Asian Journal of Chemistry; Vol. 23, No. 11 (2011), 4968-4972

ASIAN JOURNAL OF CHEMISTRY

www.asianjournalofchemistry.co.in

Quantitative Structure Activity Relationship Study on the Inhibitory Activity of Antitumor Benzothiazole Derivatives Against MCF-7 Breast Cancer Cell Line

BRIJ K. SHARMA[1,*] and YASHWANT[2,*]

[1]Department of Chemistry, S.K. Government College, Sikar-332 001, India
[2]School of Pharmaceutical Sciences, Jaipur National University, Jagatpura, Jaipur-302 025, India

*Corresponding authors: Fax: +91 1572 270039; Tel: +91 1572 251039; E-mail: bksharma_sikar@rediffmail.com; yash99yash@gmail.com

(*Received*: 21 December 2010; *Accepted*: 20 July 2011) AJC-10190

The MCF-7 breast cancer cell line inhibitory activity of benzothiazole derivatives has been quantitatively analyzed in terms of Dragon descriptors using CP-MLR. The analysis has provided a rational approach for the development of new benzothiazole derivatives as MCF-7 breast cancer cell line inhibitors. The descriptors identified in CP-MLR analysis have highlighted the role of atomic properties in respective lag of 2D-autocorrelations (MATS5p) and modified burden eigenvalues (BELe3, BELe6, BEHm3 and BELp2), maximal electrotopological positive variation (MAXDP), information content index of 2nd order neighborhood symmetry (IC2) and the 9th order Galvez topological charge (GGI9) to explain the MCF-7 breast cancer cell line inhibitory actions of benzothiazole derivatives. Thus, the descriptors identified for rationalizing the activity give avenues to modulate the structure to a desirable biological end point.

Key Words: QSAR, Benzothiazole derivatives, MCF-7 breast cancer cell line, Combinatorial protocol in multiple linear regression.

INTRODUCTION

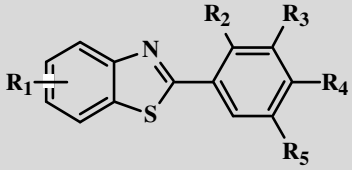
The structural similarities and the crystallographic analysis[1] of 5,6-dimethoxy-2-(4-methoxyphenyl)benzothiazole with bioactive flavones *i.e.*, quercetin and isoflavone genistein led to the supposition that the planar polyhydroxylated 2-phenylbenzothiazoles might mimic the ATP antagonistic effects of the natural products and have tyrosine kinase inhibitory properties[2]. The potent ligands for the arylhydrocarbon receptor (AhR) are the DF 203 and 5F 203 which translocate with the drug to cell nuclei[3]. The induced cytochrome P450 CYP1A1 subsequently leads to the generation of a reactive chemical intermediate(s). These intermediate(s) selectively generates DNA adducts only in sensitive tumor types such as mammary and ovarian tumor cell lines[4]. The prototype of quinols series, AW464, has also shown potent antitumor activity against renal and colon cancer cell lines that affects cell-signaling events downstream of the redox regulatory protein thioredoxin[5]. A new series of 2-phenylbenzothiazoles having oxygenated substituents in the phenyl moiety has been reported[6]. This series was exploited from the chemistry-driven approach to anticancer drug discovery[7]. In view of the importance of reported benzothiazole derivatives as antitumor agents, a quantitative structure-activity relationship is attempted on the MCF-7 breast cancer cell line inhibitory activity of these benzothiazole derivatives. The present study is aimed at rationalizing the substituent variations of these analogues to provide insight for the future endeavors.

EXPERIMENTAL

Data set: The reported 23 benzothiazole derivatives are considered as data set for the present study[6]. The structural variations of these analogues are given in Table-1. The biological actions of these compounds for MCF-7 breast cancer cell line were reported in terms of inhibitory activity (GI_{50}). These inhibitory concentrations were converted to molar scale (pGI_{50}) and are used in QSAR analysis. For the purpose of modeling study all 23 analogues have been divided into training and test sets. Out of the 23 analogues, one fourth compounds (**5**) have been placed in the test set for the validation of derived models. The training and test set compounds are also listed in Table-1.

