informa
healthcare

RESEARCH ARTICLE

Structural feature study of benzofuran derivatives as farnesyltransferase inhibitors

N. S. Hari Narayana Moorthy, Sergio F. Sousa, Maria J. Ramos, and Pedro A. Fernandes

REQUIMTE, Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, Porto, Portugal

Abstract
Ras proteins are small GTPases (G-proteins) that play a key role in cell growth and cell proliferation in the mitogen-activated protein kinase signal transduction pathway. Farnesylation is a critical step for membrane binding and the biological function of G-proteins. In the present investigation, we have studied the structural features of some molecules that are acting on the farnesyltransferase (FTase) enzyme for the inhibition of the farnesylation step in G-proteins. The benzofuran derivatives have activity against FTase inhibition and antiproliferative activity on QG56 cell lines. The result obtained from the quantitative structure-activity relationship study of these compounds shows that the models have significant predictive power and stability, as shown by statistical parameters such as R^2, Q^2, R^2_{pred}, R^2_m, F-value, Durbin–Watson, variable inflation factor values, Mahalanobis, and Cook's distances. The contribution of each descriptor for the activities (β-coefficients) reveals that the P-VSA descriptors (van der Waals surface area descriptors) such as vsa_pol, vsa_acc and SMR_VSA3 are the major contributors for the activity, along with other descriptors such as the partition coefficient, the partial charge, the atom and bond count and the adjacency, and distance descriptors. Earlier study on the FTase enzyme in our laboratory reveals that the existence of positively-charged groups on the FTase active site is important for drug design. It is also showing that the presence of hydrogen bonding donor and acceptor groups, together with negatively charged substituents is critical for improved activity by this series of molecules. These results offer important clues for the development of novel FTase inhibitors.

Keywords: Farnesyltransferase, QG56 cell lines, antiproliferative activity, QSAR, benzofuran derivatives

Introduction

Cancer is still one of the life threatening diseases and the second leading cause of death in the world. The design and development of novel, more active molecules that selectively kill tumour cells or inhibit their proliferation without the general toxicity are one of the most important goals in medicinal chemistry. The discovery of novel bioactive molecules depends on the validated targets, among which is the oncogene *Ras*. Ras proteins are small GTPases (G-protein) that plays a key role in cell growth and cell proliferation in the mitogen-activated protein kinase signal transduction pathway[1].

The post-translational modification of the oncogenic Ras protein by the farnesyl pyrophosphate (FPP) intermediate begins with the farnesylation of a carboxy terminus protein in the CAAX (tetrapeptide motif C: Cys, A: an aliphatic amino acid, X: Ser, Met, Gln, Ala at their C-terminal) sequence by farnesyltransferase (FTase). This farnesylation step is critical for membrane binding and the biological function of G-proteins[2,3] and its inhibition in the signal transduction pathway affects the G-protein function leading to inhibition of cell proliferation. FTase inhibitors have been extensively developed as anticancer agents because of their ability to block tumour growth[3]. A survey of cancer cell lines has shown that >70% of cells are sensitive to FTase inhibitors. Some FTase inhibitors, such as R115777 (tipifarnib), SCH66336 (lonafarnib), BMS 214662, L-778,123, and SCH44342 are currently being assessed in clinical trials and have demonstrated clinical efficacy for the treatment of human cancers[4–7]. The research on

RIGHTSLINK

FTase inhibitors in several other non-cancer related illnesses e.g. malaria, the African sleeping sickness, Chagas disease, leishmaniasis, and toxoplasmosis is presently underway[8–11]. Research efforts in this area are progressing on the rapid development of novel FTase inhibitors for cancer therapy[12,13].

Computational based drug design is a rapidly growing field and an important component of the medicinal chemistry discipline. It is aimed at shortening the drug discovery process, which otherwise may be much longer and expensive. Understanding the interactions between small molecules and their molecular targets should improve our ability to predict the activity of new compounds. The interaction is dependent upon the structural features of the molecules and the target. Quantitative structure-activity relationship (QSAR) is one of the possible techniques to study the structural features of the compounds needed for the interaction with the target by validated QSAR models. The development of validated QSAR models is also an important step in the biological activity prediction of molecules[14].

Our research group has been directly involved for the past 7 years in the study of FTase, focusing mainly on the catalytic mechanism of this important enzyme. In particular, we have performed a series of detailed quantum mechanical studies on the several Zn coordination spheres formed during the catalytic mechanism of FTase[15–18], in an attempt to reconcile what was initially regarded as contradicting experimental evidence arising from the available X-ray crystallographic structures[19–21], extended X-ray absorption fine structure[22] results and kinetic and mutagenesis data[23–27]. These studies served as a basis for several other analysis on important but more general properties on the global enzyme and active-site residues through extensive molecular dynamics simulations[29,30], and culminated with the finding and theoretical characterization of the transition state structure for the farnesylation process[31], a feature which was confirmed in other independent studies[32].

In the present investigation, we have performed a QSAR analysis on the structural features, especially the P_VSA descriptors and other descriptors, of novel benzofuran derivatives for FTase inhibitory and antiproliferative activity. Benzofuran derivatives, with a heterocyclic nucleus, have been shown to have a number of important activities, including anti-HIV, anticancer, antimicrobial, non-selective aromatase-inhibitory activity, etc[33,34]. The P_VSA descriptors include a novel set of descriptors called widely applicable set of descriptors` (SMR, SlogP, PEOE, and Q) derived by summing the approximate exposed surface area for each atom according to the classification based upon logP (SlogP), molar refractivity (SMR) and partial charge (PEOE and Q). The P_VSA descriptors are based on approximate van der Waals surface area (VSA) calculations using connection table approximation for an atom i, along with some atomic property P. Each descriptor in this series is defined as the atomic VSA contributions of each atom i with property P_i in the range (u,v). Thus P_VSA(u,v) can be defined as given in Equation 1.

$$P\_VSA_{(u,v)} = \Sigma V_i \delta[P_i \in (u,v)] \tag{1}$$

where V_i is atomic contribution of atom i to the VSA of the molecule[35–37].

The literature study shows that there is no QSAR work done on this series of compounds. In the present study, P_VSA descriptors along with the BCUT and the atom and bond count descriptors were used to correlate the structural properties of benzofuran derivatives responsible for FTase inhibitory activity with the validated QSAR models. The models were validated by internal validation [leave one out (LOO) crossvalidation], external validation (test and inactive compounds), distance approaches, multicollinearity, and serial autocorrelation studies were done to validate the models. The first two variables describe the predictive power of the models and the latter variables describe the stability of the models.
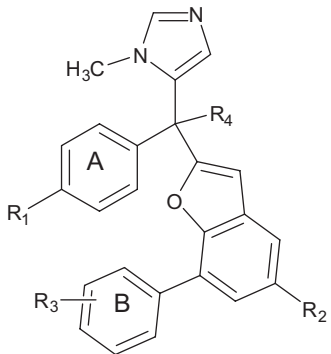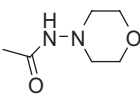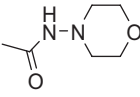
Experimental

Data set
A series of novel benzofuran derivatives, for which the FTase inhibitory activity and the antiproliferative activity on human non-small cell lung carcinoma (QG56) has been recently described in the literature[38] was considered to perform the QSAR study. This series was based on tipifarnib (R115777[8]), one of the most potent FTase inhibitors currently undergoing clinical trials, with a novel benzofuran core template replacing the quinolinone moiety of tipifarnib[38]. In this series, 30 compounds were reported with inhibitory concentration (IC$_{50}$) in the nM range (Table 1). Among the 30 compounds, only 29 compounds have defined inhibitory activity against FTase and 27 compounds have antiproliferative activity against QG56 cell lines. The inhibitory concentration of the molecules was converted into–logIC$_{50}$ or log1/IC$_{50}$ (pIC$_{50}$) to reduce the skewness of the data and further convert into free energy changes of the molecules. The IC$_{50}$ of the compounds (nM) were converted to molar concentration before calculating as –logIC$_{50}$.

Descriptor calculation
Computational studies of the compounds were performed with Molecular Operating Environment (MOE, Montreal, Canada)[39] and Statistica[40] softwares (Tulsa, OK). The template structure obtained from the available X-ray crystallographic data (PDB code: 2ZIS and 2ZIR[38]) was used to sketch the 3D structures of the molecules. MOE molecular modelling software was used to perform energy optimization and descriptor calculations. The semi-empirical MOPAC program with Austin Model 1 (AM1) Hamiltonian gradients of MOE software was used to optimize the geometry of the molecules. The QuaSAR

Table 1. Structure and activity of benzofuran derivatives considered for the present study.

