	4.721246	4.742597	-9.56	-8.14
3a	2-OEt	5.769551	5.908556	-8.69	-8.34
4a	2-F	5.221849	5.028002	-9.62	-8.65
5a	2-Cl	5.376751	5.300363	-8.92	-8.97
6a	3-I	5.821023	5.839865	-9.86	-9.47
7a	2-NO ₂	5.130768	5.218836	-7.04	-7.46
8a	2,4-Me	5.408935	5.313071	-9.51	-8.65
9a	2,5-Me	5.721246	5.758473	-9.48	-9.84



N-Phenyl ureidobenzenesulfonates (PUB-SOs)

SFOM-0004: 4-CEU, R=2-Me
 SFOM-0004: 4-CEU, R=3-Me
 SFOM-0004: 4-CEU, R=4-Me
 SFOM-0004: 4-CEU, R=4-OMe
 SFOM-0004: 4-CEU, R=4-N(Me)₂
 SFOM-0004: 4-CEU, R=4-OH



PUB-SOs substituted at position 3 and 4

X=O or NH
 R=2-Me, 2-Et, 2-Prop, 4-OH
 Series 1: n=1, R1= Cl (3 or 4-CEU)
 Series 2: n=2, R1= Cl (3 or 4-CPU)
 Series 3: n=1, R1= H (3 or 4-EU)
 SFOM-0016: 4-CEU, R=2-Et
 SFOM-0017: 4-CEU, R=2-Prop
 SFOM-0046: 4-EU, R=2-Et

10a	2,6-Me	5.431798	5.424931	-9.37	-9.35
11a	2,4-F	5.031517	5.093708	-8.94	-8.45
12a	2,6-F	5.318759	5.194877	-9.61	-8.35
13a	2,4,5-Me	5.522879	5.456428	-9.59	-9.21
14a	2,4,5-Cl	5.920819	5.850512	-10.51	-9.28
15a	2,4,6-Cl	5.769551	5.934567	-10	-9.12
16a	3-Et	5.148742	5.178438	-9.43	-8.97
17a	3-Prop	5.408935	5.296416	-10.09	-9.21
18a	3-OMe	5.376751	5.140660	-9.38	-9.05
19a	3-OEt	5.481486	5.411157	-10.08	-9.14
20a	3-F	5.107905	5.058320	-8.84	-8.29
21a	3-Cl	5.431798	5.348047	-9.53	-9.18
22a	3-Br	5.585027	5.520773	-9.61	-9.36
23a	3-I	5.744727	5.991039	-9.58	-9.33
24a	3-NO ₂	4.744727	4.914188	-7.95	-8.13
25a	3,5-Me	5.537602	5.662556	-9.91	-8.71
26a	3,4-OMe	4.823909	5.193080	-10.22	-8.57
27a	3,5-OMe	5.69897	5.726583	-9.84	-8.71
28a	3,4-F	5.05061	4.953434	-8.66	-8.29
29a	3,5-F	5.244125	5.276152	-8.7	-8.31
30a	3,5-Cl	6	5.893710	-10.32	-9.16
31a	3,5-Br	6.39794	6.243301	-9.75	-10.16
32a	3,4,5-Me	5.026872	5.106650	-10.52	-9.32
33a	3,4,5-OMe	6.180456	5.692793	-9.47	-8.46
34a	3,4,5-F	4.946922	5.061280	-8.83	-8.27
35a	4-Et	4.79588	4.724407	-9.5	-9.06
36a	4-Prop	4.886057	4.828144	-9.57	-9.56
37a	4-tertBut	4.769551	4.848530	-10.68	-9.08
38a	4-OEt	5.022276	5.084198	-9.39	-8.77
39a	4-OBu	5.148742	5.018063	-10.16	-8.69
40a	4-F	4.853872	4.996087	-9.1	-8.55
41a	4-Cl	5.136677	5.010193	-9.37	-8.82
42a	4-Br	5.026872	4.967217	-10.35	-9.24
43a	4-I	4.958607	5.021639	-10.57	-9.57
44a	4-CN	4.721246	4.629007	-9.49	-8.61
45a	H	4.721246	4.910881	-9.05	-8.46
SFOM-0004 (46a)	2-Me	5.21467	5.192087	-9.19	-8.85
SFOM-0005 (47a)	3-Me	5.142668	5.292942	-10.17	-9.01
SFOM-0006 (48a)	4-Me	4.721246	5.034783	-10.6	-8.94
SFOM-0007 (49a)	4-OMe	4.638272	4.823279	-9.07	-8.38
SFOM-0008 (50a)	4-N(Me) ₂	4.236572	4.317371	-9.55	-8.9
SFOM-0010 (51a)	4-OH	5.886057	5.088921	-9.17	-8.79
SFOM-0016 (52a)	2-Et	5.387216	5.366410	-10.28	-8.37
SFOM-0017 (53a)	2-Prop	5.677781	5.690676	-9.17	-9.11
SFOM-0046 (54a)	-	5.468521	5.274308	-8.73	-8.77

¹docking binding energy.

biological activity details of these compounds. The anti-proliferative activity against MCF7 cell line, as IC_{50} values, were used for the QSAR modeling studies.

2.2. Molecular descriptors

Using ChemBioDraw 12.0 software, two dimensional structures of the ligands were produced. Each ligand was optimized with different minimization methods including molecular mechanics (MM^+) and quantum based semi-empirical method (AM1) by means of an in house TCL script (10-12) using Hyperchem. To calculate a large number of molecular descriptors, Hyperchem, Gaussian 98 (13) and Dragon packages (14) were applied. Also, to calculate chemical parameters such as molecular volume (V), molecular surface area (SA), hydrophobicity (LogP), hydration energy (HE) and molecular polarizability (MP), Hyperchem software (Version 8, Hypercube Inc., Gainesville, FL, USA) was used. Similarly, Gaussian 98 software was applied to calculate the most positive and the negative net atomic charges, highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) energies, the average absolute atomic charge and molecular dipole moment. According to the equations developed by Thanilaivelan *et al.*, quantum chemical indices including hardness ($\eta=0.5(HOMO+LUMO)$); softness ($S=1/\eta$), electrophilicity ($\omega=\chi^2/2\eta$) and electronegativity ($\chi=-0.5(HOMO-LUMO)$) were calculated (15). Different topological, geometrical, charge, empirical and constitutional descriptors for each molecule and also 2D autocorrelations, aromaticity indices, atom-centered fragments and functional groups were calculated by Dragon.

2.3. Model development

In a data matrix with the number of molecules and descriptors as the number of rows and columns respectively, the calculated descriptors were illustrated. For producing QSAR equations, four different regression methods such as factor analysis as the data processing step for variable selection (FA-MLR), genetic algorithm-partial least squares (GA-PLS), principal component regression analysis (PCRA) and simple

multiple linear regression with stepwise variable selection (MLR), were used. These known procedures are well explained in our previous QSAR studies (16, 17).

To develop QSAR models, stepwise selection and elimination of variables was done by SPSS software (version 21; SPSS Inc., IBM, Chicago, IL, USA). To check the predictability, validity and robustness of the models, leave-one-out cross-validation procedure using MATLAB 2015 software (version 8.5; Math work Inc., Natick, MA, USA) was obtained.

FA-MLR was also conducted on the dataset. To reduce the number of variables and to detect structure in the relations between them, factor analysis (FA) was performed. Also, to identify the important predictor variables and to avoid collinearity among them, this data-processing step was used (18). Along with FA-MLR, PCRA, was also applied for database. Among X variables, collinearities as a distributing factor are not included in PCRA and also the number of variables was not more than the number of observations (19). Factor scores obtained from FA, are played the role of predictor variables. All descriptors in PCRA are important; and detecting the relevant descriptors is the end of factor analysis (17).

The PLS regression method was applied to the NIPALS-based algorithm and existed in the chemometrics toolbox of MATLAB software. Also, to obtain the desirable number of factors, according to Haaland and Thomas F-ratio criterion (20, 21), leave-one-out cross-validation procedure was applied. For PLS and GA modeling, MATLAB PLS toolbox was applied. All calculations were run on a core i7 personal computer (CPU at 8 MB) with Windows 10 operating system.