Theoretical molecular descriptors: The structures of the compounds under study have been drawn in 2D ChemDraw[8]. The drawn structures were then converted into 3D modules using the default conversion procedure implemented in the CS Chem3D Ultra. The energy of these 3D-structures was minimized in the MOPAC module using the AM1 procedure for closed shell systems. This will ensure a well defined conformer relationship among the compounds of the study. All these energy minimized structures of respective compounds have been ported to DRAGON software[9] for the computation

TABLE-1
STRUCTURES, OBSERVED AND MODELED pGI_{50} VALUES OF THE BENZOTHIAZOLE DERIVATIVES



Compound	R_1	R_2	R_3	R_4	R_5	pGI_{50}[a]				
						Obsd.[b]	Eqn. 1	Eqn. 2	Eqn. 3	Eqn. 4
1[c]	H	H	OH	OH	H	5.66	4.82	4.66	4.83	5.50
2	H	H	O-CH_2-O		H	4.24	3.70	3.98	3.78	3.89
3[c]	H	H	OMe	OMe	H	4.28	5.14	5.19	5.54	5.33
4[c]	5-F	H	H	OMe	H	6.31	6.33	5.86	5.86	6.93
5	5-F	H	Me	OMe	H	7.32	6.92	6.97	7.10	6.76
6	5-F	H	F	OMe	H	6.06	6.40	5.74	5.81	6.57
7[c]	5-F	H	Cl	OMe	H	6.12	6.24	6.79	6.20	6.01
8	5-F	H	Br	OMe	H	6.07	6.30	6.08	6.37	6.04
9	5-F	H	I	OMe	H	6.09	6.34	6.12	6.13	6.06
10	5-F	H	OH	OH	H	6.14	5.37	5.62	5.72	5.90
11	5-F	H	O-CH_2-O	H	4.68	5.17	5.58	5.18	5.50	
12	5-F	H	OH	OMe	H	6.3	6.34	6.14	6.27	5.82
13	5-F	H	OMe	OH	H	6.12	5.72	6.80	6.73	5.53
14	5-F	H	OMe	OMe	H	-[d]	6.55	6.74	6.92	7.11
15	5-F	H	OMe	H	OMe	5.91	6.36	6.33	6.54	6.95
16	5-F	H	OMe	OMe	OMe	8.72	8.50	7.79	7.86	7.85
17[c]	5-F	OMe	OMe	OMe	H	7.64	8.89	7.13	7.36	7.36
18	5-F	H	OMe	OMe	OH	6.19	6.60	7.01	7.20	6.52
19	5-F	H	OMe	MOM[e]	H	7.19	7.39	7.59	7.50	6.75
20	5-F	H	OEt	OMe	H	7.1	7.90	7.48	7.39	7.80
21	5-F	H	OEt	OEt	H	9.15	8.24	8.75	8.34	9.01
22	5-Cl	H	OMe	OMe	H	4.35	4.95	4.17	4.18	4.78
23	6-F	H	OMe	OMe	H	7.21	6.63	6.68	6.76	7.10

[a]On molar basis; [b]Taken from reference [6]; [c]Compound included in test set; [d]Uncertain activity; [e]Methoxymethyleneoxy.

of descriptors for benzothiazole derivatives (Table-1). This software offers several hundreds of descriptors from different perspectives corresponding to 0D-, 1D- and 2D-descriptor modules. The outlined modules comprised of 10 different classes, namely, the constitutional (CONST), the topological (TOPO), the molecular walk counts (MWC), the BCUT descriptors (BCUT), the Galvez topological charge indices (GALVEZ), the 2D autocorrelations (2D-AUTO), the functional groups (FUNC), the atom-centered fragments (ACF), the empirical descriptors (EMP) and the properties describing descriptors (PROP). For each of these classes the DRAGON software computes a large number of descriptors which are characteristic to the molecules under multi-descriptor environment. The definition and scope of these descriptor's classes is given in Table-2. The combinatorial protocol in multiple linear regression (CP-MLR)[10] procedure has been used in the present work for developing QSAR models. Before the application of CP-MLR procedure, all those descriptors are intercorrelated beyond 0.90 and showing a correlation of less than 0.1 with the biological endpoints (descriptor *versus* activity, $r < 0.1$)