Compound code	R_1	R_2	R_3	R_4	FTase IC_{50} (nM)	QG56 IC_{50} (nM)
8a	Cl	NO_2	H	OH	170	1885
8b	Br	NO_2	H	OH	360	1477
8c	I	NO_2	H	OH	360	>10000
8d	COO-t-Bu	NO_2	H	OH	>1000	6248
8e	COOH	NO_2	H	OH	850	>10,000
8f	NO_2	NO_2	H	OH	30	547
8g	OCH_3	NO_2	H	OH	250	4967
8h	CN	NO_2	H	OH	6.4	1477
8i	CN	NO_2	2-F	OH	3.3	89.5
8j	CN	NO_2	3-OCH_3	OH	8.5	38.8
8k	CN	NO_2	3-CN	OH	2.8	22.9
8l	CN	NO_2	3-CH_3	OH	11	23.2
8m	CN	NO_2	3-F	OH	6.3	145
8n	CN	NO_2	4-OCH_3	OH	4	36
8o	CN	NO_2	4-CN	OH	3.4	32.6
8p	CN	NO_2	4-F	OH	7.2	158
8q	CN	$COOCH_3$	3-OCH_3	OH	3.2	36.3
8r	CN	$CONH_2$	3-OCH_3	OH	0.9	8.2
8s	CN	CHO	3-OCH_3	OH	2	15.2
8t	CN	CH_2OH	3-OCH_3	OH	1	16.8
8u	CN	$CH_2N(CH_3)_2$	3-OCH_3	OH	11	142
8v	CN	(acetyl-N-morpholino)	3-OCH_3	OH	2.6	Nt
8w	Cl	(acetyl-N-morpholino)	H	OH	6.4	73.6
8x	CN	CN	H	OH	2.4	14.5
11a (S)	CN	NO_2	H	NH_2	1.5	5.9
11b (R)	CN	NO_2	H	NH_2	49	206.1
11c (S)	CN	NO_2	3-OCH_3	NH_2	0.8	1.1
11d (S)	CN	CN	3-OCH_3	NH_2	1.2	1.5
11e (S)	CN	CN	3-CN	NH_2	0.7	2
11f (S)	CN	CN	3-F	NH_2	1.1	2

FTase, farnesyltransferase.

module of MOE was used for descriptor calculations. A large number of theoretical molecular descriptors are available in the package to define the structural properties of molecules explicitly. The descriptors found in the module include physical properties (14 descriptors), subdivided surface areas (18 descriptors), atom, and bond counts (41 descriptors), Kier and Hall connectivity and κ-shape indices (16 descriptors), adjacency and distance matrix descriptors (33 descriptors), pharmacophore feature descriptors (12 descriptors), and

partial charge descriptors (50 descriptors). The P_VSA descriptors are a set of 52 2D descriptors describing electrostatic, lipophilic, steric, and pharmacophoric properties in terms of molecular surface[41,42].

Statistical analysis

The calculated descriptors were initially screened for invariant nature and insignificance using QuaSAR-Contingency module of MOE. QuaSAR-Contingency is a statistical application designed to assist in the selection of descriptors for QSAR. In order to reduce the redundant and useless information, descriptors that possess zero correlation on the dependent variable (biological activity) as well as descriptors showing intercorrelation superior to 0.5 were discarded (Supplementary Table S1 and S2). The data set was divided into training and test set in order to perform the QSAR analysis. The test set compounds (six compounds) were selected by random.

In order to quantify the correlation, QSAR models were developed using observed FTase inhibitory activity or antiproliferative activity, as dependent variables, and the calculated physicochemical descriptors as independent variables for multiple linear regression (MLR) analysis and partial least square (PLS) regression analysis. Statistica 8.0 software was used to develop statistically significant models for the complete data set that posses defined activity. Since a multiple linear model with a large number of variables can be too cumbersome to use, we have used stepwise regression to refine the model by determining the relative importance of each variable and its statistical significance. Furthermore, an equation containing an excessive number of independent variables is likely to be overfitted. Hence, the upper limit of rule of thumb (six cases per variable) was adopted in the analysis[37].

The significant models were selected for further study taking into account high correlation coefficients, F_{test}, t_{test} values and significance of the descriptors included in the model building [variable inflation factor (VIF), Durbin–Watson (DW) and β-coefficients]. The selected significant models were validated by internal (LOO) and external (test set) validation method. The distance based approaches were also used to validate the predictive ability of the models by the same software.

PLS analysis was performed for the data set with the Non-linear Iterative Partial Least Squares (NIPALS) algorithm. NIPALS is a well established iterative technique widely used in building principal component analysis and PLS analysis models. With its guaranteed convergence rate, typical accuracy, and scalability (i.e. its ability to handle large data sets), the NIPALS algorithm can construct PLS models with reliable efficiency. In the study, PLS analysis has been performed which validated the QSAR models. A maximum number of criterions of 50 and convergence criterion of 0.0001 was considered for the study.

Results and discussion

Statistical parameters

The correlation between the biological activity (FTase inhibitory activity and antiproliferative activity) and the structural features were done using QSAR techniques and the obtained results are given in Table 2. In the models, N stands for number of compounds (cases) contributed to build the respective models. The values within the parenthesis following the regression coefficient terms are the standard errors of the regression terms and the constants. R is the correlation coefficient and R^2 is the squared correlation coefficient. They describe the relative measure of the quality of fit by the regression equation. PLS-R^2 stands for the squared correlation coefficient derived from the PLS regression study. The correlation coefficient explains the variation in the observed data (experimental); its value varies from –1 to +1. The closer the R-values to 1, the better the fit of the regression equation. The R^2-value of the selected models are >0.83 against FTase inhibitory activity and are >0.88 against antiproliferative activity (QG56). The comparison of the result obtained from the statistical studies (MLR and PLS) are given in Table 3.

F is the Fischer ratio that represents the ratio between the variance of calculated and observed activities. The values within parentheses that follow the calculated F-values are the tabulated values at 99% significance. The F-value indicates that the regression relations are not a chance fit but are a significant occurrence. t Is the Student-t test and the value in the parenthesis after the calculated value, is the tabulated t-value at 0.0005 confidence level. The F statistics and the t-value of the models have a large margin of difference for the limiting values at 0.01 (99%) and 0.0005 (99.95%), respectively, which shows the models are statistically significant for further study. The contribution of each descriptor (β-coefficient) in the model is the regression coefficient that would have been obtained by adjusting all of the variables to a mean of 0 and a standard deviation of 1. It also allows us to compare the relative contribution of each independent variable in the prediction of the dependent variables[41]. In the present study, the van der Waals (P_VSA) descriptors have high contribution for both activities (FTase inhibition and antiproliferative activities) along with other descriptors such as the atom and bond count, BCUT, partial charge, and physical (partition coefficient). The graphical representation of this result is given in Figure 1A and 1B.

Validation of QSAR models

Any QSAR modelling should ultimately lead to statistically robust models capable of making accurate and reliable predictions of biological activities of new compounds. The success of any QSAR model depends on the accuracy of the input data, selection of appropriate descriptors, and statistical tools and most importantly

Table 2. Detail of the selected significant QSAR models and their statistical parameters.

Model no.	Model	Statistical parameters
Activity 1: FTase inhibitory activity		
Model 1	$pIC_{50(FT)} = 0.0433$ (± 0.0055) SMR_VSA3 $- 41.0947$ (± 9.3519) PEOE_RPC+ $+ 10.6196$ (± 0.9700)	$N = 23$, $R = 0.9126$, $R^2 = 0.8328$, Adj$R^2 = 0.8161$, $Q^2_{train} = 0.7848$, $Q^2_{test} = 0.7658$, $F_{(2,20,0.01)} = 49.8110$ (5.8490), SEE $= 0.4069$, $t_{(20,0.0005)} = 10.9480$ (3.8495), $pP = 0.0000$, PRESS$_{(train)} = 3.2890$, PRESS$_{(test)} = 0.453$, S$_{PRESS} = 0.4055$, SDEP$_{(train)} = 0.3782$, SDEP$_{(test)} = 0.2748$, $R^2_{pred} = 0.8000$, $R^2_m = 0.7252$, β-value for PEOE_RPC+ $= -0.4100$ and SMR_VSA3 $= 0.7340$ (PLS: $R^2 = 0.8280$, Eigen value: 1.1936, $Q^2 = 0.7945$)
Model 2	$pIC_{50(FT)} = 0.8488$ (± 0.0967) b_triple $- 51.9479$ (± 8.4100) PEOE_RPC+ $+ 12.3112$ (± 0.8337)	$N = 23$, $R = 0.9269$, $R^2 = 0.8592$, Adj$R^2 = 0.8451$, $Q^2_{train} = 0.8117$, $Q^2_{test} = 0.6704$, $F_{(2,20,0.01)} = 61.0280$ (5.8490), SEE $= 0.3734$, $t_{(20,0.0005)} = 14.7670$ (3.8495), $P = 0.0000$, PRESS$_{(train)} = 2.7930$, PRESS$_{(test)} = 0.6375$, S$_{PRESS} = 0.3737$, SDEP$_{(train)} = 0.3485$, SDEP$_{(test)} = 0.3260$, $R^2_{pred} = 0.7186$, $R^2_m = 0.7572$, β-value for b_triple $= 0.7380$ and PEOE_RPC+ $= -0.5200$ (PLS: $R^2 = 0.8589$, Eigen value: 1.0566, $Q^2 = 0.8291$)
Activity 2: antiproliferative activity		
Model 3	$pIC_{50(QG56)} = 0.0377$ (± 0.0047) VSA_pol -33.7387 (± 5.0556) BCUT_SlogP_3 $+ 92.1727$ (± 13.0735)	$N = 21$, $R = 0.9588$, $R^2 = 0.9193$, Adj$R^2 = 0.9104$, $Q^2_{train} = 0.8825$, $Q^2_{test} = 0.8614$, $F_{(2,18,0.01)} = 102.5700$ (6.0130), SEE $= 0.2958$, $t_{(18,0.0005)} = 7.0503$ (3.9216), $P = 0.0000$, PRESS$_{(train)} = 1.5652$, PRESS$_{(test)} = 1.1884$, S$_{PRESS} = 0.2949$, SDEP$_{(train)} = 0.2730$, SDEP$_{(test)} = 0.4450$, $R^2_{pred} = 0.8758$, $R^2_m = 0.8326$, β-value for BCUT_SlogP_3 $= -0.5100$ and VSA_pol $= 0.6050$ (PLS: $R^2 = 0.9183$, Eigen value: 1.4770, $Q^2 = 0.8697$)
Model 4	$pIC_{50(QG56)} = 0.0398$ (± 0.0064) VSA_acc -35.2638 (± 6.0386) BCUT_SlogP_3 $+ 96.6178$ (± 15.5814)	$N = 21$, $R = 0.9403$, $R^2 = 0.8841$, Adj$R^2 = 0.8712$, $Q^2_{train} = 0.8357$, $Q^2_{test} = 0.8380$, $F_{(2,18,0.01)} = 68.6370$ (6.0130), SEE $= 0.3546$, $t_{(18,0.0005)} = 6.2009$ (3.9216), $P = 0.0000$, PRESS$_{(train)} = 2.2465$, PRESS$_{(test)} = 1.3892$, S$_{PRESS} = 0.3533$, SDEP$_{(train)} = 0.3271$, SDEP$_{(test)} = 0.4812$, $R^2_{pred} = 0.8548$, $R^2_m = 0.7837$, β-value for BCUT_SlogP_3 $= -0.5300$ and VSA_acc $= 0.5640$ (PLS: $R^2 = 0.8840$, Eigen value: 1.4718, $Q^2 = 0.8260$)
Model 5	$pIC_{50(QG56)} = -1.4939$ (± 0.1097) logP(o/w) -0.0534 (± 0.0112) PEOE_VSA $+ 4 + 15.3854$ (± 0.5658)	$N = 21$, $R = 0.9599$, $R^2 = 0.9214$, Adj$R^2 = 0.9127$, $Q^2_{train} = 0.8954$, $Q^2_{test} = 0.7876$, $F_{(2,18,0.01)} = 105.5100$ (6.0130), SEE $= 0.2920$, $t_{(18,0.0005)} = 27.1930$ (3.9216), $P = 0.0000$, PRESS$_{(train)} = 1.5241$, PRESS$_{(test)} = 1.8214$, S$_{PRESS} = 0.2910$, SDEP$_{(train)} = 0.2694$, SDEP$_{(test)} = 0.5510$, $R^2_{pred} = 0.8096$, $R^2_m = 0.8355$, β-value for PEOE_VSA $+ 4 = -0.3200$ and logP(o/w) $= -0.9000$ (PLS: $R^2 = 0.9212$, Eigen value: 1.0124, $Q^2 = 0.8764$)