2.4. Variable importance in the projection (VIP)

To investigate the relative importance of the variable in the final model in GA-PLS method, variable important in objection (VIP) was applied (22). The importance of variables in PLS method are represented by VIP values. According to Erikson *et al.*, it is possible for X-variables (predictor variables) to be categorized in terms of their relevance in explaining y (predicted variable);.

then, $VIP > 1.0$ and $0.8 < VIP < 1.0$ are highly and moderately influential and $VIP < 0.8$ is less influential (23). This process were was done using XLSTAT 2017 software (24).

2.5. Model validation

To validate the regression equation, statistical parameters including correlation coefficient (R^2), root mean square error of cross-validation (RMSECV), leave-one-out cross-validation correlation coefficient (Q^2), and variance ratio (F) with certain degrees of freedom were applied. 20% of the molecules were selected as test set (prediction set) molecules to test the developed model performance. The predictive value of a QSAR model that has not been taken into account during the process of developing the model should be tested on an external set of data.

2.6. Applicability domain

Precise prediction ability of QSAR model for new compounds has been made it a widely used model (10, 17). It should be noted that no matter how significant and validated a QSAR may be, it cannot be expected to predict the modeled property for the entire space of chemicals reliably. Hence, the domain of application of QSAR before it is put into use for screening chemicals, must be defined and predictions should be considered reliable for only those chemicals that fall in this domain. The applicability domain is appraised by the leverage values for each compound of our dataset. A Williams plot (the plot of standardized residuals versus leverage values (h)) can then be used for an immediate and simple graphical detection of both the response outliers (Y outliers) and structurally influential chemicals (X outliers) in our model. The applicability domain for the graph is defined in a squared area within $\pm x$ (standard deviations) and leverage threshold h^* .

The certain features of the numerical value of leverage include 1) being greater than zero and 2) the lower the value, the higher is the confidence in the prediction. Value of 1 equals to very poor prediction and value of zero equals to perfect prediction that will not be reached. The other factor to analysis the results is warning leverage (h^*). The threshold h^* is centered on $3(k+1)/n$, with

k =the number of model parameter and n =number of training set (calibration set) compounds while $x=2$ or 3. Prediction of high leverage value ($h > h^*$) compounds may not be reliable. If a leverage is higher than warning leverage h^* , it means the predicted response is the consequence of substantial extrapolation of the model;. So, it is unreliable. In other another perspective, if being lower the compound leverage value is lower than the threshold one, the means possibility of agreement between the values observed and predicted for compounds is as high as for the calibration set of compounds (25, 26).

2.7. Docking procedure

Molecular docking werewas carried out using an in house batch script (DOCKFACE) (17, 27) of AutoDock 4.2, each ligand was subjected to MM⁺ and AM1 minimization methods using Hyperchem 8 package. With Gasteiger-Marsili procedure implanted in the AutoDock Tools program (28), the partial charges of atoms were calculated. Having non-polar hydrogens of compounds merged and the rotatable bonds assigned, the output structures were changed to PDBQT using MGLtools 1.5.6 (29).

The three dimensional crystal structure of DNA (PDB ID:1BNA) and Tubulin (PDB ID:3UT5) were retrieved from protein data bank (<http://www.rcsb.org/pdb/home/home.do>). All water molecules were removed, missing hydrogens were added and after determining the Kollman united atom charges, non-polar hydrogens were merged into their corresponding carbons using AutoDock Tools (28). As the final part of this process, desolvation parameters were assigned to each protein atom. Among the three different search algorithms performed by AutoDock 4.2 the commonly used Lamarckian Genetic Algorithm (LGA) was applied (30-32). Subsequently, the enzyme and DNA were converted to PDBQT using MGLTOOLS 1.5.6.

A maximum number of 2,500,000 energy evaluations, 150 population size, 27000 maximum generations, 100 run, a gene mutation rate of 0.02 and a crossover rate of 0.8 were used for Lamarckian GA. The grid maps of the receptors were calculated using AutoGrid tools of AutoDock

4.2. The size of grid includes both the active site and considerable proteins of the encircling surface. A grid box of 65×60×108 and 55×55×55 points in x, y and z directions was centered on the center of the ligand in the complex with a spacing of 0.375 Å for 1BNA and 3UT5, respectively. Number of points in x, y and z was 14.719, 20.979 and 8.824 and for 3UT5 was -19.934, 131.986 and 116.717 respectively. Using AutoDock Tools the grid and docking parameter files, gpf and dpf, was produced. With a root mean square deviation (RMSD) tolerance of 2Å, cluster analysis was done on the docked results. Co-crystal ligand within pdb file of Tubulin (3UT5) was observed by a viewer and treated as other ligands. All the docking protocols were done on validated structures with RMSD values below 2 Å.

According to docking results, ligand-receptor interactions were detected using AutoDock tools program (ADT, Version 1.5.6), VMD software (33) and PyMOL molecular graphics program (34).

3. Results and Discussion

In this paper, to calculate structural parameters affecting anti-proliferative activity against MCF7 cell line of PUB-SOs derivatives, a detailed QSAR study has been conducted. Four well known QSAR methods such as stepwise MLR, FA-MLR, PCRA, and GA-PLS were applied for modeling the relationship between the biological activity and molecular descriptors, in these compounds.

3.1. MLR modeling

To model MLR, different stepwise selection-based MLR analyses were run using pool of all calculated descriptors. Table 2 shows the summarized results of QSAR models. Correlation coefficient (r^2) matrix for the descriptors used in different MLR equations is shown in Table 3. Performance of MLR equations is impaired by collinear descriptors and prediction ability is decreased by such models. Collinear descriptors degrade the performance of MLR equations and such models have lowered prediction ability. The correlation

Table 2. The results of different QSAR models with different types of dependent variables.

Model	Eq.no.	MLR Equation	n ¹	R ² c	Q ²	Rmscv	Cvcv	F	R ² p
MLR	1	pIC ₅₀ =8.889 AT5p (±0.831)-39.686 MSD (±5.817)-21.630 PW ₃ (±4.864)-0.463 Softness (±0.171)-11.341 MATS6m(±2.379) +4.463 MATS5m(±2.458) +1.265 IC ₂ (±0.327)-0.438 HOMA (±0.179)-0.752 IC ₃ (±0.362)+21.591(±3.722)	43	0.94	0.80	0.14	2.61	54.0	0.87
FA-MLR	2	pIC ₅₀ =0.869 X4v (±0.133)+ 3.294 GATS1e (±0.372)+8.426 MATS1e (±1.343)+0.994 MATS7p (±0.374)-8.153 AT58v (±1.996)+0.007 G(Cl..Cl)(±0.002)-87.173 X1A (±13.071)-12.021 LP1 (±4.124)+ 0876 IC ₂ (±0.211)-1.819 MATS5e (±0.250)+77.747 PW ₅ (±10.037)+1.596 MATS6e (±0.310)+0.643 J3D (±0.272)-58.989 (±13.588)	43	0.95	0.79	0.24	4.25	52.6	0.86
PCRA	3	pIC ₅₀ =0.244 FAC4 (±0.034)+0.204 FAC1 (±0.034)+0.106 FAC15 (±0.034)+0.090 FAC13 (±0.034)-0.089 FAC7 (±0.034) +0.075 FAC8 (±0.034)+0.075 FAC3 (±0.034)+5.283 (±0.034)	43	0.72	0.64	0.27	5.03	17.2	0.61
GA-PLS	4	pIC ₅₀ =1.081X5v (±0.201)-0.012 L/BW (±0.002)-2.270 MATS2e (±0.298)-9.671 AT57e(±1.316)+2.808 MATS5p (±0.420)-0.007 piPC07(±0.001)-2.938 MATS5v (±0.631)+0.15 piPC05 (±0.004)+0.179 SEigZ (±0.059)+12.422(±1.281)	43	0.97	0.72	0.24	4.55	89.8	0.81

¹Number of molecules of training set used to derive the QSAR models. R²: Regression Coefficient for training set (calibration set). Q²: Regression Coefficient for Leave One Out Cross Validation. RMScv: Root Mean Square Error of cross validation. R²p: Regression Coefficient for prediction set (test set).

coefficient (r^2) matrix for the descriptors used in MLR equation 1, shows that no significant correlation exists between pairs of descriptors (Table 3).