TABLE-2
DESCRIPTOR CLASSES[a] USED, ALONG WITH THEIR DEFINITION AND SCOPE FOR MODELING THE MCF-7 BREAST CANCER CELL LINE INHIBITORY ACTIVITY OF BENZOTHIAZOLE DERIVATIVES

Descriptor class (acronyms)	Definition and scope
Constitutional (CONST)	Dimensionless or 0D descriptors; independent from molecular connectivity and conformations
Topological (TOPO)	2D-descriptor from molecular graphs and independent conformations
Molecular walk counts (MWC)	2D-descriptors representing self-returning walks counts of different lengths
Modified Burden eigenvalues (BCUT)	2D-descriptors representing positive and negative eigenvalues of the adjacency matrix, weights the diagonal elements and atoms
Galvez topological charge indices (GALVEZ)	2D-descriptors representing the first 10 eigenvalues of corrected adjacency matrix
2D-autocorrelations (2D-AUTO)	Molecular descriptors calculated from the molecular graphs by summing the products of atom weights of the terminal atoms of all the paths of the considered path length (the lag)
Functional groups (FUNC)	Molecular descriptors based on the counting of the chemical functional groups
Atom centered fragments (ACF)	Molecular descriptors based on the counting of 120 atom centered fragments, as defined by Ghose-Crippen
Empirical (EMP)	1D-descriptors represent the counts of non-single bonds, hydrophilic groups and ratio of the number of aromatic bonds and total bonds in an H-depleted molecule
Properties (PROP)	1D-descriptors representing molecular properties of a molecule

[a]Reference [9].

were excluded. This has reduced the total dataset of the compounds from 442 to 85 descriptors as relevant ones for the MCF-7 breast cancer cell line inhibitory activity. A brief description of the computational procedure is given below.

Model development: The CP-MLR is a 'filter' based variable selection procedure for model development in QSAR studies[10]. Its procedural aspects and implementation are discussed in some of the publications[11-15]. It involves selected subset regressions. In this procedure a combinatorial strategy with appropriately placed 'filters' has been interfaced with MLR to result in the extraction of diverse structure-activity models, each having unique combination of descriptors from the dataset under study. In this, the contents and number of variables to be evaluated are mixed according to the predefined confines. Here the 'filters' are significance evaluators of the variables in regression at different stages of model development. Of these, filter-1 is set in terms of inter-parameter correlation cutoff criteria for variables to stay as a subset (filter-1, default value 0.3 and upper limit ≤ 0.79). In this, if two variables are correlated higher than a predefined cutoff value the respective variable combination is forbidden and will be rejected. The second filter is in terms of *t*-values of regression coefficients of variables associated with a subset (filter-2, default value 2.0). Here, if the ratio of regression coefficient and associated standard error of any variable is less than a predefined cutoff value then the variable combination will be rejected. Since successive additions of variables to multiple regression equation will increase successive multiple correlation coefficient (r) values, square-root of adjusted multiple correlation coefficient of regression equation, r-bar, has been used to compare the internal explanatory power of models with different number of variables. Accordingly, a filter has been set in terms of predefined threshold level of r-bar (filter-3, default value 0.71) to decide the variables' 'merit' in the model formation. Finally, to exclude false or artificial correlations, the external consistency of the variables of the model have been addressed in terms of cross-validated R^2 or Q^2 criteria from the leave-one-out (LOO) cross-validation procedure as default option (filter-4, default threshold value $0.3 \leq Q^2 \leq 1.0$). All these filters make the variable selection process efficient and lead to unique solution. In order to collect the descriptors with higher information content and explanatory power, the threshold of filter-3 was successively incremented with increasing number of descriptors (per equation) by considering the r-bar value of the preceding optimum model as the new threshold for next generation.