FTase, farnesyltransferase; QSAR, quantitative structure-activity relationships; PLS, partial least square; SEE, standard error estimate, SDEP, squared deviation error of prediction.

Table 3. Comparison of validation parameter obtained from MLR and PLS analysis.

Model no.	MLR		PLS		D^2			CD		
	R^2	Q^2	R^2	Q^2	Min	Max	Mean	Min	Max	Mean
Model 1	0.83	0.78	0.83	0.79	0.0280	6.2590	1.9130	0.0007	0.2190	0.0458
Model 2	0.86	0.81	0.86	0.83	0.0064	6.4834	1.9130	0.0000	0.3360	0.0546
Model 3	0.92	0.88	0.92	0.87	0.1360	12.5009	1.9048	0.0001	0.5600	0.0738
Model 4	0.88	0.84	0.88	0.83	0.0665	12.5750	1.9048	0.0000	0.4967	0.0677
Model 5	0.92	0.90	0.92	0.88	0.0218	8.5775	1.9048	0.0001	0.2282	0.0527

CD, Cook's distance; MLR, multiple linear regression analysis; PLS, partial least square analysis; D^2, Mahalanobis distance.

validation of the developed model. The derived models were validated to examine the self-consistency between them, which implies a quantitative assessment of the model robustness and its predictive power. In this analysis, LOO crossvalidation techniques (internal validation method) and test set method (external validation method) were used to find out the predictive power of the models.

The results obtained from the internal validation method (LOO) show that Q^2 (the cross-validated correlation coefficient), which provides the statistical significant and predictability of the models has very satisfactory values. Q^2 is used as a criterion of both robustness and predictive ability of the model. A high Q^2 (for instance $Q^2 > 0.5$) may be considered as an indicator of significant predictivity of the models[43–45]. Q^2 is calculated as Equation 2.

$$Q^2 = 1 - \frac{\sum_{i=1}^{n}(y_{exp} - y_{pred})^2}{\sum_{i=1}^{n}(y_{exp} - \bar{y}_{pred})^2}, \tag{2}$$

where, y_{exp} and y_{pred} are the observed and predicted values for the dependent variables, respectively and $\bar{y}$ is the average observed value.

The Q^2-value (Q^2_{train}) obtained from the MLR and PLS are >0.78 against both the activities (FTase inhibitory and antiproliferative activity), which indicates that the selected models have sufficient predictive power and self-consistency. The Q^2-value calculated from the test set compounds (Q^2_{test}) > 0.67 against FTase inhibitory activity and >0.78 against antiproliferative activities. These results show that the selected models have significant predictive ability. The relationship between the predicted and observed activity values (Table 4) are represented graphically in Figure 2 (graph 1–5).

The cross-validated correlation coefficient calculated for the models shows the selected models have significant Q^2-values, but it is not sufficient to consider the selected models are predictive. In order to establish its predictivity, additional validation parameters such as $PRESS_{(train\ and\ test)}$, S_{PRESS} (cross-validated standard error of prediction), $SDEP_{(train\ and\ test)}$ (squared deviation error of prediction), R^2_{pred} and R^2_m were calculated. The low S_{PRESS} and SDEP-value for the models developed with the training set reveals the models are statistically significant for the activity prediction. It is supported by the SDEP-value obtained from the test set compounds is also <0.6 confirm that the abovementioned statement for the predictive ability of the models. The PRESS-value for the models 1 and 2 are >2 for training and <1 for test sets while the models 3–5 have the values <2.2 for both training and test set. R^2_{pred}-value for the models of FTase inhibitory activity is >0.71 and is >0.81 for antiproliferative activity. Also the R^2_m-value for all the models is >0.7 shows the predictive ability of the models. These validation parameters also revealed that the developed models possess significant predictive ability, and which are satisfied the criteria (Q^2 > 0.5, R^2 > 0.6, R^2_{pred} > 0.5, and R^2_m > 0.5) recommended by Golbraikh and Tropsha[43] and Roy and Roy[46,47].

Distance based approaches are also a way of validation of the models. They represent the distance from each point to a particular point. Cook's distances indicate the distances between the computed B values (standard coefficient values) and the values one would have obtained if the respective case has been excluded (LOO). All distances should be of about equal magnitude, otherwise there is reason to believe that the respective case(s) biased the estimation of the regression coefficients[48]. The maximum Cook's distance value of the models is <0.56 which is <1 (squared Cook's distances[46]) and the Cook's distances of all the compounds have almost equal magnitude (<1), showing that the equation has significant predictive ability for FTase inhibitory and antiproliferative activity.

Mahalanobis distances methods (D^2) identify the interpolation region by assuming that the data have a normal distribution.
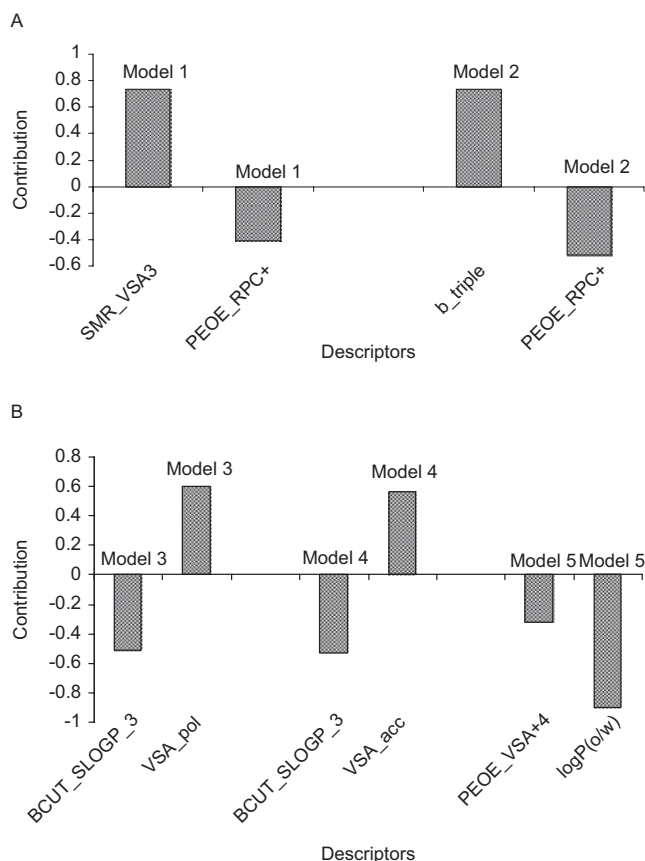

Figure 1. Contribution chart of descriptors in the models for the activity. (A) Contribution of descriptors in farnesyltransferase inhibitory activity. (B) Contribution of descriptors in antiproliferative activity (QG56).

$$MD = DM(x,y) = (x - \mu)^T \sum{}^{-1}(x - \mu), \tag{3}$$

where $\sum^{-1}$ is the inverse of the covariance matrix[41,49,50].

