

Predictive QSAR workflow for the *in silico* identification and screening of novel HDAC inhibitors

Georgia Melagraki · Antreas Afantitis ·
Haralambos Sarimveis · Panayiotis A. Koutentis ·
George Kollias · Olga Igglessi-Markopoulou

Received: 4 September 2008 / Accepted: 16 January 2009
© Springer Science+Business Media B.V. 2009

Abstract A linear Quantitative Structure–Activity Relationship (QSAR) is developed in this work for modeling and predicting HDAC inhibition by 5-pyridin-2-yl-thiophene-2-hydroxamic acids. In particular, a five-variable model is produced by using the Multiple Linear Regression (MLR) technique and the Elimination Selection-Stepwise Regression Method (ES-SWR) on a database that consists of 58 recently discovered 5-pyridin-2-yl-thiophene-2-hydroxamic acids and 69 descriptors. The physical meaning of the selected descriptors is discussed in detail. The validity of the proposed MLR model is established using the following techniques: cross validation, validation through an external test set and Y-randomization. Furthermore, the domain of applicability which indicates the area of reliable predictions is defined. Based on the produced model, an *in silico*-screening study

explores novel structural patterns and suggests new potent lead compounds.

Keywords HDAC · Hydroxamic acids · QSAR ·
In silico screening · Histone deacetylases

Introduction

Histone deacetylase (HDAC) inhibitors such as suberoylanilide hydroxamic acid (SAHA) and trichostatin A (TSA), inhibit the activity of the enzymes known as histone deacetylases. These enzymes remove an acetyl group from histones, which allows them to bind DNA and inhibit gene transcription [1,2]. Despite the youth of the field of HDAC inhibitors as therapeutic agents, an impressive body of data describes the ability of these molecules to modulate a wide variety of cellular functions, transcriptional activity in cells, cell cycling, angiogenesis, apoptosis and differentiation which are key components of tumour proliferation [3,4]. HDAC inhibitors may improve the efficacy of existing cancer therapies and, because they target the transcription of specific disease-causing genes, they may offer new therapeutic approaches to cancer therapy [5,6].

The exponential growth in the level of research activity surrounding histone deacetylases (HDACs) witnessed over the past decade has now started to produce success in the clinic, particularly in the field of oncology [7]. SAHA, a hydroxamic acid-containing HDAC inhibitor has been recently approved by the FDA for once—daily oral treatment of advanced cutaneous T-cell lymphoma (CTCL). The anticipation over the next few years is that as first generation HDAC inhibitors produce clinical benefits and second generation inhibitors are rationally designed with improved specificity, this field will emerge as a new treatment for cancer [8].

G. Melagraki · A. Afantitis
Department of ChemoInformatics, NovaMechanics Ltd, Larnaca,
Cyprus

G. Melagraki · A. Afantitis
Cyano Research Corporation Ltd, P.O. Box 28670, 2081 Nicosia,
Cyprus

G. Melagraki (✉) · A. Afantitis · H. Sarimveis ·
O. Igglessi-Markopoulou
School of Chemical Engineering, National Technical University
of Athens, Athens, Greece
e-mail: geomel@mail.ntua.gr