Model validation: In this study, the data set is divided into training set for model development and test set for external prediction. Goodness of fit of the models was assessed by examining the multiple correlation coefficient (r), the standard deviation (s), the F-ratio between the variances of calculated and observed activities (F).

The internal validation of derived model was ascertained through the cross-validated index, Q^2, from leave-one-out and leave-four-out procedures. The leave-one-out method creates a number of modified data sets by taking away one compound from the parent data set in such a way that each observation has been removed once only. Then one model is developed for each reduced data set and the response values of the deleted observations are predicted from these models. In leave-four-out procedure, a group of four compounds is randomly kept outside the analysis each time in such a way that all the compounds, for once, become the part of the predictive groups. A value greater than 0.5 of Q^2-index hints toward a reasonable robust model. The external validation or predictive power of derived model is based on test set compounds. r^2_{Test} is the squared correlation coefficient between the observed and predicted data of the test-set. A value greater than 0.5 of r^2_{Test} suggests that the model obtained from training set has a reliable predictive power.

Y-Randomization: Chance correlations, if any, associated with the CP-MLR models were recognized in randomization test[16,17] by repeated scrambling of the biological response. The data sets with scrambled response vector have been reassessed by multiple regression analysis (MRA). The resulting regression equations, if any, with correlation coefficients better than or equal to the one corresponding to the unscrambled response data were counted. Every model has been subjected to 100 such simulation runs. This has been used as a measure to express the per cent chance correlation of the model under scrutiny.

RESULTS AND DISCUSSION

In multi-descriptor class environment, exploring for best model equation(s) along the descriptor class provides an opportunity to unravel the phenomenon under investigation. In other words, the concepts embedded in the descriptor classes relate the biological actions revealed by the compounds. For the purpose of modeling study, 5 compounds have been included in the test set for the validation of the models derived from 17 training set compounds. A total number of 85 significant descriptors from 0D-, 1D- and 2D-classes have been subjected to CP-MLR analysis with default 'filters' set in it. Statistical models in three descriptor(s) have been derived successively to achieve the best relationship correlating MCF-7 cell line inhibitory activity. These models (with 85 descriptors) were identified in CP-MLR by successively incrementing the filter-3 with increasing number of descriptors (per equation). For this the optimum r-bar value of the preceding level model has been used as the new threshold of filter-3 for the next generation. The highest significant models in three descriptors are given below.

pGI_{50} = 0.508(0.134)MAXDP + 12.200(3.050)BELe3 + 12.603(3.724)BELe6 - 21.236

n = 17, r = 0.921, s = 0.575, F = 24.065, Q^2_{LOO} = 0.573, Q^2_{L4O} = 0.709, r^2_{randY}(sd) = 0.439(0.167), r^2_{Test} = 0.547 (1)

pGI_{50} = -4.509(1.408)BEHm3 + 29.529(4.082)GGI9 - 10.290(1.802)MATS5p + 18.754

n = 17, r = 0.919, s = 0.579, F = 23.632, Q_{2LOO} = 0.763, Q^2_{L4O} = 0.586, r^2_{randY}(sd) = 0.387(0.143), r^2_{Test} = 0.588 (2)

pGI_{50} = 19.775(5.216)GGI9 - 6.411(1.830)MATS5p + 101.460(32.399)BELp2 - 187.842

n = 17, r = 0.918, s = 0.585, F = 23.096, Q^2_{LOO} = 0.736, Q^2_{L4O} = 0.659, r^2_{randY}(sd) = 0.420(0.154), r^2_{Test} = 0.615 (3)

pGI_{50} = 0.704(0.152)MAXDP + 19.006(3.281)BELe3 - 3.061(1.054)IC2 - 8.369

n = 17, r = 0.910, s = 0.610, F = 20.902, Q^2_{LOO} = 0.652, Q^2_{L4O} = 0.661, r^2_{randY}(sd) = 0.394(0.149), r^2_{Test} = 0.761 (4)