Mahalanobis distances improve the prediction accuracy and speed up a solution for QSAR. The higher the Mahalanobis distances for a case (molecule), the more the independent variables diverge from the average values. The results reveal that the cases in the developed models are normally distributed at 95% significance level.

Multicollinearity is a statistical phenomenon used in multiple regression models in which two or more predictor descriptors are highly correlated. In this situation, the regression coefficients of the descriptors may change erratically in response to small changes in the model or the data. Multicollinearity does not reduce the predictive power or reliability of the model as a whole; it only affects calculations regarding individual predictors.

To confirm the absence of multicollinearity, the VIF was calculated for each parameter in the regression. More precisely, the VIF is an index which measures how much the variance of a coefficient (square of the standard deviation) is increased because of collinearity. VIF

denotes the fact that the variance of the standardized regression coefficients can be computed as the product of the residual variance (for the correlation transformed model) and it can be calculated as follows Equation 4.

$$\text{VIF} = \frac{1}{1 - R^2} \text{ or } \frac{1}{\text{Tolerance}}, \tag{4}$$

where R^2 is the multiple correlation coefficient of one parameter's effect regressed on the remaining parameters. If R equals to zero (i.e. no correlation between X and the remaining independent variables) then VIF equals 1, the minimum value. A value greater than 10 is an indication of potential multicollinearity problems. Not uncommonly, a VIF of 10 or even one as low as 4 (equivalent to a tolerance level of 0.10 or 0.25) have been used as rules of thumb to indicate excessive or serious multicollinearity. In the selected models, the VIF value between 1 and 1.3 shows that the descriptors in the selected models are free from multicollinearity[51,52]. The results are provided in Table 5.

A DW test was employed to check the serial correlation of residuals (correlation of adjacent residuals). The

DW statistics is useful for evaluating the presence or absence of a serial correlation of residuals (i.e. whether or not residual for adjacent cases are correlated, indicating that the observations or cases in the data file are not independent).

$$d = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}, \tag{5}$$

where e_t is the residual associated with the observation at time t, since d is approximately equal to $2(1-r)$, where r is the sample autocorrelation of the residuals, $d = 2$ indicates no autocorrelation. The value of d always lies between 0 and 4. If the DW statistic is substantially less than 2, there is evidence of positive serial correlation and a value toward 4 indicates negative autocorrelation[53,54].

The tabulated upper and lower bound values of Durbin–Watson were considered to test the hypothesis of zero autocorrelation against the positive and negative autocorrelations. In the present study, the DW values

Table 4. Observed and predicted activity of the significant models.

Compound code	Activity 1 [pIC$_{50(FT)}$]	Model 1	Model 2	Activity 2 [pIC$_{50(QG56)}$]	Model 3	Model 4	Model 5
8A	6.77	6.81	6.68	5.72	6.22	6.25	5.90
8B	6.44	6.62	6.63	5.83	6.22	6.25	5.60
8C	6.44	6.61	6.62	—	6.28*	6.31*	5.01*
8D	—	7.37*	7.41*	5.20	5.15	5.13	5.07
8E	6.07	6.85	6.75	—	7.30*	7.40*	7.09*
8F	7.52	7.19	7.35	6.26	6.17	6.20	6.89
8G	6.60	7.27	7.17	5.30	6.03**	6.06**	6.05**
8H	8.19	8.05**	8.07**	5.83	6.02**	6.03**	6.30**
8I	8.48	7.85	7.96	7.05	6.97	7.04	7.07
8J	8.07	8.26	8.19	7.41	7.02	7.10	7.31
8K	8.55	8.70	8.80	7.64	7.62	7.73	7.75
8L	7.96	7.88	7.81	7.63	7.42**	7.33**	7.18**
8M	8.20	7.81	7.90	6.84	6.96	7.03	7.01
8N	8.40	8.26	8.19	7.44	7.03	7.11	7.36
8O	8.47	8.70	8.80	7.49	7.62	7.73	7.80
8P	8.14	7.81	7.90	6.80	6.96	7.03	7.07
8Q	8.49	8.29	8.04	7.44	7.62	7.73	7.30
8R	9.05	8.96	8.90	8.09	8.30	7.74	8.10
8S	8.70	9.12	9.10	7.82	7.63	7.73	7.32
8T	9.00	9.26	9.28	7.77	7.48	7.04	7.44
8U	7.96	8.12**	8.17**	6.85	6.98	7.05	7.14
8V	8.59	8.85**	8.80**	—	8.07*	7.83*	9.58*
8W	8.19	8.00	8.01	7.13	7.26**	6.97**	7.08**
8X	8.62	8.98**	9.06**	7.84	7.72	7.83	7.71
11A	8.82	7.99	8.13	8.23	7.79	7.73	8.18
11B	7.31	7.40**	7.38**	6.69	6.89**	7.13**	7.18**
11C	9.10	8.65**	8.52**	8.96	8.24**	8.28**	8.19**
11D	8.92	9.03	8.93	8.82	8.64	8.64	8.60
11E	9.15	9.51	9.58	8.70	9.25	9.28	9.04
11F	8.96	8.62	8.69	8.70	8.59	8.58	8.31

Activity 1: farnesyltransferase inhibitory activity; activity 2: antiproliferative activity (QG56).
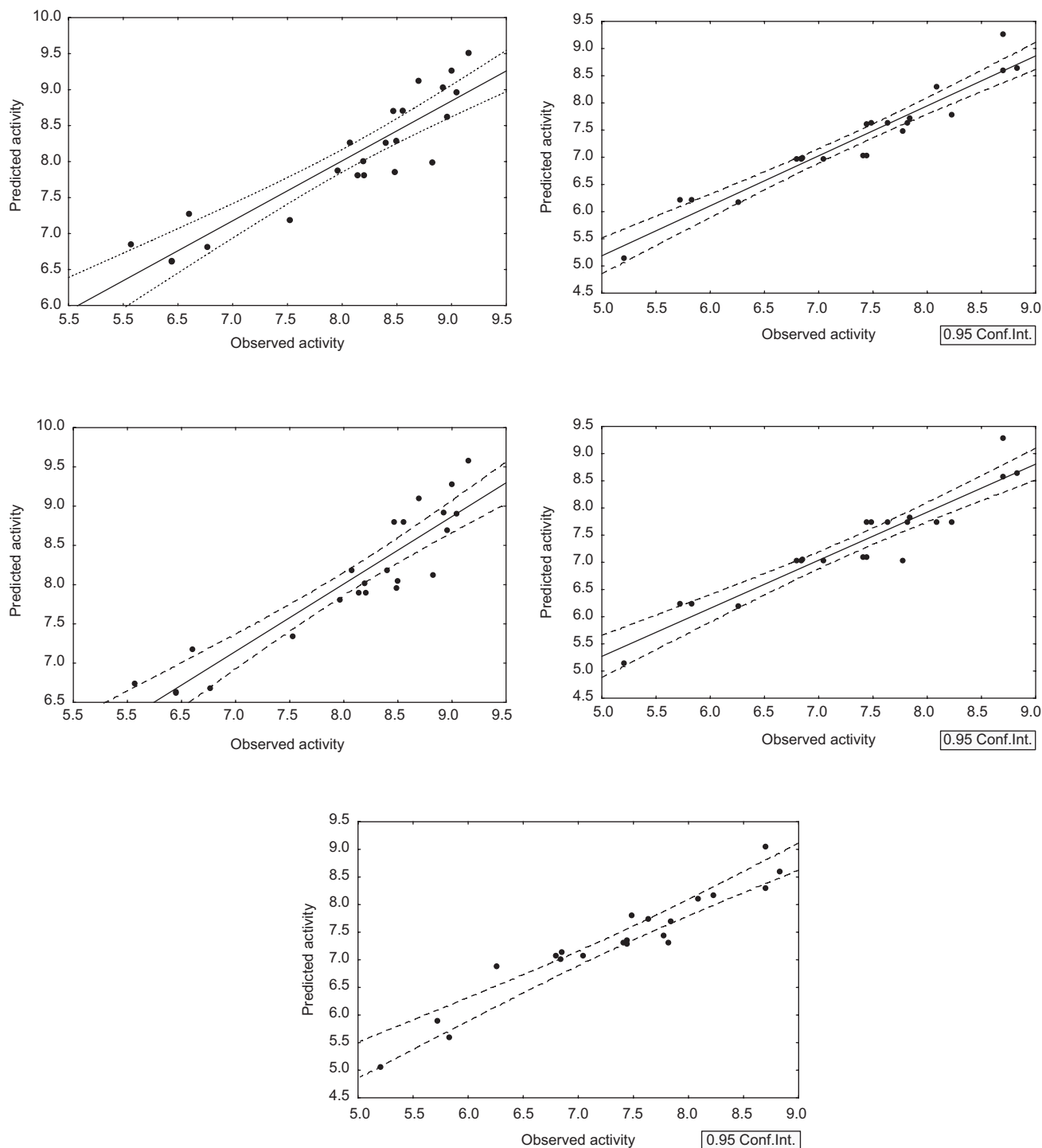*Inactive compounds.
**Test set compounds.