The statistical parameters calculated for an obtained QSAR model such as R^2 , correlation coefficient (R^2_p) of prediction set, F at specified degrees of freedom, Q^2 , $C_{v_{cv}}$ and RMS_{cv} , having been used to validate the goodness of fit of the resulted QSAR equations, are represented in Table 2. The selected variables in table 2 demonstrate that Aromaticity (HOMA), topological (MSD, PW3, IC2, IC3), 2D autocorrelations (ATS5P, MATS5m, MATS6m), quantum (softness) descriptors affect the anti-proliferative activity of the studied compounds. Table 2 shows that none of the suggested QSAR models were obtained by chance and the best set of calculated descriptors was selected by MLR (equation 1) because of its greatest statistical parameters. Therefore the best predictive results were observed.

3.2. FA-MLR and PCRA

Seven factors of loading for compounds (after VARIMAX rotation) were shown in Table 4 upon their anti-proliferative activity (factor 1, 3, 4, 7, 8, 13 and 15). As it was observed in table 4, about 49% of variances in the original data matrix could be explained by the selected seven factors. Table 4 shows that descriptors including MW, ATS3m, ATS4m, ATS7v, ATS8v, ATS3p, ATS4p, ATS5p, GATS7m, GATS1p, GATS4p, SEigZ, nx, X4v, X5v and I-099 are highest loading values for factor 1 and the highest loading values for factor 3 are associated with Sv, Se, nSK, RBF, X4, X5, X0v, S2K, S3K, Lop, ICR, piPC08, MATS8m,

MATS2v, MATS4e, MAXDP, nCp and Eph whereas MATS7p, MATS8p, GATS8v, GATS7e and GATS8p are the descriptors of high loading of factor 8. Table 4 revealed that, factors 1 and 3 are moderately loaded with anti-proliferative activity. It is interesting to note that, factor 1 is with highest loadings from constiitutional (MW, nX), 2D autocorrelations (ATS3m, ATS4m, ATS7v, ATS8v, ATS3p, ATS4p, ATS5p, GATS7m, GATS1p, GATS4p), topological (SEigZ, X4v, X5v) and atom-centered fragments (I-099) while factor 3 is comprised of information about constiitutional (Sv, Se, nSK, RBF), topological (X4, X5, X0v, MAXDP, S2K, S3K, Lop, ICR, piPC08), 2D autocorrelations (MATS8m, MATS2v, MATS4e), functional (nCp) and quantum (Eph) descriptors. Table 2, equation 2 show the FA-MLR equation that has been made by highly loaded descriptors.

3.3. PCRA

Equation 3 is obtained from factor scores as predictor parameters in multiple regression equation via forward selection method (PCRA). Unlike selected descriptors, factor scores include information from different descriptors, thus the risk for data missing is reduced. By using the principle component method, seven (Table 4) factors scores were considered as independent parameters for developing QSAR equations. The variables used in Eq. 3 shows statistical quantities similar to those obtained by the FA-MLR method. But, it indicates partly higher calibration and lower cross-validation statistics with respect to Eq.2.

In Table 4, Factor score 1 signifies the importance of MW, ATS3m, ATS4m, ATS7v, ATS8v,

Table 3. Correlation coefficient (R^2) matrix for descriptors represented in multiple linear regression eqn 1.

	ATS5p	MSD	PW3	softness	MATS6m	MATS5m	IC2	IC3	HOMA
ATS5p	1	-0.130	-0.138	0.264	0.119	-0.105	0.071	0.179	-0.049
MSD		1	-0.282	-0.089	0.203	0.225	-0.114	0.167	-0.088
PW3			1	-0.110	0.088	-0.053	-0.045	-0.215	-0.007
softness				1	0.094	0.225	-0.231	-0.106	0.396
MATS6m					1	0.303	-0.202	-0.196	-0.062
MATS5m						1	-0.319	-0.327	0.000
IC2							1	0.290	0.030
IC3								1	0.078
HOMA									1

Table 4. Factor loadings of some significant descriptors after VARIMAX rotation.

Descriptor	F1	F3	F4	F7	F8	F13	F15	Communalities
HE	000	0.059	0.217	-0.570	-0.197	-0.171	-0.108	1.000
MW	0.873	0.359	-0.024	-0.019	0.009	-0.017	-0.007	1.000
Sv	-0.151	0.886	-0.018	-0.138	-0.040	-0.045	0.988	1.000
Se	-0.421	0.832	0.128	-0.015	-0.130	-0.023	-0.028	1.000
nSK	-0.098	0.916	0.160	0.155	-0.059	-0.016	-0.029	1.000
RBF	-0.299	0.810	0.108	0.055	0.005	0.114	0.005	1.000
nx	0.563	-0.206	-0.015	-0.258	0.028	-0.057	-0.036	1.000
MSD	-0.033	-0.244	-0.897	-0.058	-0.048	0.055	-0.015	1.000
jhetz	0.311	-0.074	0.822	0.055	0.125	0.013	-0.041	1.000
Jhetv	0.275	-0.067	0.664	0.007	-0.007	-0.315	-0.111	1.000
X4	-0.004	0.802	0.441	0.083	-0.081	0.026	-0.018	1.000
X5	-0.077	0.723	0.340	0.127	0.184	-0.096	0.054	1.000
X0v	0.480	0.828	0.111	-0.051	-0.086	-0.029	-0.026	1.000
X4v	0.652	0.512	0.327	-0.084	-0.129	-0.044	0.020	1.000
X5v	0.598	0.456	-0.375	-0.074	0.037	-0.112	0.004	1.000
S2K	0.118	0.853	0.015	0.036	0.010	0.050	-0.028	1.000
S3K	0.259	0.774	-0.349	-0.007	-0.092	0.140	0.054	1.000
PW3	-0.074	0.198	0.639	0.055	0.110	-0.246	-0.176	1.000
PW4	-0.024	0.273	0.815	-0.037	0.094	0.036	0.929	1.000
PW5	0.014	0.100	0.512	-0.055	0.486	-0.019	0.162	1.000
Lop	-0.215	-0.630	-0.202	0.140	-0.053	0.062	0.058	1.000
ICR	-0.135	0.629	-0.530	0.069	0.246	0.033	-0.213	1.000
IDDE	-0.052	-0.145	-0.049	-0.150	-0.063	-0.091	0.591	1.000
IVDE	0.199	0.150	0.247	-0.506	-0.070	-0.062	0.068	1.000
vindex	0.131	-0.526	0.774	-0.073	0.126	-0.050	0.026	1.000
SIC3	0.327	-0.647	-0.105	0.068	0.065	0.015	0.088	1.000
SEigZ	0.823	0.001	0.005	-0.003	0.084	0.034	0.015	1.000
piPC08	-0.076	0.826	0.029	0.236	-0.122	-0.061	-0.090	1.000
ATS3m	0.874	-0.212	0.063	-0.022	0.102	0.024	-0.018	1.000
ATS4m	0.857	-0.143	0.161	-0.046	-0.016	0.073	0.057	1.000
ATS7v	0.784	-0.370	-0.166	0.048	0.043	-0.054	-0.021	1.000
ATS8v	0.794	0.027	-0.172	-0.023	-0.354	-0.040	-0.134	1.000
ATS3p	0.890	-0.078	0.064	0.002	0.081	-0.069	0.0081	1.000
ATS4p	0.921	-0.113	0.169	-0.062	0.106	0.127	-0.120	1.000
ATS5p	0.836	-0.339	0.147	-0.030	-0.024	0.120	0.119	1.000
MATS5m	0.279	0.176	-0.538	-0.056	0.111	-0.100	-0.227	1.000
MATS8m	0.104	-0.571	0.227	0.084	0.570	0.045	-0.057	1.000
MATS2v	0.026	-0.644	0.287	-0.097	-0.050	-0.059	0.037	1.000
MATS1e	-0.198	0.043	0.061	0.535	-0.129	-0.090	-0.036	1.000
MATS3e	-0.313	0.123	0.388	-0.061	-0.165	-0.508	-0.133	1.000
MATS4e	-0.233	0.535	-0.207	-0.238	-0.100	-0.047	0.012	1.000
MATS7p	0.203	0.211	-0.067	-0.031	-0.739	0.004	-0.001	1.000
MATS8p	0.240	-0.165	-0.124	-0.024	0.859	0.005	0.008	1.000
GATS7m	0.542	0.037	-0.025	-0.099	0.039	-0.051	-0.103	1.000

Continued.