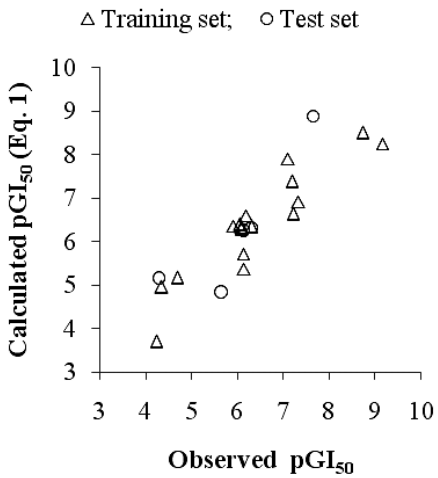
In above regression equations, the values given in the parentheses are the standard errors of the regression coefficients. The r^2_{randY}(sd) is the mean random squared multiple correlation coefficient of the regressions in the activity (Y) randomization study with its standard deviation from 100 simulations. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation. The signs of the regression coefficients suggest the direction of influence of explanatory variables in the models.

Most of the descriptors which have taken part in above models (BELe3, BELe6, BEHm3 and BELp2) belong to BCUT class of Dragon descriptors. The BCUT descriptors are the first 8 highest and the lowest absolute eigenvalues, BEHwk and BELwk, respectively, for the modified Burden adjacency matrix. Here w refers to the atomic property and k to the eigenvalue rank. The ordered sequence of the highest and the lowest eigenvalues reflect upon the relevant aspects of molecular structure, useful for similarity searching. It is evident from the sign of regression coefficients of these descriptors that all descriptors but BEHm3 have contributed positively to the activity. Thus, advocate that bigger values of lowest eigenvalues weighted by atomic Sanderson electronegativities (BELe3 and BELe6) and atomic polarizabilities (BELp2) and a lower value of highest eigenvalue weighted by atomic masses (BEHm3) would be beneficiary to the activity.

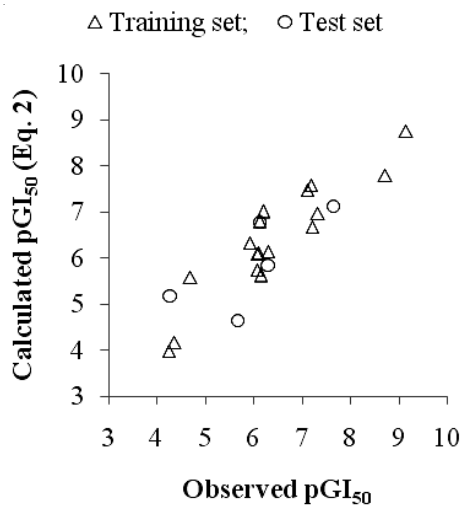
The descriptors MAXDP and IC2 belong to TOPO class of Dragon descriptors. The TOPO class descriptors are based on a graph representation of the molecule and are numerical quantifiers of molecular topology obtained by the application of algebraic operators to matrices representing molecular graphs and whose values are independent of vertex numbering or labeling. They can be sensitive to one or more structural features of the molecule such as size, shape, symmetry, branching and cyclicity and can also encode chemical information concerning atom type and bond multiplicity. The descriptor MAXDP (maximal electrotopological positive variation) has shown positive contribution to the activity suggesting that a higher value of this descriptor would be favourable to the activity. On the other hand another descriptor from this class IC2, the information content index of 2nd order neighborhood symmetry, has shown negative correlation to the activity. Thus a lower value of this descriptor will be supportive to the activity.