Figure 2. Graph 1–5: observed and predicted activity of the models. (A) Graph 1: Observed versus predicted activity of model 1 ($r=0.9126$). (B) Graph 2: Observed versus predicted activity of model 2 ($r=0.9269$). (C) Graph 3: Observed versus predicted activity of model 3 ($r=0.9588$). (D) Graph 4: Observed versus predicted activity of model 4 ($r=0.9403$). (E) Graph 5: Observed versus predicted activity of model 5 ($r=0.9599$).

are closer to 2 which show that the values are above the positive autocorrelation and below the negative autocorrelation of the tabulated upper and lower bound value at 5% significance level (Table 5). The following criteria were adopted to check the serial correlation of the models[55,56].

- If the value is less than the lower limit then the hypothesis is in favour of positive first-order correlation.
- If the value lies between the lower and upper limits then the test is inconclusive.
- If the value is above the upper limit, there is no autocorrelation or may be negative autocorrelation.

Table 5. Redundancy (VIF, tolerance) and Durbin–Watson statistics of the QSAR models.

Model	Descriptors	Tolerance	R^2	VIF	DW Calculated	DW Tabulated
Model 1	PEOE_RPC+	0.9570	0.0430	1.0449	2.3500	1.078–1.660
	SMR_VSA3	0.9570	0.0430	1.0449	1.6500*	
Model 2	b_triple	0.9965	0.0035	1.0035	2.1059	1.078–1.660
	PEOE_RPC+	0.9965	0.0035	1.0035	1.8941*	
Model 3	BCUT_SlogP_3	0.7716	0.2284	1.2960	1.8337	1.026–1.669
	VSA_pol	0.7716	0.2284	1.2960		
Model 4	BCUT_SlogP_3	0.7773	0.2227	1.2865	2.0252	1.026–1.669
	VSA_acc	0.7773	0.2227	1.2865	1.9748*	
Model 5	PEOE_VSA + 4	0.9996	0.0004	1.0004	2.3965	1.026–1.669
	logP(o/w)	0.9996	0.0004	1.0004	1.6035*	

*The value obtained from subtracting DW value from 4 (4–d).

DW, Durbin–Watson; QSAR, quantitative structure-activity relationships; VIF, variable inflation factor.

- If the test statistics >2, the quantity 4–d is computed and the values with the upper and lower limit of positive autocorrelation compared.

Discussion of descriptors in the QSAR model
FTase inhibitory activity

Model 1 is a biparametric model, built with a subdivided surface area descriptor (SMR_VSA3) and a partial charge descriptor (PEOE_RPC+). The subdivided surface area descriptors are based on an approximate accessible VSA calculation (in $Å^2$) for each atom, v_i along with other atomic property, P_i. The v_i-values are calculated using a connection table approximation. The properties (P_i) of small molecules can be calculated as the sum of the contributions of each of the atoms in the molecule as per Equation 6.

$$P\_VSA_k = -V_i\delta[P_i \in (a_k-1, a_k)] \quad k = 1,2,3, \&, n \qquad (6)$$

where $a_o < a_k < a_n$ are interval boundaries such that (a_o, a_n) bound are values of P_i in any molecule. Each VSA type descriptor can be characterized as the amount of surface area with P in a certain range[37,57]. SMR_VSA3 is defined to be the sum of the v_i over all atoms i. P_i denotes the contribution to molar refractivity for atom i as calculated in the SMR descriptor, calculated in a specified range, from 0.35 to 0.39. This descriptor intends to reflect the polarizability and the atomic contribution to the molar refractivity[42]. The positive sign of the coefficient in this descriptor suggests that the polarizability on the VSA of the molecule is favourable for the FTase inhibitory activity. It reveals that the inhibitors should have polarizable groups in their structure to interact with the receptor (FTase) significantly. This point emphasizes the polar nature of the binding pocket of FTase and the importance of polar ligand-receptor interactions for a good activity.

The partial charge descriptor (PEOE_RPC+) provides the largest positive q_i divided by the sum of the positive q_i (the relative positive partial charge). The Partial Equalization of Orbital Electronegativities (PEOE) is a method of calculating atomic partial charges, in which charge is transferred between bonded atoms until equilibrium. The amount of charge transferred at each iteration is damped with an exponentially decreasing scale factor to guarantee convergence. The amount of charge transferred, dq_{ij}, between atoms i and j when $X_i > X_j$ is

$$dq_{ij} = \frac{\left(\frac{1}{2^k}\right)(X_i - X_j)}{X_j^+}, \qquad (7)$$

where X_j^+ is the electronegativity of the positive ion of atom j, X_i is the electronegativity of atom i (quadratically dependent on partial charge) and k is the iteration number of the algorithm. PEOE is the electronegativity concept as per the following equation.

$$\chi_v = \frac{1}{2}\left(I_v + E_v\right) \qquad (8)$$

In this equation, the electronegativity is related to its ionization potential, I, and its electron affinity E. The electronegativity of an atom further depends on the charge of other atoms in this orbital and also the charge of the same atom in other orbitals[41,58,59]. The coefficient in the descriptor carries a negative sign, which shows that the relative positive charge of the molecules is detrimental for the FTase inhibitory activity. This result illustrates the importance of positively-charged groups on the enzyme active site for the interaction with this series of molecules. In fact, the two available X-ray crystallographic structures for molecules in this series (PDB code 2ZIR and 2ZIS for molecules 8k and 8w, respectively[38]) show the existence of particularly short Zn-N bond-lengths[60] between these molecules and the catalytically relevant Zn(II) metal ion in the FTase active site. The positively-charged Arg202β amino acid residue represents another positively-charged potential point of interaction for this series of molecules, well portrayed for molecules 8k and 8w in the corresponding X-ray structures, for which an interaction with the R1 substituent has been observed[38]. This observation could also partially explain, from an atomistic point of view, the very good correlation obtained for this descriptor in the QSAR analysis. The important role of this amino acid residue for molecule binding to

the FTase active site has been previously demonstrated in structural and mutagenesis studies[61–65], and was the subject of particular attention in some of our molecular dynamics studies[29], suggesting that against natural peptidic CAAX substrates, this residue establishes important interaction with the negatively charged terminal carboxylate group, and has a very important contribution for enzyme-substrate affinity. In addition to this amino acid residue, several other positively-charged amino acid residues have been previously implicated in substrate/inhibitor binding[61,62,65]. Examples include Lys164α, Arg291β and Lys294 β. However, for this particular set of molecules no direct implication by these residues could be inferred in the present study.

Model 2 is composed of an atom and bond count descriptor (b_triple) and the partial charge descriptor (PEOE_RPC+). The atom and bond count descriptor signify the number of triple bonded groups in the molecules and the positive contribution of the descriptor suggests that the presence of the triple bond is favourable for the FTase inhibitory activity. The data set shows that the CN group is the only triple bonded group present in the molecules, which suggests that the presence of CN groups in the molecules is favourable for FTase inhibitory activity. It is evidenced by the compounds in the data set that those with triple bonded groups (CN) (8h–8x and 11a–11f) possess significant activity, more than the other compounds. Some compounds such as 8x and 11d–11f with 2 or CN groups have considerable increase in FTase inhibitory activity. Structural analysis of the available X-ray crystallographic structures on molecules on this series of compounds[38] have suggested that the hydrogen bonding between the cyano group (triple bonded group) on the A-ring (R1) and Arg202β improved the enzyme inhibitory activity, even though for the interaction between the cyano group on the B-ring (R3 site) and the enzyme no obvious interactions had been noticed, with the two X-ray structures available for molecules in the series suggesting that the B-ring interacts with the active-site amino acid residues Leu96β, Trp102β, and Tyr361β and with the end of the isoprenoid portion of the FPP molecule.

The second descriptor, relative partial positive charge descriptor (PEOE_RPC+), suggests that the partial positive charges in the molecules are detrimental for the activity and has already discussed in the model 1. This biparametric model indicates that the presence of triple bonded groups and reduced partial positive charge are favourable for the FTase inhibitory activity. The models 1 and 2 were developed with the positively contributed SMR_VSA3 and b_triple descriptors, respectively and the negatively contributed PEOE_RPC+ descriptor. Note that both SMR_VSA3 and b_triple descriptors reflect the beneficial effect of polar groups in the ligand for the affinity for FTase.

In order to interpret the effect of the descriptors in a given compound for FTase inhibitory activity, we have multiplied the descriptors with their respective regression coefficients [PEOE_RPC+ × regression coefficient (PEOE_RPC + RC) and SMR_VSA3 × regression coefficient (SMR_VSA3RC); Supplementary Table S3].

For the compounds studied PEOE_RPC + RC varies between a minimum of 3.0698 (compound 8t) and a maximum of 4.50040 (compound 8c), while SMR_VSA3RC changes between 0.4932 (8b, 8c, and 8f) and 3.0636 (compound 11e). As both contributions differ in sign, the module of the ratio PEOE_RPC + RC/SMR_VSA3RC was determined to asses for each individual compound the relative importance of the contributions arising for each compound in the global molecule activity. For all compounds the contribution from the PEOE_RPC + RC is dominant, varying between a maximum ratio of 9.1315 (compound 8c) and a minimum ratio of 1.3628 (11e) (due to the negative contribution of the PEOE_RPC + RC term, the lower the value, the more favourable the activity). These conclusions illustrate the high importance of positive partial charge in the molecule (negative contribution), compared with polarizability and the atomic contribution to the molar refractivity for ligand activity.