P. A. Koutentis
Department of Chemistry, University of Cyprus, P.O. Box 20537,
1678 Nicosia, Cyprus

A. Afantitis · G. Kollias
Biomedical Sciences Research Center “Alexander Fleming”, Athens,
Greece

For the rational design of novel HDAC inhibitors, quantitative structure–activity relationships (QSAR) models [9–16] and in silico-screening [17–22] could be useful. In the past, several attempts have been made to build QSAR models for HDAC inhibition. Chen et al. applied CoMFA calculations for modeling a series of side-chain analogues of Apicidin natural product [23]. Xie et al. studied a diverse set of HDAC inhibitors from various different literature sources using QSAR and classification techniques [24]. Wang et al. developed a linear QSAR model for a 19 TSA and SAHA-like hydroxamic acids [25]. Liu et al. studied a series of sulfonamide hydroxamic acid HDAC inhibitors applying CoMFA calculations [26]. Guo et al. using docking simulations, CoMFA and CoMSIA analyses, constructed a predictive model for 29 indole amide hydroxamic acid-based HDAC inhibitors [27]. Ragno et al. presented a 3D-QSAR model using a broad molecular diversity training set of HDACIs and applying docking simulations and GRID/GOLPE combinations [28]. More recently Ragno et al. reported a 3D-QSAR study for class II-HDAC inhibitors [29]. Kukalni et al. studied various hydroxamic acid analogues applying CoMFA, CoMSIA and Genetic Function Approximation techniques. These studies lead to the conclusion that molecular shape analysis (MSA), thermodynamic and structural descriptors are important for inhibition of HDACs [30,31]. Jaismal et al. developed QSAR models on a new set of sulfonamide derivatives applying linear free energy related (LFER) approach of Hansch to explain the structural requirements of sulfonamide derivatives for histone deacetylase inhibition [32]. More recently, Vadeilevan et al. studied a total of 20 well-defined HDAC inhibitors to generate pharmacophore models and then a similarity analysis has been carried out to identify possible new chemotypes [33]. Finally, Chen et al. developed a 3D chemical-feature-based QSAR pharmacophore and the interactions between the benzamide MS-275 and HDAC were explored [34].

In this work, a series of 58 5-pyridin-2-yl-thiophene-2-hydroxamic acids with HDAC inhibitory activity, recently discovered by Argenta Discovery Ltd was studied [35,36]. First, a quantitative structure–activity relationship was explored with 69 different physicochemical, topological and structural descriptors being considered as inputs to the model. Among different candidates, a linear five-parameter QSAR model was selected as the most accurate and reliable using a rigorous and systematic variable selection method. A virtual screening study was then conducted to identify novel biologically active patterns by insertion, deletion and substitution of different substitutes of the original molecules. The study led to the identification of novel structures, which are potent HDAC inhibitors according to the QSAR model. The structures were filtered using the domain of applicability of the QSAR model.

Materials and methods

The structures of the 58 compounds together with the corresponding biological activity used in this study are shown in Table 1. The notation (**6a**, **6b**, etc.) corresponds to the studies published by Price et al. [35,36]. Initially, for developing the QSAR model, the structures were fully optimized using the PM6 method [37]. Then 73 physicochemical constants, topological and structural descriptors were calculated using Chem3D [38], the newly released MOPAC2007 [39] and Topix [40] (Table 2).

For evaluation purposes the original data set was partitioned into a training set and a validation set using the popular Kennard and Stones algorithm [41]. This algorithm has been applied with great success in many recent QSAR studies [42–46] and has been highlighted as one of the best ways to build training and test sets [47].

For every structure in the training set, the most appropriate calculated descriptors were selected by using the ES-SWR variable selection method. ES-SWR is a popular stepwise technique that combines Forward Selection (FS-SWR) and Backward Elimination (BE-SWR) [48]. The descriptors were examined in terms of their efficacy to predict the activity of the investigated derivatives and the most statistically significant descriptors were selected. Our first objective was to determine the best variables which produced the most significant linear QSAR models linking compound structure with the HDAC inhibition.

The QSAR model was evaluated for its robustness, accuracy and reliability. To illustrate this, the following evaluation techniques were used: the leave-one-out (LOO) cross-validation procedure, validation through an external test set and Y-randomization [49–53].

Specifically for the external validation based on the validation set, the following criteria were used:

$$R_{\text{pred}}^2 > 0.6 \quad (1)$$

$$\frac{(R_{\text{pred}}^2 - R_o^2)}{R_{\text{pred}}^2} \text{ or } \frac{(R_{\text{pred}}^2 - R_o'^2)}{R_{\text{pred}}^2} \leq 0.1 \quad (2)$$

$$k \text{ or } k' \approx 1 \quad (3)$$

In Eqs. 2 and 3, R_{pred}^2 is the coefficient of determination between experimental values and model predictions on the validation set. Mathematical definitions of R_o^2 , $R_o'^2$, k and k' are based on regression of the observed activities against predicted activities and regression of the predicted activities against observed activities. The definitions of the aforementioned statistical indices are presented in detail in reference 50.