GATS8v	-0.048	-0.181	0.373	0.021	-0.713	0.050	0.104	1.000
GATS3e	0.239	0.012	-0.315	-0.187	-0.197	0.688	0.053	1.000
GATS7e	-0.270	0.188	0.317	-0.142	-0.545	-0.124	-0.157	1.000
GATS1p	-0.857	0.277	0.147	0.012	-0.045	0.044	0.021	1.000
GATS4p	-0.630	0.074	-0.044	0.032	0.177	0.109	0.178	1.000
GATS7p	-0.333	0.220	0.035	0.002	0.570	-0.050	-0.050	1.000
GATS8p	-0.058	0.324	-0.153	0.016	0.868	-0.018	-0.018	1.000
HOMA	-0.026	-0.046	0.064	0.842	0.019	0.038	0.066	1.000
J3D	-0.362	0.411	0.615	-0.045	-0.057	-0.105	-0.035	1.000
MAXDP	-0.102	0.544	0.509	-0.086	-0.105	-0.110	0.107	1.000
ASP	-0.128	0.025	-0.865	-0.032	0.144	0.072	0.000	1.000
L/Bw	-0.049	0.076	-0.833	-0.029	0.202	-0.091	-0.213	1.000
nCp	-0.408	0.619	0.267	-0.103	-0.115	-0.022	0.001	1.000
C-006	-0.016	0.382	-0.311	-0.067	-0.014	0.574	0.111	1.000
H-052	-0.185	0.192	0.068	-0.045	-0.037	0.698	-0.332	1.000
I-099	0.529	-0.126	-0.109	-0.003	0.025	-0.030	-0.086	1.000
Eph	-0.120	0.610	-0.200	-0.189	0.212	0.072	-0.093	1.000

ATS3p, ATS4p, ATS5p, GATS7m, GATS1p, GATS4p, SEigZ, nx, X4v, X5v and I-099 descriptors. Factor score 3 indicates the importance of Sv, Se, nSK, RBF, X4, X5, X0v, S2K, S3K, Lop, ICR, piPC08, MATS8m, MATS2v, MATS4e, MAXDP, nCp and Eph descriptors. Factor score 4 indicates the importance of MSD, jhetz, Jhetv, PW3, PW4, PW5, ICR, vindex, MATS5m, J3D, MAXDP and ASP. Factor score 7 indicates the importance of HE, IVDE and MATS1e descriptors. Factor score 8 signifies the importance of MATS7p, MATS8p, GATS8v, GATS7e and GATS8p descriptors. Factor score 13 indicates the importance of GATS3e, C-006 and H-052 descriptors. Factor score 15 signifies the importance of IDDE descriptor.

3.4. GA-PLS

The breakdown of data matrix for descriptors in PLS analysis is into orthogonal matrices having an inner relation between dependent and independent variables. In PLS analysis in contrast to MLR analysis, the multicollinearity problem for descriptors is removed. Modeling in PLS is consistent with noisy data better than MLR because a least number of latent variables are used in this type of analysis and also a lot of different GA-PLS runs have been performed via different initial sets of population. The statistical parameters calculat-

ed for this model are represented in Table 3.

As it was shown in Table 3, a combination of 2D autocorrelations (ATS7e, MATS2e, MATS5p and MATS5v), topological (SEigZ, X5v, piPC05, piPC07) and geometrical (L/BW) descriptors have been selected by GA-PLS to account for the anti-proliferative activity. In this table, Eq. 4 with high statistical quality parameters was obtained from the pool of calculated descriptors (i.e., $R^2=0.97$ and $Q^2=0.72$) and, the predictive R^2 value for the test set was found to be 0.81.

The optimum GA-PLS with best fitness includes 79 indices. The PLS estimate of coefficients for the descriptors are given in Figure 1. To calculate anti-proliferative activity of N-phenyl ureidobenzenesulfonates derivatives, a combined set of quantum, topological, Aromaticity and 2D-autocorrelations descriptors has been used by GA-PLS.

VIP was calculated for each descriptor to determine the importance of the 79 selected GA-PLS descriptors. Figure 2 show the VIP analysis of PLS equation. VIP shows that 23 descriptors with $VIP>1.0$ such as ATS5v, GATS1p, GATS5m, ATS6v, ATS1e, JhetZ, T(N..N), X5v, X5sol, X3sol, X1AV, X4, piPC06, piPC07, Vindex, HVcpX, SEigZ, L/BW, nBR, Br094, AMW, pol and MW are the most important indices in the QSAR equation

Table 5. Definitions of molecular descriptors present in the models.

Descriptors	Brief description
MATS5m	Moran autocorrelation of lag 5 weighted by mass 2D autocorrelations
MATS6m	Moran autocorrelation of lag 6 weighted by mass 2D autocorrelations
MATS1e	Moran autocorrelation of lag 1 weighted by Sanderson electronegativity
MATS2e	Moran autocorrelation of lag 2 weighted by Sanderson electronegativity
MATS5e	Moran autocorrelation of lag 5 weighted by Sanderson electronegativity
MATS6e	Moran autocorrelation of lag 5 weighted by Sanderson electronegativity
MATS5p	Moran autocorrelation of lag 5 weighted by polarizability
MATS7p	Moran autocorrelation of lag 7 weighted by polarizability
MATS5v	Moran autocorrelation of lag 5 weighted by van der Waals volume
ATS5p	Broto-Moreau autocorrelation of lag 5 (log function) weighted by polarizability
ATS7e	Broto-Moreau autocorrelation of lag 7 (log function) weighted by Sanderson electronegativity
ATS8v	Broto-Moreau autocorrelation of lag 8 (log function) weighted by van der Waals volume
GATS1e	Geary autocorrelation of lag 1 weighted by Sanderson electronegativity
MSD	mean square distance index (Balaban)
PW3	path/walk 3 - Randic shape index
PW5	path/walk 5 - Randic shape index
softness	softness
IC2	Information Content index (neighborhood symmetry of 2-order)
IC3	Information Content index (neighborhood symmetry of 3-order)
HOMA	Harmonic Oscillator Model of Aromaticity index
X2V	valence connectivity index of order 2
X4V	valence connectivity index of order 4
X5V	valence connectivity index of order 5
X1A	average connectivity index of order 1
G(Cl..Cl)	sum of geometrical distances between Cl..Cl
LP1	lovasz-pclikan index (leading eigenvalue) eigenvalue-based index
J3D	3D-Balaban index
L/BW	length-to-breadth ratio by WHIM
piPCO5	molecular multiple path count of order 5
piPCO7	molecular multiple path count of order 7
SEigZ	the eigenvalue sum from Z weighted distance matrix

derived by PLS analysis. In addition, $0.8 < VIP < 1.0$ is moderately influential and $VIP < 0.8$ is less influential.

Finally it should be noted that the brief description of the descriptors which were inserted in the QSAR models are listed in Table 5.

3.5. *In silico* screening

Instead of expensive and time-consuming *in vivo* experiments, it is possible to apply *in silico* screening in initial steps of drug development because it is able to accelerate the speed of discovery, anticipate and explore new pharmaceutical com-

pounds. To explore and detect active compounds among molecular databases, a strong technique, virtual screening, may be used. This technique is used through deletion, insertion and substitution of different substitutes on the parent molecules and is able to investigate the influences of the structural modifications on biological activity (10, 17).