The compounds (8a–8g) have PEOE_RPC + RC values between 3.88 and 4.50 and the value for 8w is 3.40. The other compounds in the series have the values <4.26, showing that the decreased PEOE_RPC + RC value of compound 8w is due to the small value of PEOE_RPC+. The SMR_VSA3RC value for the compounds 8a–g is between 0.49 and 0.72 while 8w is 0.78. All the other compounds in the series have SMR_VSA3RC values above 1.35. Compound 8w has a comparatively higher SMR_VSA3RC value (0.7827) and lower PEOE_RPC + RC value (3.3985) than the low active compounds in the series (PEOE_RPC + RC/SMR_VSA3RC ratio of 4.3418).

Other active compounds in the series posses proportionally higher SMR_VSA3RC and smaller PEOE_RPC + RC values. Compound 11e, the most active compound in the series, has values of 3.06 and 4.18 for SMR_VSA3RC and PEOE_RPC + RC, respectively, resulting a comparatively low ratio of only 1.3628.

Compounds 8k, 8o, 8r, 8s, 8t, 8v, 8x, 11d, and 11f have also rather low ratio values, and display also very significant inhibitory activities. The low active compounds (8a–g) have ratio values >5.64 and the other active compounds have ratio values <4.34. The fact that the entire most active compounds in the series (8k, 8o, 8r, 8s, 8t, 8x, 11d, 11e, and 11f) have PEOE_RPC + RC/SMR_VSA3RC below 2 clearly illustrates the importance of the PEOE_RPC+ and SMR_VSA3 terms for FTase inhibitory activity.

In model 2, the b_triple descriptor contributes positively along with the negatively contributing PEOE_RPC+ descriptor. This suggests, that the presence of triple bonds (CN groups) in the molecule is favourable for the FTase inhibitory activity. In the data set, all the compounds except 8a–g and 8w have CN groups in their structure. Compound 8w does not have a triple bonded group (CN) in its structure, but has significant FTase inhibitory activity experimentally. This may be due to the influence of the other physicochemical properties of the compounds. It suggests that the presence of a CN (triple bond) is

favourable for the activity but it does not fully determine the FTase inhibitory activity of the compounds.

In order to find out the role of b_triple and PEOE_RPC+ descriptors of the compounds for FTase inhibitory activity, the product of the descriptors with their respective regression coefficients [PEOE_RPC+ × regression coefficient (PEOE_RPC + RC1) and b_triple × regression coefficient (b_tripleRC)] were calculated.

The calculated PEOE_RPC + RC1 values varies between a minimum of 3.8805 (compound 8t) and a maximum of 5.6935 (compound 8c), while b_tripleRC value for the compounds do not have CN groups posses 0 and the CN group containing compounds have b_tripleRC values of 0.8488 (single CN group), 1.6976 (two CN groups) (8k, 8o, 8x, 11d, and 8f) and 2.5464 for compound 11e (posses three CN groups). As both contributions differ in sign, the module of the ratio PEOE_RPC + RC1/b_tripleRC was determined to asses for each individual compound the relative importance of the contributions arising for each compound in the FTase inhibitory activity. For all compounds the contribution from the PEOE_RPC + RC1 is dominant, varying between a maximum ratio of 6.3527 (compound 8h) and a minimum ratio of 2.0727 (11e) (due to the negative contribution of the PEOE_RPC + RC1 term, the lower the value, the more favourable the activity). But the compounds 8a–g and 8w have a value of 0 because the b_tripleRC value of these compounds also 0. These conclusions illustrate the importance of positive partial charge in the molecule (negative contribution), along with the triple bonded group (CN) for ligand activity.

The compounds (8a–8g) have PEOE_RPC + RC1 values between 5.69 and 4.90 and the value for 8w is 4.30. The other compounds in the series have values <5.39. The decreased PEOE_RPC + RC1 value of compound 8w is due to the small value of PEOE_RPC+. The active compounds in the series posses proportionally higher b_tripleRC and smaller PEOE_RPC + RC1 values. Compound 11e, the most active compound in the series, has values of 2.5464 and 5.2779 for b_tripleRC and PEOE_RPC + RC1, respectively, resulting a comparatively low ratio of only 2.0727.

Compounds 8k, 8o, 8r, 8s, 8t, 8v, 8x, 11d, and 11f have also rather low ratio values, and display also very significant inhibitory activities. The low active compounds (8a–g) have ratio values 0 (because of zero b_tripleRC values). The fact that all the most active compounds in the series (8k, 8o, 8r, 8s, 8t, 8x, 11d, 11e, and 11f) have PEOE_RPC + RC1/b_tripleRC below 5 clearly illustrates the importance of the PEOE_RPC+ and b_triple terms for FTase inhibitory activity.

Antiproliferative activity (QG56)

Model 3 was developed with the pharmacophore atom type (vsa_pol) and the adjacency and distance matrix (BCUT_SlogP_3) descriptors. The pharmacophore atom type descriptor is the VSA (P_VSA) descriptors that considers only the heavy atoms of a molecule and assign a type to each atom. Therefore, hydrogen atoms are suppressed during the calculation. The feature set in this

type includes donor, acceptor, polar (both donor and acceptor), positive (base), negative (acid), hydrophobes, and others. The vsa_pol is an approximation to the sum of VSA (Å^2) of polar atoms (atoms that are both hydrogen bond donors and acceptors[42]). It reveals that the polar groups, such as hydrogen bond donor and acceptor groups, in the van der Waals surface of the molecule are needed for hydrogen bonding with the polar groups present in the active site.

The other descriptor is an adjacency and distance matrix descriptor (BCUT_SlogP_3), in which the BCUT metrics are an extension of Burden's parameters. These are based on a combination of the atomic number for each atom and a description of the nominal bond-type for adjacent and non-adjacent atoms, and incorporate both connectivity information and atomic properties (e.g. atomic charge, polarizability, hydrogen bond abilities) relevant to intermolecular interactions[66]. Some BCUT metrics can be generated, based on the connectivity and atomic information, and on the scaling factors controlling the relative balance of these two kinds of information. It can also capture sufficient structural features of molecules to yield useful measurements of molecular diversity[67]. The Burden's parameter BCUT_SlogP_3 has a BCUT descriptor that uses the atomic contribution to logP (calculated with the Wildman and Crippen SlogP method) instead of the partial charges[55] The partition coefficient (logP) of small molecules can be calculated as the sum of the contributions of each of the atoms in the molecule.

$$P_{calc} = \Sigma n_i a_i \tag{9}$$

where P_{calc} is the property to be calculated (logP), n_i is the number of atoms type i present in the molecule and a_i is the contribution for atoms of type i[39,57]. The negative contribution of the descriptor explains that the hydrophobicity of the molecules should be small to favour the antiproliferative activity of the benzofuran derivatives. It is in agreement with the fact that the active site of the receptor is hydrophilic and that passive diffusion through the cell membrane is not limiting on this series of analogues.

Model 4 also has the same type descriptors as model 3, namely the pharmacophore type descriptor (vsa_acc) and the adjacency and distance matrix (BCUT_SlogP_3) descriptor. In this model, the pharmacophore type descriptor is different from the earlier model. The pharmacophore descriptor (vsa-acc) is a VSA descriptor, which calculates the approximation to the sum of VSA (Å^2) of pure hydrogen bond acceptors (not counting acidic atoms and atoms that are both hydrogen bond donors and acceptors such as -OH). The descriptor reveals that the compounds that have pure hydrogen bond acceptor groups are favourable for the antiproliferative activity. Models 3 and 4 explain that the hydrogen bonding groups are necessary to establish stable drug receptor interactions by hydrogen bonding. The significance of the BCUT_SlogP_3 has already been explained in model 3.

Model 5 was found to be the most important two-variable correlation for modelling antiproliferative activity of benzofuran derivatives. It describes the role of the partial charge descriptors [PEOE_VSA4) and the physical descriptor (logP(o/w)] for the antiproliferative activity of benzofuran derivatives. The partial charge descriptor (PEOE_VSA4) provides the partial charge of an atom (q_i) on the VSA (Å^2) of an atom (v_i). It is calculated in a range of q_i (partial charge of an atom i) value between 0.20 and 0.25. This PEOE_VSA4 descriptors uses PEOE method for the partial charge calculation on the VSA[42]. The descriptor PEOE_VSA4 has a negative coefficient in the model, suggesting that the positive charged groups present on the VSA of the molecule are detrimental for the activity. It reveals that a decrease in the descriptor value may improve the antiproliferative activity of the compounds. It highlights the importance of positively charge groups in the active site, showing also that active molecules should bear negative charges (or at least not bears positive charges).

The second molecular descriptor incorporated into the model is the logarithm of the octanol/water partition coefficient [logP(o/w)]. The descriptor logP(o/w) is a measure of overall hydrophobicity of the molecule and therefore the negative coefficient associated with this term implies that a decrease in the lipophilicity of the molecule will cause a corresponding increase in the antiproliferative activity of the derivatives.

The models 3–5 describe the physicochemical properties responsible for the antiproliferative activity on QG56 cell lines. In these models, the hydrophobicity (BCUT_SlogP_3 and logP(o/w) and polar descriptors (vsa_acc, vas_pol and PEOE_VSA4) contributed for the activity prediction. The hydrophobic descriptors in the models are not merely a determinant for effective binding to the receptors. In fact, they are also important for the molecule to reach its target (cross the biological barriers, avoid sequestrion by other enzymes or receptors, avoid premature metabolization, etc.).