The remaining descriptors MATS5p and GGI9 are from 2D-AUTO and GALVEZ class, respectively. The 2D autocorrelations descriptors are molecular descriptors which describe how a considered property is distributed along a topological molecular structure. The 2D-AUTO descriptors have their origin in autocorrelation of topological structure of Broto-Moreau (ATS), of Moran (MATS) and of Geary (GATS). The computation of these descriptors involve the summations of different autocorrelation functions corresponding to the different fragment lengths and lead to different autocorrelation vectors corresponding to the lengths of the structural fragments. Also a weighting component in terms of a physico-chemical property has been embedded in this descriptor. As a result these descriptors address the topology of the structure or parts thereof in association with a selected physicochemical property. In these descriptors' nomenclature, the penultimate character, a number, indicates the number of consecutively connected edges considered in its computation and is called as the autocorrelation vector of lag n (corresponding to the number of edges in the unit fragment). The very last character of the descriptor's nomenclature indicates the physicochemical property considered in the weighting component-m for mass or v for volume or e for Sanderson electronegativity or p for polarizability - for its computation. It is evident from the sign of regression coefficients of the participating descriptor that descriptor MATS5p from this class has contributed negatively to the activity. Thus a lower value of descriptor MATS5p (Moran autocorrelation of lag 5 weighted by atomic polarizabilities) will be in favour of activity. The GALVEZ descriptors are the Galvez topological charge indices and have their origin in the first ten eigenvalues of the polynomial of corrected adjacency matrix of the compounds. All the GALVEZ class descriptors belong to two categories. Of this one category corresponds to the topological charge index of order n (GGIn) and the other to the mean topological charge index of order n (JGIn), where 'n' represents the order of eigen value. The positive influence of descriptor GGI9 (topological charge index of 9th order) from this class to the activity suggested that a higher value of 9th order charge index would be beneficiary to the activity.

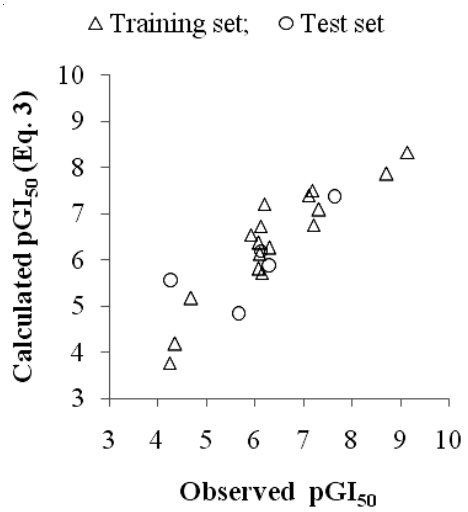
These models (eqns. 1-4) have accounted for nearly 83 per cent variance in the observed activities. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation. The values greater than 0.5 of Q^2-index is in accordance to a reasonable robust QSAR model. The pGI_{50} values of training and test set compounds calculated using eqns. 1-4 have been included in Table-1. The models (1-4) are validated with an external test set of five compounds listed in Table-1. The predictions of the test set compounds based on external validation are found to be satisfactory as reflected in the test set r^2 (r^2_{Test}) values. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Fig. 1. Thus the descriptors identified for rationalizing the activity give avenues to modulate the structure to a desirable biological end point.



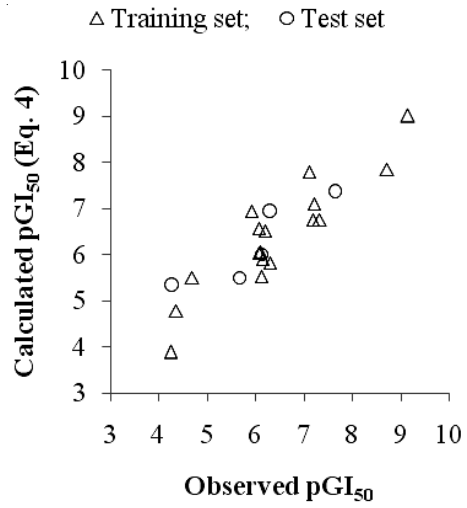
(a) Plot of observed *versus* calculated pGI_{50} values for training and test set compounds



(b) Plot of observed *versus* calculated pGI_{50} values for training and test set compounds



(c) Plot of observed *versus* calculated pGI_{50} values for training and test set compounds



(d) Plot of observed *versus* calculated pGI_{50} values for training and test set compounds

Fig. 1. Plot of observed *versus* calculated pGI_{50} values for training and test set compounds

Conclusion

This study has provided a rational approach for the development of new antitumor benzothiazole derivatives as MCF-7 breast cancer cell line inhibitors. The descriptors identified in CP-MLR analysis have highlighted the role of atomic properties in respective lag of 2D-autocorrelations (MATS5p) and Modified Burden eigenvalues (BELe3, BELe6, BEHm3 and BELp2), maximal electrotopological positive variation (MAXDP), information content index of 2nd order neighbourhood symmetry (IC2) and the 9th order Galvez topological charge (GGI9) to explain the MCF-7 breast cancer cell line inhibitory actions of benzothiazole derivatives. Thus the descriptors identified for rationalizing the activity give avenues to modulate the structure to a desirable biological end point.