For the use of the model in screening new compounds, the domain application of QSAR model was determined. The applicability domain (AD) of QSAR model was applied to verify the prediction reliability, to recognize the troublesome compounds and to predict the compounds with ac-

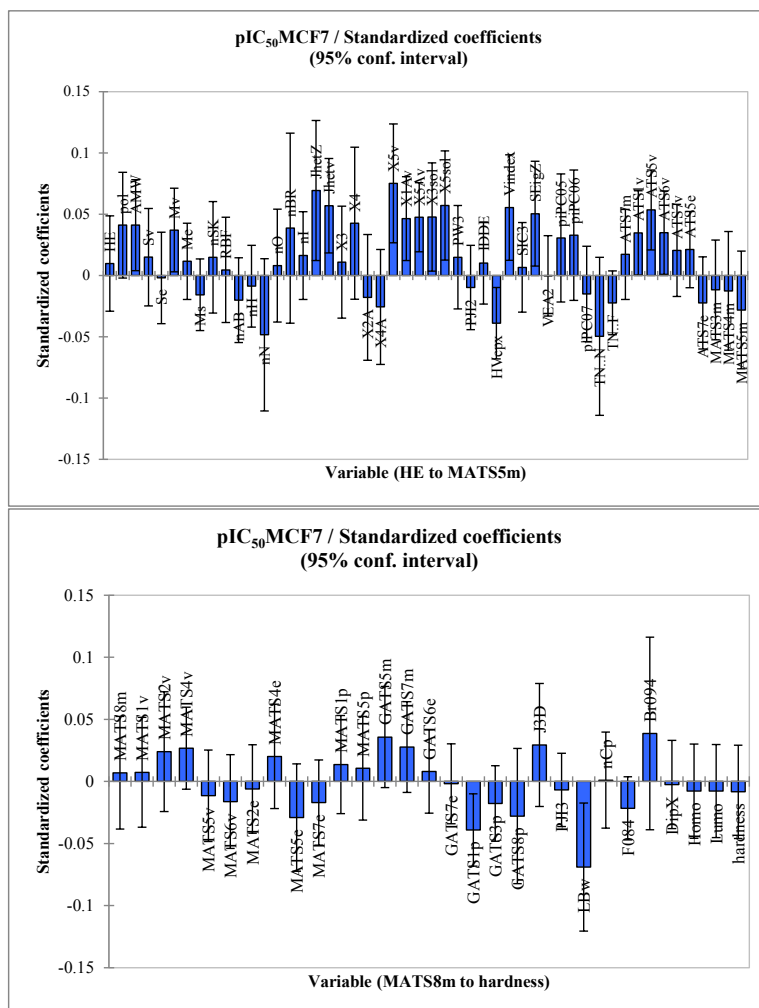


Figure 1. Regression coefficients for the variables used in GA-PLS model.

ceptable activity that falls within this domain. We employed the important descriptors selected from MLR model (as the best studied model because of its greatest statistical parameters) to design new active compounds.

According to analysis done on AD model in the Williams plot Of the MLR model (Figure 3) using the whole dataset, it was demonstrated that neither one of the compounds are an obvious outlier. As it was depicted in figure 3, none of the compounds have leverage (h) values greater than the threshold leverages (h^*). The warning leverage (h^*), was found to be 0.69. To the best of our knowledge, the compounds that had a standardized residual more than three times of the standard deviation units were considered to be outliers. For both the calibration set and test set of targets, the presented model matches the high quality parameters with good fitting power and the capability of

assessing external data. Moreover, almost all of the compounds were within the applicability domain of the proposed model and were evaluated accurately. While chemicals with a leverage value higher than h^* were considered to be influential or high leverage chemicals (20, 21).

To design new compounds with improved potential anti-proliferative activity against MCF7 cell line, in the studied QSAR model, the in silico screening should be used. Then, the in silico screening was applied by substituting diverse groups in different places. The results of this investigation are summarized in Table 6. As it was shown in Table 6, 45 novel compounds were designed and their predicted anti-proliferative activities based on MLR equation, as well as their docking binding energies on DNA and tubulin were obtained. Leverage values show that all of the designed compounds were within the applicability

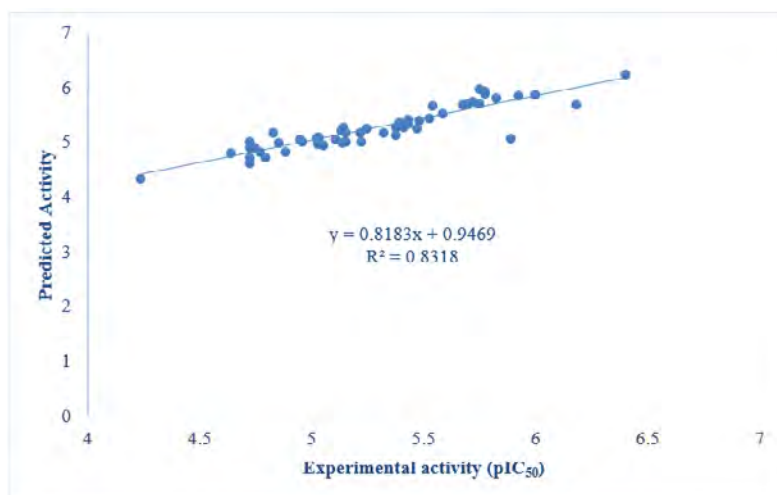


Figure 4. Plots of cross-validated predicted values of activity by MLR model against the experimental values.

poses of designed compounds are within the range of -7.72 to -10.62 Kcal.mol⁻¹ for DNA and -7.33 to -9.84 Kcal.mol⁻¹ for tubulin. Here, compound 27c shows great docking binding energies to both tubulin and DNA compared to the others.

Types of the interaction of these compounds to their targets were also investigated. As indicated in figure 5A, the NH group of ureid moiety of compound 7a is involved in hydrogen bond interaction with residue LysB254 in the active site of tubulin. The carbonyl group of urea interacts via acceptor hydrogen bonds with GlnB247 and AlaA181. A hydrogen bond interaction of the oxygen of the sulfonate group with GlnB247 and

AlaA181 amino acids also existed. The docked model suggests that the most energetically favorable conformation of the docked pose of 7a interacts with the base pairs in minor groove of 1BNA (figure 5B). It interacts via the NH group of the ureid moiety and the phenyl ring attached to the ureid moiety interacts with G10, the oxygen of the sulfonate group with G12 and the phenyl ring attached to the sulfonate moiety with C3 base pairs in the minor groove of DNA.

Types of interactions of compound 18a with both targets are depicted in figure 6. One of the NH groups of the ureid moiety of this compound is involved in hydrogen bond interaction

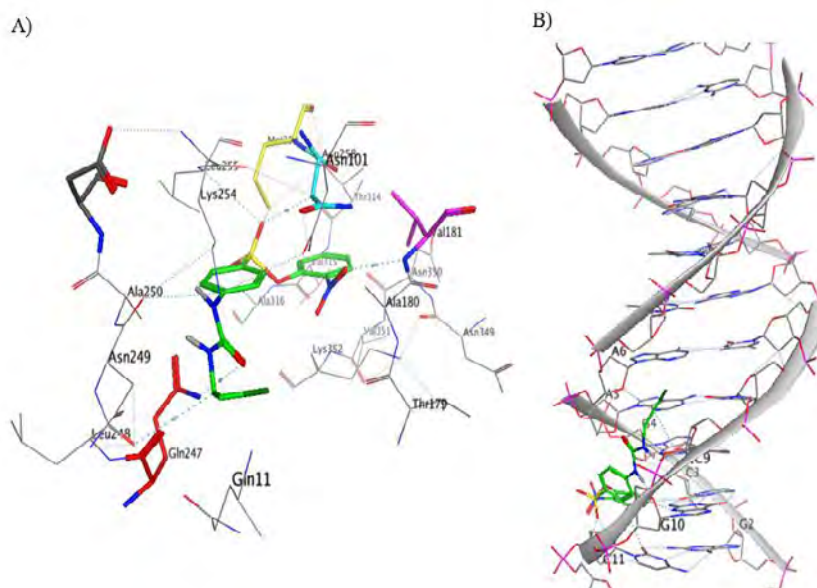


Figure 5. A) The structure of 7a surrounded by the key residues in the active site of tubulin. B) Molecular docking simulation studies of the interaction between 7a and 1BNA.