The QSAR model was finally used to identify novel active compounds via an in silico screening procedure, and thus the

Table 1 HDAC inhibitory activity—observed and predicted using Eq. 4

	Compound	IC ₅₀ (μM)	−logIC ₅₀ observed	−logIC ₅₀ predicted (training set)	−logIC ₅₀ predicted (validation set)	Leverage (limit = 0.4)
1	ADS100380	0.750	0.125	0.31		
2	6a	0.153	0.815	1.09		
3	6b	0.100	1.000	1.14		
4	6c	0.034	1.469	1.67		
5	6d	0.022	1.658	1.56		
6	6e	0.020	1.699	1.33		
7	6f	0.021	1.678	1.39		
8 ^a	6g	0.033	1.481		1.20	0.10
9	6h	0.069	1.161	1.63		
10	6i	0.019	1.721	1.73		
11 ^a	6j	0.029	1.538		2.33	0.09
12	6k	0.013	1.886	1.59		
13	6l	0.107	0.971	1.61		
14	6m	0.005	2.301	2.20		
15 ^a	6n	0.011	1.959		2.32	0.08
16	6o	0.006	2.222	2.49		
17	6p	0.005	2.301	1.82		
18	6q	0.048	1.319	2.01		
19	6r	0.009	2.046	2.22		
20 ^a	6s	0.012	1.921		2.27	0.19
21	6t	0.007	2.155	1.70		
22 ^a	6u	0.008	2.097		2.18	0.21
23	6v	0.008	2.097	1.89		
24	6w	0.006	2.222	1.94		
25	3a	2.500	−0.398	−0.20		
26 ^a	3b	0.900	0.046		−0.14	0.37
27 ^a	3c	0.430	0.614		−0.01	0.39
28 ^a	3d	1.130	−0.053		−0.06	0.37
29	3e	1.260	−0.100	−0.26		
30 ^a	3f	0.186	0.730		−0.10	0.36
31	5a	0.016	1.796	1.55		
32 ^a	5b	0.016	1.796		1.17	0.08
33	5c	0.035	1.456	1.46		
34	5d	0.030	1.523	1.78		
35	5e	0.012	1.921	1.66		
36	5f	0.018	1.745	1.35		
37	5g	0.011	1.959	1.57		
38	7a	0.359	0.445	1.11		
39	7b	0.581	0.236	0.68		
40 ^a	9	0.080	1.097		1.24	0.11
41	13a	0.009	2.046	1.67		
42	13b	0.031	1.509	1.62		
43	5h	0.009	2.046	1.90		
44	5i	0.007	2.155	1.50		
45	5j	0.008	2.097	2.29		
46	5k	0.011	1.959	1.90		
47	5l	0.007	2.155	1.76		

Table 1 continued

	Compound	IC ₅₀ (μM)	−logIC ₅₀ observed	−logIC ₅₀ predicted (training set)	−logIC ₅₀ predicted (validation set)	Leverage (limit=0.4)
48 ^a	5m	0.009	2.046		2.14	0.26
49	5n	0.012	1.921	2.03		
50	5o	0.006	2.222	1.95		
51	5p	0.017	1.770	1.79		
52	5q	0.008	2.097	2.12		
53	5r	0.009	2.046	2.10		
54 ^a	5s	0.014	1.854		1.95	0.10
55	5t	0.008	2.097	2.24		
56	5u	0.004	2.398	2.50		
57	13c	0.013	1.886	2.08		
58	13d	0.009	2.046	2.45		

^a Validation set

definition of its domain of applicability is of particular importance. The *Extent of Extrapolation* method was adopted for defining the domain of applicability of the produced method, based on the calculation of the leverages for the components in the available data set as described in reference 49. A comprehensive in silico screening procedure was performed next to determine a variety of potential new lead compounds by introducing structural modifications on the original dataset. Throughout the screening procedure, only the predictions that fall into the domain of applicability were considered reliable.

Results and discussion

The Kennard-Stone algorithm was initially used to partition the dataset of the 58 derivatives into a training set (45 compounds) and a validation set (13 compounds, see Table 1 note). The following MLR QSAR model was developed next, by applying the ES-SWR algorithm on the training data:

$$\log(1/IC_{50}) = 5.69 - 0.0762 \text{ DPL} - 0.000441 \text{ PMIX} - 2.11 \text{ ShpC} + 3.87 \text{ TopoJ} - 4.35 \text{ ChiInf0} \quad (4)$$

$$n = 45, S = 0.34, R^2 = 0.78, \text{RMS} = 0.34, R_{\text{adj}}^2 = 0.75, Q^2 = 0.68, \text{PRESS} = 6.59, F = 26.870$$

It should be noted that the LOO cross-validation methodology, which produces the Q^2 and PRESS statistical indices was performed by applying for each fold the entire variable selection model development procedure. More specifically, for each training subset resulting from the deletion of one training example, the ES-SWR algorithm was applied and the produced model was used to obtain the response of the

single example that was not utilized in the development of this particular model.