The calculated proportionality of the descriptors (BCUT_SlogP_3RC1 (model 3)/vsa_polRC and BCUT_SlogP_3RC2 (model 4)/vsa_accRC) shows some interesting trends regarding the relative contribution of the descriptor values for the antiproliferative activity (Table S3). The ratio value of BUCT_SlogP_3RC1/vsa_polRC varies between 119.36 and 119.56 for compounds 8a–c, 8f, 8g, while between the other compounds it changes as much as between 29.69 and 62.12. In a similar fashion, the ratio BCUT_SlogP_3RC2/vsa_accRC varies between 278.33 and 400.79 for compounds 8a–c, 8f, 8g, and changes between 38.60 and 120.41 for the other compounds.

Compounds 11d–f have considerably smaller ratio values in the series (37.08, 29.69, and 38.69 for BCUT_SlogP_3RC1/vsa_polRC and 51.62, 38.25, and 54.73 for BCUT_SlogP_3RC2/vsa_accRC) and display very significant activities in nM range. In the same way, compound 8v also has significant value for both descriptors i.e. 35.33 for vsa_polRc ratio and 54.12 for

vsa_accRC ratio. Unfortunately, that compound was not tested experimentally. In addition, for both cases (ratios BCUT_SlogP_3RC1/vsa_polRC and BCUT_SlogP_3RC2/vsa_accRC), compounds 8d and 8e have large ratio values, but they are experimentally inactive or display very low activity, suggesting that other structural properties not included in the models may cause its low antiproliferative activity.

Model 5 describes the importance of log partition coefficient and PEOE_VSA4 for antiproliferative activity. In this model, the descriptors are negatively contributing for the antiproliferative activity. The calculated logP(o/w) RC of the compounds shows that compounds 8a–d have values >9 and that the same compounds have PEOE_VSA4RC values of 0.55. All other compounds in the series posses logP(o/w)RC values <8 and have the same PEOE_VSA4RC value of the abovementioned compounds. The exceptions are compounds 8r, 8s, and 11a–f. Compound 8r and 8t have small logP(o/w)RC values of 6.0 and 6.8, respectively. In these studies, compounds 8d and 8e are predicted as active compounds but experimentally they have low active/inactive. Structurally, these compounds posses COO-*t*-Bu and COOH groups in their structure (in A-ring). None of the compounds in the series has these groups in their structure, in addition to these two compounds. So the bulkiness or other physicochemical effect of these compounds may be preventing the compounds to reach the site for binding.

The higher complexity associated with the antiproliferative activity against cancer cell lines measurement, when compared against the much more specific enzyme inhibitory activity analysis, prevent the establishment of a direct relationship between the results and the possible conformation of the molecules in the FTase active site; as in antiproliferative activity analysis the FTase-molecule interactions may be just one in a number of factors responsible for the overall activity observed. Nevertheless, the main conclusions derived from this analysis are in line with the results obtained for the FTase inhibitory activity, and with the structural analysis made from the X-ray crystallographic structures of molecules in this series bound to FTase, suggesting that in spite of the natural differences observed, FTase-molecule binding is mainly responsible for the antiproliferative activity.

The results obtained for the contribution of each descriptor for the prediction (β-coefficient) of the activities show that the VSA (P_VSA) descriptors such as pharmacophore descriptors [VSA_pol (0.6050) and VSA_acc (0.5640)], subdivided surface area descriptors [SMR_VSA3 (0.7340)] are the major contributors for the activities (Figure 2). Except for model 2, the remaining models have P_VSA as one of the contributing descriptors to explain structural features of the inhibitors for the FTase inhibitory and antiproliferatve activity. The study provides the information that the negatively charged descriptors and hydrogen bond acceptor and donor groups on the VSA are favourable for the activities,

while the lipophilicity and polarizability on the VSA of the molecule are detrimental for the activities.

Conclusion

From the study, it is concluded that the QSAR models developed by MLR methods and PLS methods are statistically significant. The significant models were selected by correlation coefficient, Fischer value, *t* value and other statistical relevance of the contributed descriptors. The stability and predictive power of the models were evaluated by parameters such as cross-validated correlation coefficient, Cook's distance and Mahalanobis distance, VIF and DW, obtained from the validation studies and which can be used for further study. The significant QSAR model developed from the studies shows that the subdivided surface area (SMR_VSA3), the atom and bond count descriptors and the partial positive charge descriptors (PEOE_RPC+) contribute for the FTase inhibitory activity, either positively or negatively. It reveals that the molar refractivity (polarizability) on the VSA and the number of triple bonds in the molecules are favourable for the FTase inhibitory activity and the partial positive charge of the molecules is unfavourable for the FTase activity. The results obtained from the antiproliferative activity shows that the partition coefficient [BCUT_SlogP and logP(o/w)] and PEOE_VSA4 of the molecules are detrimental for the activity, while the hydrogen bond acceptor and donor groups on the van der Waals surface of the molecules are favourable for the antiproliferative activity. The experimental results obtained from both evaluations (FTase inhibitory activity and QG56 cancer cell lines) show that this series of compounds have significant activity against FTase inhibitory activity, as well as having significant anticancer activity. The study shows that the P_VSA descriptors (subdivided surface area, pharmacophore type) are the main contributors for the inhibitory activities (FTase and antiproliferative), along with the atom and bond count and the partial charge descriptors, for the FTase inhibitory activity, and the adjacency and distance matrix descriptors and the conformational charge descriptors, for antiproliferative activity. The study highlights the importance of the positively-charged groups on the active site of the enzyme or receptor (possibly the Arg202β amino acid residue and the Zn(II) ion), and the existence of hydrogen bond donor and acceptor groups on the VSA of the molecules, and of negatively charged groups in their structures. The proposed models can be used for the design and development of novel FTase inhibitors as anticancer molecules, particularly when complemented with detailed structural analysis of the available X-ray structures of molecules in the series considered. These models will also be applied in future studies considering also protein-ligand docking, molecular dynamics simulations, and detailed free energy calculations in order to develop new inhibitors for this important enzyme.

Supporting information

The correlation matrix between the physicochemical descriptors and with the biological activity (FTase inhibitory activity and antiproliferative activity), (Table S1 and S2) and the calculated physicochemical descriptors values, which are contributed in the models (Table S3) are provided as supporting materials.

Declaration of interest

References

1. Bell IM. Inhibitors of farnesyltransferase: a rational approach to cancer chemotherapy? J Med Chem 2004;47:1869–1878.
2. Hancock JF, Magee AI, Childs JE, Marshall CJ. All Ras proteins are polyisoprenylated but only some are palmitoylated. Cell 1989;57:1167–1177.
3. Jackson JH, Cochrane CG, Bourne JR, Solski PA, Buss JE, Der CJ. Farnesyl modification of Kirsten-Ras Exon 4B protein is essential for transformation. Proc Natl Acad Sci USA 1990;87:3042–3046.
4. Bell IM. Inhibitors of protein prenylation 2000. Expert Opin Ther Patents 2000;10:1813–1831.
5. Tanaka R, Rubio A, Harn NK, Gernert D, Grese TA, Eishima J et al. Design and synthesis of piperidine farnesyltransferase inhibitors with reduced glucuronidation potential. Bioorg Med Chem 2007;15:1363–1382.
6. Ganguly AK, Doll RJ, Girijavallabhan VM. Farnesyl protein transferase inhibition: a novel approach to anti-tumor therapy. the discovery and development of SCH 66336. Curr Med Chem 2001;8:1419–1436.
7. Equbal T, Silakari O, Rambabu G, Ravikumar M. Pharmacophore mapping of diverse classes of farnesyltransferase inhibitors. Bioorg Med Chem Lett 2007;17:1594–1600.
8. Venet M, End D, Angibaud P. Farnesyl protein transferase inhibitor ZARNESTRA R115777 - history of a discovery. Curr Top Med Chem 2003;3:1095–1102.
9. Buckner FS, Yokoyama K, Nguyen L, Grewal A, Erdjument-Bromage H, Tempst P et al. Cloning, heterologous expression, and distinct substrate specificity of protein farnesyltransferase from *Trypanosoma brucei*. J Biol Chem 2000;275:21870–21876.
10. Ohkanda J, Buckner FS, Lockman JW, Yokoyama K, Carrico D, Eastman R et al. Design and synthesis of peptidomimetic protein farnesyltransferase inhibitors as anti-*Trypanosoma brucei* agents. J Med Chem 2004;47:432–445.
11. Nguyen UT, Cramer J, Gomis J, Reents R, Gutierrez-Rodriguez M, Goody RS et al. Exploiting the substrate tolerance of farnesyltransferase for site-selective protein derivatization. Chembiochem 2007;8:408–423.
12. Basso AD, Kirschmeier P, Bishop WR. Lipid posttranslational modifications. Farnesyl transferase inhibitors. J Lipid Res 2006;47:15–31.
13. Sousa SF, Fernandes PA, Ramos MJ. Farnesyltransferase inhibitors: a detailed chemical view on an elusive biological problem. Curr Med Chem 2008;15:1478–1492.
14. Gramatica P. Principles of QSAR models validation: internal and external. QSAR Comb Sci 2007;26:694–701.

15. Sousa SF, Fernandes PA, Ramos MJ. Farnesyltransferase: theoretical studies on peptide substrate entrance—thiol or thiolate coordination? J Mol Struct (Theochem) 2005;729:125–129.

16. Sousa SF, Fernandes PA, Ramos MJ. Farnesyltransferase—new insights into the zinc-coordination sphere paradigm: evidence for a carboxylate-shift mechanism. Biophys J 2005;88:483–494.