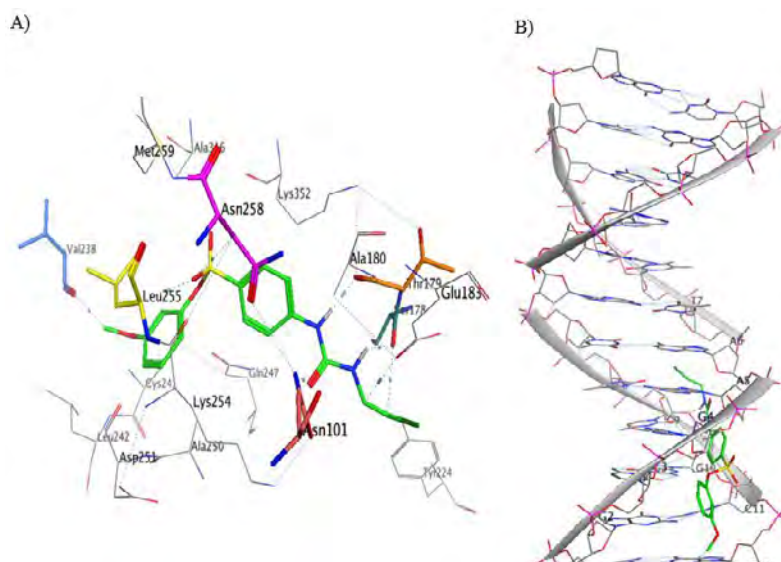


Figure 6. A) The structure of 18a surrounded by the key residues in the active site of tubulin. B) Molecular docking simulation studies of the interaction between 18a and 1BNA.

with residue SerA178 and the other NH group interacts with residue ThrA179 in the active site of tubulin. The carbonyl group of urea interacts via acceptor hydrogen bonds with AsnA101. The oxygen of sulfonate group and the methoxy group on the phenyl ring are involved in hydrogen bond interactions with residues LeuB255 and ValB238, respectively (figure 6A). This compound interacts with the base pairs in the minor groove of 1BNA (figure 6B). It interacts via one of the NH groups

A)

of the ureid moiety with G10 and the other NH groups with G9 base pairs. The phenyl ring attached to the sulfonate group and the oxygen of the sulfonate group interact with G12 and C11 base pairs in minor groove of DNA.

The interaction of the designed compounds based on *in silico* screening with both targets were also determined.

As indicated in figure 7A, the carbonyl group of the urea moiety of compound 38c inter-

B)

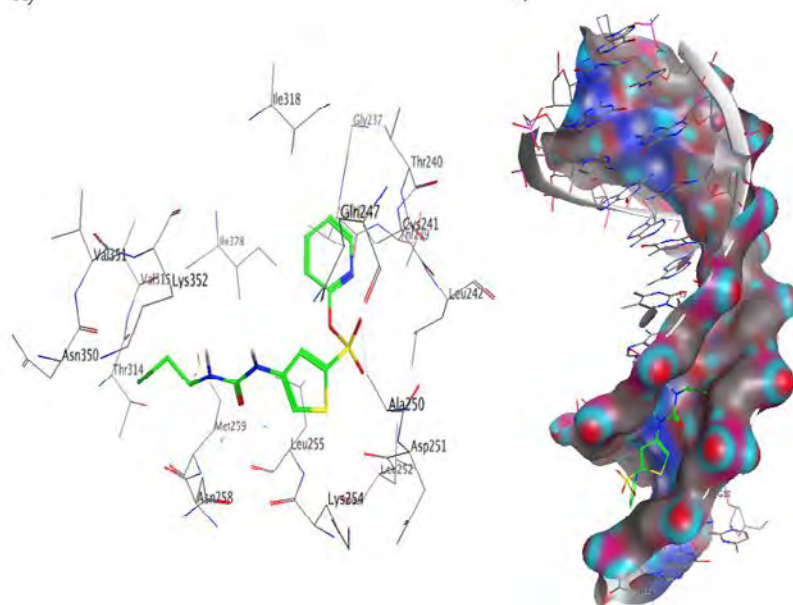
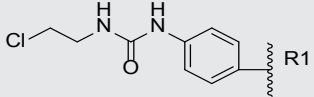
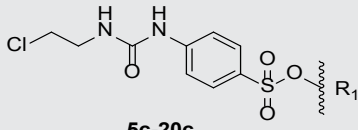
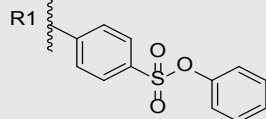
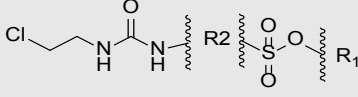
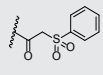
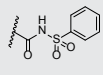
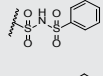
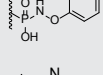
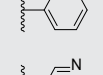
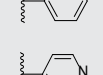
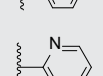
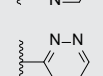
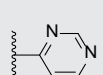
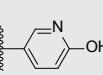
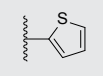
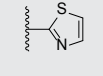
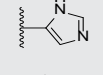
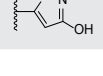
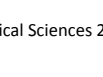


Figure 7. A) The structure of 38c surrounded by the key residues in the active site of tubulin. B) The structure of 38c in DNA minor groove.












Table 6. Structural modification of studied compounds and their predicted activities and docking binding energies for tubulin inhibitory and DNA binding based on MLR equation.

Name	R ¹	R ²	QSAR		Docking	
			pIC ₅₀ pred	leverage	ΔE (kcal/mol) DNA	ΔE (kcal/mol) Tu-bulin
1c-4c						
5c-20c						
21c-27c						
28c-32c						
1c		-	4.13218	0.104278	-9.58	-8.96
2c		-	4.9922779	0.1174002	-10.47	-8.73
3c		-	5.5665589	0.3069526	-9.33	-9.4
4c		-	4.9590493	0.1921305	-8.84	-7.41
5c		-	5.2363122	0.2563010	-8.58	-8.44
6c		-	5.0913227	0.3599163	-8.88	-8.46
7c		-	4.9712558	0.5123768	-8.92	-8.14
8c		-	5.2651320	0.5392542	-8.56	-8.2
9c		-	5.0426637	0.4657614	-8.95	-8.38
10c		-	5.1343357	0.5823411	-8.31	-7.95
11c		-	5.2468897	0.3428497	-8.77	-8.62
12c		-	5.5287544	0.3922224	-8.77	-8.43
13c		-	5.7314762	0.6115747	-8.71	-8.3
14c		-	5.0422809	0.4137238	-8.11	-7.67
15c		-	4.8795738	0.5603296	-9.55	-8.06

Continued.

16c		-	4.6522294	0.5012231	-8.43	-7.61
17c		-	5.6518219	0.2244990	-9.79	-9.45
18c		-	5.6522115	0.1923932	-10.52	-9.34
19c		-	5.5255279	0.3501535	-8.56	-8.28
20c		-	5.7150829	0.3098086	-9.39	-8.28
21c		-	5.1411216	0.4650149	-8.14	-8.83
22c		-	5.0913299	0.3380399	-8.52	-8.36
23c		-	5.0992968	0.3408480	-8.09	-8.48
24c		-	4.9823906	0.2774701	-8.37	-9.36
25c		-	5.1827703	0.3006536	-10.41	-9.22
26c		-	5.84488271	0.3309945	-10.09	-9.48
27c		-	5.3610461	0.3291805	-10.62	-9.84
28c			4.9494058	0.4360185	-8.38	-7.86
29c			5.0980903	0.3304486	-8.49	-8.32
30c			4.8277382	0.5560025	-9.33	-8.07
31c			4.8873655	0.4892056	-7.72	-7.51
32c			5.0397236	0.6520727	-8.24	-7.71
33c			5.0388827	0.6354044	-8.72	-8.18
34c			5.7113154	0.5350038	-8.54	-8.13
35c			5.8864446	0.6815654	-8.34	-7.96
36c			5.3788083	0.3893294	-8.3	-7.67
37c			5.0365843	0.6528003	-8.44	-8.53
38c			6.4691508	0.5279159	-8.33	-8.15

Continued.