ACKNOWLEDGEMENTS

The authors are thankful to their respective institutions for providing necessary facilities to complete this study.

REFERENCES

1. P.C. Yates, C.J. McCall and M.F.G. Stevens, *Tetrahedron*, **47**, 6493 (1991).
2. M.F.G. Stevens, C.J. McCall, P. Lelieveld, P. Alexander, A. Richter and D.E. Davies, *J. Med. Chem.*, **37**, 1689 (1994).
3. (a) A.L. Loaiza-Pérez, V. Trapani, C. Hose, S.S. Singh and J. Trepel, *Mol. Pharmacol.*, **61**, 13 (2002); (b) V. Trapani, V. Patel, H.P. Ciolino, G.C. Yeh, C. Hose, J.B. Trepel, M.F.G. Stevens, E.A. Sausville and A.L. Loaiza-Pérez, *Br. J. Cancer*, **88**, 599 (2003).
4. C.O. Leong, M. Suggitt, D.J. Swaine, M.C. Bibby, M.F.G. Stevens and T.D. Bradshaw, *Mol. Cancer Ther.*, **3**, 1565 (2004).
5. (a) T.D. Bradshaw, C.S. Matthews, J. Cookson, E.-H. Chew, M. Shah, K. Bailey, A. Monks, E. Harris, A.D. Westwell, G. Wells, C.A. Laughton and M.F.G. Stevens, *Cancer Res.*, **65**, 3911 (2005); (b) A. Mukherjee, T.D. Bradshaw, A.D. Westwell, M.F.G. Stevens, J. Carmichael and S.G. Martin, *Br. J. Cancer*, **92**, 30 (2005).
6. C.G. Mortimer, G. Wells, J.-P. Crochard, E.L. Stone, T.D. Bradshaw, M.F.G. Stevens and A.D. Westwell, *J. Med. Chem.*, **49**, 179 (2006).
7. A.D. Westwell and M.F.G. Stevens, *Drug Discov. Today*, **9**, 625 (2004).
8. ChemDraw Ultra 6.0 and Chem3D Ultra, Cambridge Soft Corporation, Cambridge, USA.
9. R. Todeschini and V. Consonni, Dragon Software (Version 1.11-2001) Milano, Italy (2001).
10. Y.S. Prabhakar, *QSAR Comb. Sci.*, **22**, 583 (2003).
11. S. Sharma, Y.S. Prabhakar, P. Singh and B.K. Sharma, *Eur. J. Med. Chem.*, **43**, 2354 (2008).
12. B.K. Sharma, P. Pilania and P. Singh, *J. Enzym. Inhibit. Med. Chem.*, **24**, 607 (2009).
13. S. Sharma, B.K. Sharma, P. Pilania, P. Singh and Y.S. Prabhakar, *J. Enzym. Inhibit. Med. Chem.*, **24**, 1024 (2009).
14. B.K. Sharma, P. Pilania, P. Singh and Y.S. Prabhakar, *SAR QSAR Environ. Res.*, **21**, 169 (2010).
15. B.K. Sharma, P. Pilania, K. Sarbhai, P. Singh and Y.S. Prabhakar, *Mol. Divers*, **14**, 371 (2010).
16. S.-S. So and M. Karplus, *J. Med. Chem.*, **40**, 4347 (1997).
17. Y.S. Prabhakar, V.R. Solomon, R.K. Rawal, M.K. Gupta and S.B. Katti, *QSAR Comb. Sci.*, **23**, 234 (2004).