17. Sousa SF, Fernandes PA, Ramos MJ. Theoretical studies on farnesyltransferase: the distances paradox explained. Proteins 2007;66:205–218.

18. Sousa SF, Fernandes PA, Ramos MJ. Theoretical studies on farnesyl transferase: evidence for thioether product coordination to the active-site zinc sphere. J Comput Chem 2007;28:1160–1168.

19. Park HW, Boduluri SR, Moomaw JF, Casey PJ, Beese LS. Crystal structure of protein farnesyltransferase at 2.25 angstrom resolution. Science 1997;275:1800–1804.

20. Long SB, Casey PJ, Beese LS. Reaction path of protein farnesyltransferase at atomic resolution. Nature 2002;419:645–650.

21. Long SB, Hancock PJ, Kral AM, Hellinga HW, Beese LS. The crystal structure of human protein farnesyltransferase reveals the basis for inhibition by CaaX tetrapeptides and their mimetics. Proc Natl Acad Sci USA 2001;98:12948–12953.

22. Tobin DA, Pickett JS, Hartman HL, Fierke CA, Penner-Hahn JE. Structural characterization of the zinc site in protein farnesyltransferase. J Am Chem Soc 2003;125:9962–9969.

23. Pompliano DL, Rands E, Schaber MD, Mosser SD, Anthony NJ, Gibbs JB. Steady-state kinetic mechanism of Ras farnesyl:protein transferase. Biochemistry 1992;31:3800–3807.

24. Furfine ES, Leban JJ, Landavazo A, Moomaw JF, Casey PJ. Protein farnesyltransferase: kinetics of farnesyl pyrophosphate binding and product release. Biochemistry 1995;34:6857–6862.

25. Fu HW, Beese LS, Casey PJ. Kinetic analysis of zinc ligand mutants of mammalian protein farnesyltransferase. Biochemistry 1998;37:4465–4472.

26. Hightower KE, De S, Weinbaum C, Spence RA, Casey PJ. Lysine(164) alpha of protein farnesyltransferase is important for both CaaX substrate binding and catalysis. Biochem J 2001;360:625–631.

27. Pickett JS, Bowers KE, Hartman HL, Fu HW, Embry AC, Casey PJ et al. Kinetic studies of protein farnesyltransferase mutants establish active substrate conformation. Biochemistry 2003;42:9741–9748.

28. Sousa SF, Fernandes PA, Ramos MJ. Enzyme flexibility and the catalytic mechanism of farnesyltransferase: targeting the relation. J Phys Chem B 2008;112:8681–8691.

29. Sousa SF, Fernandes PA, Ramos MJ. Molecular dynamics analysis of farnesyltransferase: a closer look into the amino acid behavior. Int J Quant Chem 2008;108:1939–1950.

30. Sousa SF, Fernandes PA, Ramos MJ. Molecular dynamics simulations on the critical states of the farnesyltransferase enzyme. Bioorg Med Chem 2009;17:3369–3378.

31. Sousa SF, Fernandes PA, Ramos MJ. The search for the mechanism of the reaction catalyzed by farnesyltransferase. Chemistry 2009;15:4243–4247.

32. Ho MH, De Vivo M, Dal Peraro M, Klein ML. Unraveling the catalytic pathway of metalloenzyme farnesyltransferase through QM/MM computation. J Chem Theor Comput 2009;5:1657–1666.

33. Rida SM, El-Hawash SA, Fahmy HT, Hazzaa AA, El-Meligy MM. Synthesis of novel benzofuran and related benzimidazole derivatives for evaluation of *in vitro* anti-HIV-1, anticancer and antimicrobial activities. Arch Pharm Res 2006;29:826–833.

34. Galal SA, Abd El-All AS, Abdallah MM, El-Diwani HI. Synthesis of potent antitumor and antiviral benzofuran derivatives. Bioorg Med Chem Lett 2009;19:2420–2428.

35. Thai KM, Ecker GF. Similarity-based SIBAR descriptors for classification of chemically diverse hERG blockers. Mol Divers 2009;13:321–336.

36. Labute P. A widely applicable set of descriptors. J Mol Graph Model 2000;18:464–477.

37. Balaji S, Karthikeyan C, Moorthy NSHN, Trivedi P. QSAR modelling of HIV-1 reverse transcriptase inhibition by benzoxazinones using a combination of P_VSA and pharmacophore feature descriptors. Bioorg Med Chem Lett 2004;14:6089–6094.

38. Asoh K, Kohchi M, Hyoudoh I, Ohtsuka T, Masubuchi M, Kawasaki K et al. Synthesis and structure-activity relationships of novel benzofuran farnesyltransferase inhibitors. Bioorg Med Chem Lett 2009;19:1753–1757.

39. Molecular Operating Environment (MOE). Chemical Computing Group Inc. Montreal, H3A 2R7 Canada, 2007.

40. Systatica. (8.0). StatSoft Inc. Tulsa, OK, USA, 2008.

41. Lin A. QuaSAR-descriptors. Chemical Computing Group Inc. Montreal, H3A 2R7 Canada, 2002.

42. Baurin N, Mozziconacci JC, Arnoult E, Chavatte P, Marot C, Morin-Allory L. 2D QSAR consensus prediction for high-throughput virtual screening. An application to COX-2 inhibition modeling and screening of the NCI database. J Chem Inf Comput Sci 2004;44:276–285.

43. Golbraikh A, Tropsha A. Beware of Q^2! J Mol Graph Model 2002;20:269–276.

44. Eriksson L, Jaworska J, Worth AP, Cronin MT, McDowell RM, Gramatica P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environ Health Perspect 2003;111:1361–1375.

45. Tropsha A, Gramatica P, Gombar VK. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb Sci 2003;22:69–77.

46. Roy PP, Roy K. On some aspectes of variable selected for partial least squares regression models. QSAR Comb Sci 2008;27:302–313.

47. Pratim Roy P, Paul S, Mitra I, Roy K. On two novel parameters for validation of predictive QSAR models. Molecules 2009;14:1660–1701.

48. Cook RD. Influential observations in linear-regression. J Am Stat Assoc 1979;74:169–174.

49. Vilar S, González-Díaz H, Santana L, Uriarte E. QSAR model for alignment-free prediction of human breast cancer biomarkers based on electrostatic potentials of protein pseudofolding HP-lattice networks. J Comput Chem 2008;29:2613–2622.

50. Gnanades R, Kettenri JR. Robust estimates, residuals, and outlier detection with multiresponse data. Biometrics 1972;28:81–124.

51. O'Brien RM. A caution regarding rules of thumb for variance inflation factors. Qual Quant 2007;41:673–690.

52. Neter J, Wasserman W, Kunter MH. Applied linear statistical models: regression, analysis of variance, and experimental designs. 3rd ed. Homewood (IL): Irwin, 1990.

53. Gujarati DN. Qualitative response regression models. In: Basic Econometrics, 4th ed. Boston: McGraw-Hill, 2003.

54. Prasanna S, Manivannan E, Chaturvedi SC. QSAR analyses of conformationally restricted 1,5-diaryl pyrazoles as selective COX-2 inhibitors: application of connection table representation of ligands. Bioorg Med Chem Lett 2005;15:2097–2102.

55. Durbin J, Watson GS. Testing for serial correlation in least squares regression. I. Biometrika 1950;37:409–428.

56. Durbin J, Watson GS. Testing for serial correlation in least squares regression. II. Biometrika 1951;38:159–178.

57. Wildman SA, Crippen GM. Prediction of physicochemical parameters by atomic contributions. J Chem Inf Comput Sci 1999;39:868–873.

58. Mulliken RS. A new electroaffinity scale; together with data on valence states and on valence ionization potentials and electron affinities. J Chem Phys 1934;2:782–793.

59. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—A rapid access to atomic charges. Tetrahedron 1980;36:3219–3228.

60. Tamames B, Sousa SF, Tamames J, Fernandes PA, Ramos MJ. Analysis of zinc-ligand bond lengths in metalloproteins: trends and patterns. Proteins 2007;69:466–475.

61. Long SB, Casey PJ, Beese LS. Cocrystal structure of protein farnesyltransferase complexed with a farnesyl diphosphate substrate. Biochemistry 1998;37:9612–9618.

62. Strickland CL, Windsor WT, Syto R, Wang L, Bond R, Wu Z et al. Crystal structure of farnesyl protein transferase complexed with a

CaaX peptide and farnesyl diphosphate analogue. Biochemistry 1998;37:16601–16611.

63. Kral AM, Diehl RE, deSolms SJ, Williams TM, Kohl NE, Omer CA. Mutational analysis of conserved residues of the beta-subunit of human farnesyl:protein transferase. J Biol Chem 1997;272:27319–27323.

64. Hightower KE, Casey PJ, Fierke CA. Farnesylation of nonpeptidic thiol compounds by protein farnesyltransferase. Biochemistry 2001;40:1002–1010.

65. Dunten P, Kammlott U, Crowther R, Weber D, Palermo R, Birktoft J. Protein farnesyltransferase: structure and implications for substrate binding. Biochemistry 1998;37:7907–7912.

66. Gao H. Application of BCUT metrics and genetic algorithm in binary QSAR analysis. J Chem Inf Comput Sci 2001;41:402–407.

67. González MP, Terán C, Teijeira M, Besada P, González-Moa MJ. BCUT descriptors to predicting affinity toward A3 adenosine receptors. Bioorg Med Chem Lett 2005;15:3491–3495.

RIGHTS LINK