39c			6.3150507	0.4960081	-8.51	-8.71
40c			6.3813354	0.6069968	-9.16	-8.29
41c			6.3260508	0.5245212	-8.47	-7.7
42c			6.1913096	0.6489881	-8.65	-7.35
43c			5.7160477	0.4960274	-8.25	-7.33
44c			5.9852362	0.4678657	-8.48	-7.87
45c			6.1021267	0.6130011	-9.11	-7.85

acts with residues Leu255 and Asp258 in the active site of tubulin. There existed a hydrogen bond interaction between the oxygen of the sulfonate group with residue Cys241. Its thiophene ring via its sulfur group interacts with residue Asp251. The docked model suggests that the most energetically favorable conformation of the docked pose of this compound interacts with the base pairs in minor groove of DNA (figure 7B). It interacts via one of the NH groups of its ureid moiety with A5 and the other NH group through arene-hydrogen bond with A6 base pairs. One of the oxygen groups of sulfonate moiety interacts with C11 and the other with A5 base pairs. The sulfur group of thiophene ring interacts with C9 and G10 base pairs in the minor groove of 1BNA.

As depicted in figure 8A, the NH group of ureid moiety of compound 45c is involved in hydrogen bond interaction with residue Gln247 in the active site of tubulin. The carbonyl group of urea moiety of this compound interacts with residue lys254. There existed a hydrogen bond interaction between the oxygen of sulfonate group with residues Leu255 and Met259. The pyrole ring of this compound via its nitrogen group interacts with residue Asp258. On the other hand, the docked pose of this compound interacts with the base pairs in minor groove of DNA (figure 8B). It interacts via the NH group of its ureid moiety with A5, the carbonyl group of urea moiety with G2,

one of the oxygen groups of the sulfonate moiety with G10 and the other with A6 base pairs. The nitrogen group of pyrole ring interacts with C11 and the sulfur group of thiazole ring interacts with A5 and T8 base pairs in the minor groove of DNA.

The results obtained from this docking study indicate that the important amino acids inside the active site cavity that are in charge of essential interactions with tubulin are Ala30, Lys B254, Asn B258, Met B259, Asn A101, Glu A183, Thr A179, Leu B255, Ser A178 and Gln B247. And the most important base pairs inside the minor groove of DNA being responsible for essential intercalations with DNA are G2, G4, G10, G12, A5, A6, C9 and C11 base pairs.

The application of relative operating characteristic curve (ROC) as a helpful metric tool to weigh the validity of docking procedures was first reported by Triballeau et al in computational medicinal chemistry (35). Nowadays, it was widely used as a validating procedure (36). First of all, about 106 DNA intercalator and 74 tubulin inhibitors were retrieved from ChEMBL database as SMILES format (37-39). The structures based on their experimental activities are categorized into two subsets of active ligands and inactive decoys. 26 ligands and 80 decoys for DNA and 20 ligands and 54 decoys for tubulin were generated. Subsequently, through a shell script using openbabel 2.3.2, the primary 3D generation of the structures

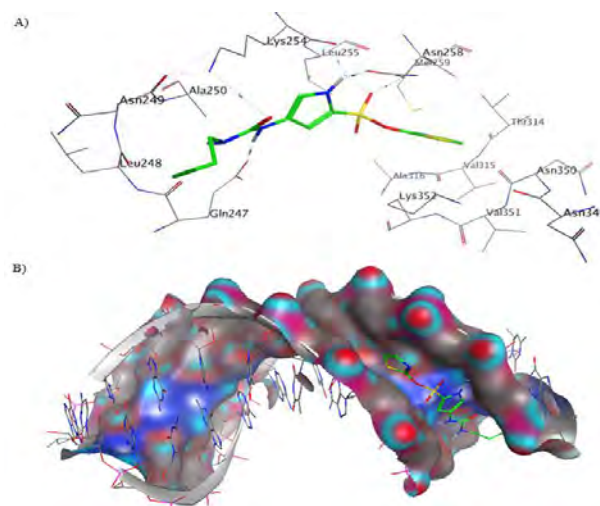


Figure 8. A) The structure of 45c surrounded by the key residues in the active site of tubulin. B) Molecular docking simulation studies of the interaction between 45c and 1BNA.

as mol2 format was provided (40). For all structures the ionization states at PH=7 were also calculated. Using batch scripting in windows operating system, the shell script was obtained. This screening method should be able to discriminate between active ligands and inactive decoys. ROC value is the area under the curve (AUC) for the plot of the true positive rate (TPR or sensitivity) against the false positive rate (FPR or 1- specificity) at various threshold settings. The ROC curve is thus the sensitivity as a function of 1- specificity. The AUC for ROC is calculated by trapezoidal integration

method as implemented in our in house ROC application (36). The more ROCAUC value means that the docking protocol is more able to discriminate between ligands and decoys. As it was shown in figure 9, the AUC of 0.776 for DNA and 0.812 for tubulin shows that our applied docking protocol was a validated protocol.

On the other hand, to evaluate the efficiency and quality of docking protocol with another tool, Enrichment Factor (EFmax) was used. Its calculations were based on the Li et al. work (41). EFmax factor in comparison to ROC curves,

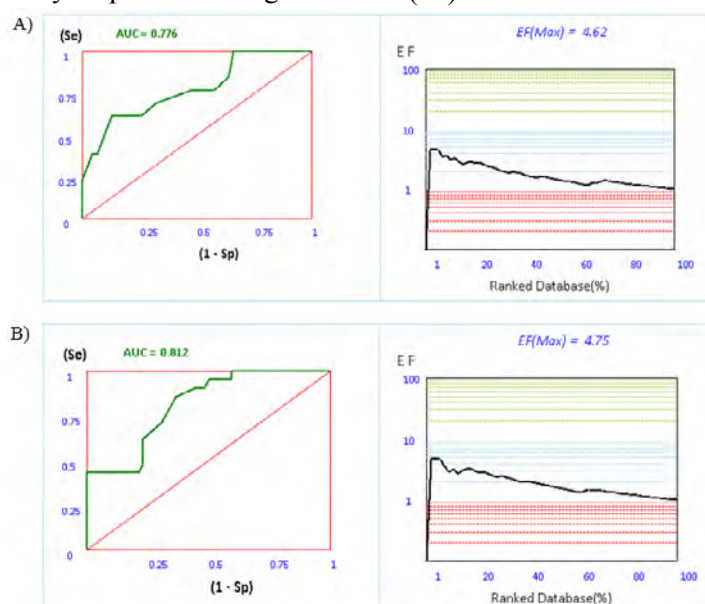


Figure 9. A) ROC and EF diagrams for docking with DNA B) ROC and EF diagrams for docking with tubulin.

is highly dependent on the number of actives in a data set (35). It means that early enrichment can be easily obtained if the number of active ligands is increasing in a dataset. Hence, the ROC should be considered most importantly.

4. Conclusion

Here, using a series of chemometric methods such as MLR, GA-PLS, FA-MLR and PCRA, quantitative relationships between molecular structure and anti-proliferative activity against MCF7 cell line for a set of N-phenyl ureidobenzenesulfonates (PUB-SOs) derivatives were obtained. Via different criteria such as cross-validation, validation through Y-randomization and root mean square error of cross validation (RM-SECV), the reliability, accuracy and predictability of the proposed models were assessed. The role of charge, geometrical, topological, 2D autocorrelations and quantum descriptors on anti-proliferative activity were acquired. A comparison between the different statistical methods employed, indicated that MLR represented superior results. According

5. References

- Gagné-Boulet M, Moussa H, Lacroix J, Côté M-F, Masson J-Y, Fortin S. Synthesis and biological evaluation of novel N-phenyl ureidobenzenesulfonate derivatives as potential anticancer agents. Part 2. Modulation of the ring B. *Eur J Med Chem.* 2015;103:563-73.
- Hoeijmakers JH. DNA damage, aging, and cancer. *N Engl J Med.* 2009;361:1475-85.
- Jackson SP, Bartek J. The DNA-damage response in human biology and disease. *Nature.* 2009;461:1071-8.
- Biedermann KA, Sun J, Giaccia AJ, Tosto LM, Brown JM. scid mutation in mice confers hypersensitivity to ionizing radiation and a deficiency in DNA double-strand break repair. *Proc Natl Acad Sci U S A.* 1991;88:1394-7.
- Aplan PD. Causes of oncogenic chromosomal translocation. *Trends Genet.* 2006;22:46-55.
- Negritto M. Repairing Double-Strand DNA Breaks. *Nature Education.* 2010;3:26
- Turcotte V, Fortin Sb, Vevey F, Coulombe Y, Lacroix J, Côté M-F, *et al.* Synthesis, biological evaluation, and structure-activity relationships of

to the developed QSAR model, *in silico* screening was applied and new compounds such as 3c, 12c, 17-20c, 26c, 34c, 35c and 38-45c with potential anti-proliferative activity were suggested for synthesis. Molecular docking simulations were also performed on these compounds to elucidate their interactions and to gain some insight into their molecular binding mode with DNA and tubulin as their targets.