The possibility of having included outliers in our dataset was investigated by calculating the standard residuals. Standardized residuals greater than 2.5 or less than −2.5 are considered large and indicate the exclusion of the respective data from the data set. The calculated standardized residuals were within the above upper and lower limits for all the compounds, and thus, none of them were excluded from the data set as outlier.

From the above equation we can conclude that the most significant descriptors according to the ES-SWR algorithm are Dipole (DPL), Principal Moment of Inertia along X axis (PMIX), Shape Coefficient (ShpC), Balaban topological index (TopoJ) and Randic Information index order 0 (ChiInf0). The correlation matrix (Table 3), strongly supports the fact that the five selected descriptors are not highly correlated. Moreover, Variance Inflation Factor (VIF) values for the five descriptors, shown also in Table 3, indicate that the model contains no multicollinearity. The chemical meaning of the five descriptors is briefly described in the sequel.

The principal moments of inertia (PMI) ($\text{g/mol} \text{ \AA}^2$) are physical quantities related to the rotational dynamics of a module [48]. The PMIs are defined by the diagonal elements of the inertia tensor matrix when the Cartesian coordinate axes are the principal axes of the module, with the origin located at the center of mass of the module. In this case the off-diagonal elements of the inertia tensor matrix are zero and the three diagonal elements, I_{xx} , I_{yy} , and I_{zz} correspond to the moments of inertia about the X-, Y-, and Z-axes of the module.

Dipole (DPL) is the electric dipole moment. The electric dipole is a vector quantity, which encodes displacement with respect to the centre of gravity of positive and negative

Table 2 Physicochemical, topological and structural descriptors

ID	Description	Notation	ID	Description	Notation
1	Molar Refractivity	MR	36	Kier-Hall Information 0	KiInf0
2	Diameter	Diam	37	Xu2	Xu2
3	Partition Coefficient (Octanol Water)	ClogP	38	Kier-Hall Information 2	KiInf2
4	Molecular Topological Index	TIndx	39	Balaban Topological	TopoJ
5	Principal Moment of Inertia Z	PMIZ	40	Xu1	Xu1
6	Number of Rotatable Bonds	NRBo	41	Number of Rings	NRings
7	Principal Moment of Inertia Y	PMIY	42	Xu3	Xu3
8	Polar Surface Area	PSAr	43	Bertz	Bertz
9	Principal Moment of Inertia X	PMIX	44	Randic Mod	ChiMod
10	Radius	Rad	45	AtomCompTot	AtomCompTot
11	NBranch	NBranch	46	Number of Clusters	NClusters
12	Shape attribute	ShpA	47	Zagreb2	Zagreb2
13	Total Energy	TotE	48	Wiener	Wiener
14	Shape coefficient	ShpC	49	ScHultz	ScHultz
15	LUMO Energy	LUMO	50	AtomCompMean	AtomCompMean
16	Sum of Valence Degrees	SVDe	51	Kappa3	Kappa3
17	HOMO Energy	HOMO	52	Zagreb1	Zagreb1
18	Total Connectivity	TCon	53	Wiener Distance	WienerDistCode
19	Cluster Count	ClsC	54	Quadratic	Quadr
20	Total Valence Connectivity	TVCon	55	DistEqMean	DistEqMean
21	Randic 1	Chi1	56	Kappa1	Kappa1
22	Wiener Index	WIndx	57	InfMagnitDistTot	InfMagnitDistTot
23	Randic 3	Chi3	58	Kappa2	Kappa2
24	Randic 0	Chi0	59	Gordon	Gordon
25	Randic Information 1	ChiInf1	60	Wiener Information	InfWiener
26	Randic 2	Chi2	61	Electronic Energy	ElectE
27	Randic Information 3	ChiInf3	62	DistEqTotal	DistEqTotal
28	Randic Information 0	ChiInf0	63	Dipole	DPL
29	Kier-Hall 1	Ki1	64	Polarity	Polarity
30	Randic Information 2	ChiInf2	65	Sum of Degrees	SDe
31	Kier-Hall 3	Ki3	66	Heat of Formation	HeatForm
32	Kier-Hall 0	Ki0	67	KiCl3	KiCl3
33	Kier-Hall Information 1	KiInf1	68	Molecular Weight	MW
34	Kier-Hall 2	Ki2	69	ChiCl3	ChiCl3
35	Kier-Hall Information 3	KiInf3			

charges in a molecule [48]. The DPL encodes information about the charge distribution in molecules and is important for modeling polar interactions. Large substituents decrease DPL value which is desirable.

The shape coefficient of a chemical compound is given by: $ShpC = (D - Rad)/Rad$, where the diameter (D) is the maximum such value for all atoms and is held by the most outlying atom(s). The radius (Rad) is the minimum such value and is held by the most central atom(s) [54].

The Randic Information Index, order 0, is a topological information index. Topological information indices encode

both topostructural and topochemical information about the structure [48]. More specifically, these indices encode information on the adjacency and distance of atoms in the molecular structure and quantify information on topology and specific chemical properties of atoms such their chemical identity and hybridization state. In a recent work of our group, topological information descriptors were used in QSAR studies with great success [55]. Topological Information indices have several advantages such as unique representation of the compound and high discriminating power (isomer discrimination).

Table 3 Correlation matrix for the five selected descriptors

	DPL	ShpC	TopoJ	ChiInf0	PMIX	VIF ^a
DPL	1					1.6
ShpC	0.019	1				1.1
TopoJ	0.453	-0.056	1			3
ChiInf0	0.034	-0.189	0.670	1		3
PMIX	0.049	0.26	-0.102	-0.441	1	1.5

^a VIF less than 10 indicate that the model contains no multicollinearity

In addition to the aforementioned indices, the Balaban topological index also significantly influenced the inhibition activity [56]. Topological indices give information not only about the atomic constitution of a compound but also about the presence and character of chemical bonds.

According to the produced QSAR equation (Eq. 4) a high value of Balaban index (TopoJ) gives a positive contribution to the inhibition. On the other hand, a high value of the Principal Moment of Inertia along X axis (PMIX), Dipole (DPL), Randic Information Index, order 0 (ChiInf0) and Shape Coefficient (ShpC) give a negative contribution to the inhibition.

The HDAC inhibition predictability of the proposed model (Eq. 4) was evaluated by using the external set of 13 compounds (Table 1). The QSAR model showed significant predictive ability as indicated from the results of each of the following statistics (Eqs. 1–3):

$$R_{\text{pred}}^2 = 0.83 > 0.6$$

$$\frac{(R_{\text{pred}}^2 - R_o^2)}{R_{\text{pred}}^2} = -0.14 < 0.1,$$

$$\frac{(R_{\text{pred}}^2 - R_o'^2)}{R_{\text{pred}}^2} = -0.19 < 0.1$$

$$k = 0.90, k' = 1.02$$

Finally, the model was further validated by applying the Y-randomization. Several random shuffles of the Y vector were performed and the low R^2 and Q^2 values that were obtained show that the good results in our original model are not due to a chance correlation or structural dependency of the training set. The results of the Y-randomization from 10 shuffles of the Y-vector gave R^2 and Q^2 values in the ranges of 0.0–0.3 and 0.0–0.2, respectively.

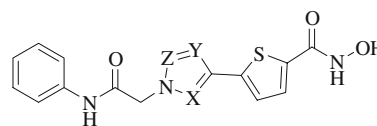
It needs to be emphasized that no matter how robust, significant and validated a QSAR model may be, it cannot be expected to reliably predict the modeled activity for the entire universe of chemicals. Therefore the model's domain of applicability was defined through the leverages for each compound. The extrapolation method was applied to the compounds that constitute the test set. The results are presented in Table 1. None of the 13 compounds fell outside from the domain of the model (warning leverage limit 0.4).

The proposed method, due to the high predictive ability, could be a useful aid to the costly and time consuming experiments for determining the HDAC inhibitory activity by 5-pyridin-2-yl-thiophene-2-hydroxamic acids. The method can also be used to screen virtual compounds in order to identify derivatives with desired activity. In this case, the applicability domain will serve as a valuable tool to filter out “dissimilar” combinations. Within this context a virtual screening study was performed and is presented next.