Acknowledgements

The authors would like to thank the research deputy of Ahvaz Jundishapur University of medical sciences who supported this work. Collaboration of the medicinal chemistry department, faculty of pharmacy, Ahvaz Jundishapur University of Medical Sciences, in providing the required facilities for this work is greatly acknowledged. This paper is a part of thesis by Fariba Aliyan.

Conflict of Interest

None declared.

novel substituted N-phenyl ureidobenzenesulfonate derivatives blocking cell cycle progression in S-phase and inducing DNA double-strand breaks. *J Med Chem.* 2012;55:6194-208.

8. Fortin S, Bouchon B, Chambon C, Lacroix J, Moreau E, Chezal J-M, *et al.* Characterization of the Covalent Binding of N-Phenyl-N'-(2-chloroethyl) ureas to β -Tubulin: Importance of Glu198 in Microtubule Stability. *J Pharmacol Exp Ther.* 2011;336:460-7.

9. Fortin S, Wei L, Moreau E, Lacroix J, Côté M-F, Petitclerc É, *et al.* Design, synthesis, biological evaluation, and structure-activity relationships of substituted phenyl 4-(2-oxoimidazolidin-1-yl) benzenesulfonates as new tubulin inhibitors mimicking combretastatin A-4. *J Med Chem.* 2011 ;54:4559-80.

10. Fereidoonzhad M, Faghieh Z, Jokar E, Mojaddami A, Rezaei Z, Khoshneviszadeh M. QSAR, Molecular Docking and protein ligand interaction fingerprint studies of N-phenyl dichloroacetamide derivatives as anticancer agents. *Trends in Pharmaceutical Sciences.* 2016;2:159-76.

11. Fereidoonzhad M, Faghieh Z, Mojaddami A, Sakhteman A, Rezaei Z. A Comparative

Docking Studies of Dichloroacetate Analogues on Four Isozymes of Pyruvate Dehydrogenase Kinase in Humans. *Indian J Pharm Edu Res.* 2016;50:S32-S8.

12. Fereidoonzhad M, Faghieh Z, Mojaddami A, Tabaei S, Rezaei Z. Novel Approach Synthesis, Molecular Docking and Cytotoxic Activity Evaluation of N-phenyl-2, 2-dichloroacetamide Derivatives as Anticancer Agents. *J Sci I R Iran* 2016;27:39-49.

13. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, *et al.* Gaussian 09. Wallingford, CT, USA: Gaussian, Inc.; 2009.

14. A. Mauri VC, M. Pavan, R. Todeschini: D., RAGON Software: An Easy Approach to Molecular Descriptor Calculations. *MATCH Commun Math Comput Chem.* 2006;56:237-48.

15. Thanikaivelan P, Subramanian V, Raghava Rao J, Unni Nair B. Application of quantum chemical descriptor in quantitative structure activity and structure property relationship. *Chem Phys Lett.* 2000;323:59-70.

16. Khoshneviszadeh M, Faghieh Z, Jokar E, Mojaddami A, Rezaei Z. QSAR, Molecular Docking and protein ligand interaction fingerprint studies of N-phenyl dichloroacetamide derivatives as anticancer agents. *Trends in Pharmaceutical Sciences.* 2016;2(2).

17. Zare S, Fereidoonzhad M, Afshar D, Ramezani Z. A comparative QSAR analysis and molecular docking studies of phenyl piperidine derivatives as potent dual NK 1 R antagonists/serotonin transporter (SERT) inhibitors. *Computational Biology and Chemistry.* 2017;67:22-37.

18. Franke R, Gruska A, van de Waterbeemd H. Chemometrics Methods in molecular design. *Methods and Principles in Medicinal Chemistry.* 1995;2:113-9.

19. Bhattacharya P, Roy K. QSAR of adenosine A 3 receptor antagonist 1, 2, 4-triazolo [4, 3-a] quinoxalin-1-one derivatives using chemometric tools. *Bioorg Med Chem Lett.* 2005;15:3737-43.

20. Asadollahi T, Dadfarnia S, Shabani AMH, Ghasemi JB, Sarkhosh M. QSAR Models for CXCR2 Receptor Antagonists Based on the Genetic Algorithm for Data Preprocessing Prior to Application of the PLS Linear Regression Method and Design of the New Compounds Using *In Silico* Virtual Screening. *Molecules.* 2011;16:1928-55.

21. Khoshneviszadeh M, Edraki N, Miri R, Foroumadi A, Hemmateenejad B. QSAR Study of 4-Aryl-4H-Chromenes as a New Series of Apoptosis Inducers Using Different Chemometric Tools. *Chem Biol Drug Des.* 2012;79:442-58.

22. Olah M, Bologna C, Oprea TI. An automated PLS search for biologically relevant QSAR descriptors. *Comput Aided Mol Des.* 2004;18:437-49.

23. Bishop JM. The molecular genetics of cancer. *Science.* 1987;235:305-11.

24. Addinsoft S. XLstat 2012: Leading data analysis and statistical solution for microsoft excel. Addinsoft SRL. 2012.

25. Roy K, Kar S, Das RN. Chapter 7 - Validation of QSAR Models. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment.* Boston: Academic Press; 2015.p.231-89.

26. Weaver S, Gleeson MP. The importance of the domain of applicability in QSAR modeling. *J Mol Graphics Modell.* 2008;26:1315-26.

27. Mojaddami A, Sakhteman A, Fereidoonzhad M, Faghieh Z, Najdian A, Khabnadideh S, *et al.* Binding mode of triazole derivatives as aromatase inhibitors based on docking, protein ligand interaction fingerprinting, and molecular dynamics simulation studies. *Res Pharm Sci.* 2017;12:21-30.

28. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron.* 1980;36:3219-28.

29. Morris GM, Huey R, Olson AJ. Using AutoDock for Ligand-Receptor Docking. *Current Protocols in Bioinformatics:* John Wiley & Sons, Inc.; 2002.

30. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, *et al.* Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem.* 1998;19:1639-62.

31. Hamed A, Khoshnoud MJ, Tanideh N, Abbasi F, Fereidoonzhad M, Mehrabani D. Reproductive toxicity of Cassia absus seeds in female rats: possible progesteronic properties of chaksine and b-sitosterol. *Pharm Chem J.* 2015;49:268-74.

32. Faghieh Z, Fereidoonzhad M, Tabaei SMH, Rezaei Z, Zolghadr AR. The binding of small carbazole derivative (P7C3) to protofibrils

of the Alzheimer's disease and β -secretase: Molecular dynamics simulation studies. *Chem Phys*. 2015;459:31-9.

33. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J M Graph*. 1996;14:33-8, 27-8.

34. DeLano WL. PyMOL. DeLano Scientific, San Carlos, CA. 2002;700.

35. Triballeau N, Acher F, Brabet I, Pin JP, Bertrand HO. Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J Med Chem*. 2005;48:2534-47.

36. Rezaei Z, Fereidoonzhad M, Faghieh Z, Sadeghpur H, Mojaddami A, Sakhteman A. Comparison of docking procedures and its efficiency for Betasecretase, Aromatase and Pyruvate dehydrogenase kinase inhibitors. *Trends in Pharmaceutical Sciences*. 2017;3(1):31-42.

37. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, *et al*. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res*. 2012;40(Database issue):D1100-7.

38. Wassermann AM, Bajorath J. Binding-DB and ChEMBL: online compound databases for drug discovery. *Expert Opin Drug Discov*. 2011;6:683-7.

39. Willighagen EL, Waagmeester A, Spjuth O, Ansell P, Williams AJ, Tkachenko V, *et al*. The ChEMBL database as linked open data. *J Cheminform*. 2013;5:23.

40. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. *J Cheminform*. 2011;3:33.

41. Li H, Zhang H, Zheng M, Luo J, Kang L, Liu X, *et al*. An effective docking strategy for virtual screening based on multi-objective optimization algorithm. *BMC bioinformatics*. 2009;10:58.