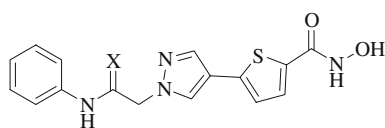
In silico screening

The primary objective of the in silico screen was to determine whether the developed QSAR model could differentiate structures as more or less active than those used for the training and validation sets. The secondary objective was to identify which structural modifications could be tolerated using the domain of applicability. The ultimate role of the in silico screen was as a guide to the identification of the most promising new synthetic targets.

An initial effort to make structural modifications based on an understanding of the five descriptors was complex. Even minor structural modifications indicated difficult to predict antagonistic and/or synergistic effects between variable and multiple descriptors. As such a template based new compound design was initiated. The study was performed starting with compound **6f** ($IC_{50} = 0.021\mu\text{M}$) since this offered good HDAC inhibition and looked promising as a useful scaffold. The modifications incorporated in the virtual screening study were chosen based on their synthetic viability. None of the structures proposed involve the use of unusual ring fragments or functional groups that can not be prepared using established protocols. Since the chemistry of five membered azoles was well understood and introducing practical modifications here was considered synthetically viable, the in silico

**Table 4** Structural modification of azole and predicted activities

ID	X	Y	Z	$-\log IC_{50}$ predicted	Leverage-limit
6f	N	CH	CH	1.39	0.25
1v	CH	CH	N	1.63	0.32
2v	CH	N	CH	1.45	0.30
3v	N	CH	N	1.80	0.29
4v	N	N	CH	1.39	0.23
5v	CH	N	N	1.49	0.26
6v	N	N	N	1.38	0.26

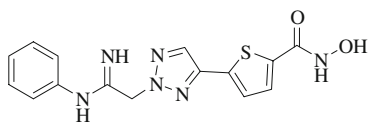
**Table 5** Structural modification of amide and predicted activities

ID	X	$-\log IC_{50}$ predicted	Leverage-limit
1v	O	1.63	0.32
7v	NH	1.68	0.33
8v	NMe	1.13	0.22
9v	NOH	1.06	0.18
10v	NOMe	1.12	0.24
11v	NNH ₂	1.18	0.27
12v	NNHMe	0.83	0.06
13v	NNMe ₂	1.11	0.15
14v	S	1.35	0.25

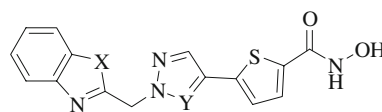
screen began with the replacement of the 1,3-disubstituted pyrazole by other azole heterocycles (Table 4).

The model tolerated the introduction of various azoles since all those studied were within the domain of applicability. The 1,4-disubstituted pyrazole (**1v**) and the 1,3-disubstituted 1,2,3-triazole (**3v**) showed the best activities, 1.63 and 1.80, respectively. Despite the higher predicted activity of the 1,2,3-triazole (**3v**) the pyrazole structure (**1v**) was initially also carried forward for further structural modification until a clear differentiation between the activity of the two azoles became apparent. This decision was based on the ease of preparation of 1,4-disubstituted pyrazoles versus 1,3-disubstituted 1,2,3-triazoles. The next step, chosen somewhat arbitrarily, involved modifying the amide functionality (Table 5).

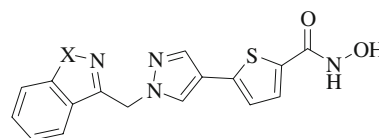
The model tolerated all the structural modifications made to the amide functionality of the 1,4-disubstituted pyrazole (**1v**) with the best predicted activity found for the amidine **7v** (X = NH). The same modification, amide into amidine, for the 1,2,3-triazole (**3v**) led to very large improvement in predicted activity (2.90) but fell outside of the domain of applicability (−0.07).

ID **15v** (Pred. = 2.90, Leverage - limit = -0.07)

Interestingly fusion of the amide functionality to the phenyl to afford a series of benz-1,3-azoles was also tolerated by the QSAR model and allowed for increased diversification of the structures. This included isomerisation of the benz-1,3-azoles into the benz-1,2-azoles. The structural

**Table 6** Predicted activities of selected benz-1,3-azoles

ID	X	Y	$-\log IC_{50}$ predicted	Leverage-limit
16v	O	CH	2.09	0.21
17v	NH	CH	1.95	0.29
18v	NMe	CH	2.15	0.18
19v	S	CH	1.96	0.29
20v	O	N	2.91	−0.02
21v	NH	N	2.66	0.04
22v	NMe	N	2.35	0.01
23v	S	N	2.52	0.13

**Table 7** Predicted activities of selected benz-1,2-azoles

ID	X	Y	$-\log IC_{50}$ predicted	Leverage-limit
24v	O	CH	2.64	0.15
25v	NH	CH	2.65	0.06
26v	NMe	CH	2.10	0.24
27v	S	CH	2.79	−0.01
28v	SO	CH	2.33	0.17
29v	O	N	2.59	0.19
30v	NH	N	2.73	0.13
31v	NMe	N	2.16	0.23
32v	S	N	2.73	0.04
33v	SO	N	2.25	0.27

modification was also tolerated by inclusion of the 1,2,3-triazole moiety which provided compounds with on average increased predicted activities (Tables 6, 7).

Since only two of the above structures were not tolerated by the domain of applicability, structures **20v** and **27v**, further variation of this part of the structure was pursued in order to find the limits of the model. Removal of the benzo fusion led to a large increase of predicted activity but all structures were shown to be outliers (Tables 8, 9).

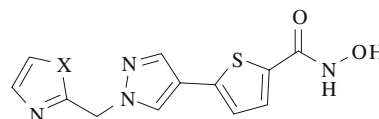
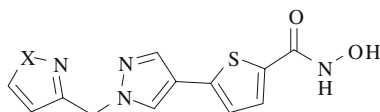
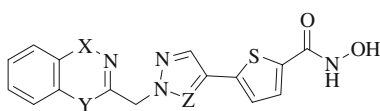


Table 8 Predicted activities of selected 1,3-azoles

ID	X	$-\log IC_{50}$ predicted	Leverage-limit
34v	O	3.72	-0.80
35v	NH	3.47	-0.43
36v	NMe	2.77	-0.13
37v	S	3.58	-0.56

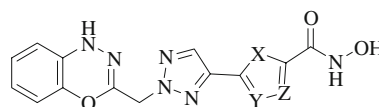
**Table 9** Predicted activities of selected 1,2-azoles

ID	X	$-\log IC_{50}$ predicted	Leverage-limit
38v	O	3.76	-0.84
39v	NH	3.44	-0.42
40v	NMe	3.44	-0.47
41v	S	3.78	-0.89

**Table 10** Predicted activities of selected benzo-1,2,4-heteroazines

ID	X	Y	Z	$-\log IC_{50}$ predicted	Leverage-limit
42v	O	NH	CH	2.31	0.24
43v	O	NMe	CH	2.41	0.23
44v	NH	O	CH	2.94	0.09
45v	NH	NH	CH	1.91	0.32
46v	NH	NMe	CH	1.95	0.26
47v	NMe	O	CH	2.54	0.08
48v	NMe	NH	CH	2.25	0.17
49v	NMe	NMe	CH	2.14	0.14
50v	S	NH	CH	1.82	0.34
51v	O	NH	N	2.85	0.20
52v	O	NMe	N	2.65	0.09
53v	NH	O	N	3.02	0.14
54v	NH	NH	N	2.94	0.16
55v	NH	NMe	N	2.45	0.23
56v	NH	S	N	2.72	0.09
57v	NMe	O	N	2.52	0.15
58v	NMe	NH	N	2.53	0.21
59v	NMe	NMe	N	2.24	0.20
60v	S	NH	N	2.84	0.14

An alternative fusion of the amide functionality to afford benzo-1,2,4-heteroazines was next investigated (Table 10). At best the introduction of a benzo-1,3,4-oxadiazine moiety

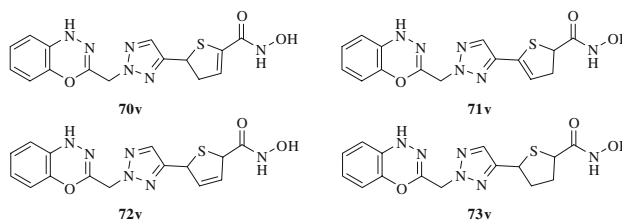
**Table 11** Predicted activities of selected benzo-1,3,4-oxadiazines

ID	X	Y	Z	$-\log IC_{50}$ predicted	Leverage-limit
61v	O	CH	CH	2.97	0.15
62v	NH	CH	CH	2.95	0.15
63v	NMe	CH	CH	2.25	0.20
64v	S	CH	CH	3.02	0.14
65v	S	N	CH	2.87	0.19
66v	S	CH	N	2.82	0.16
67v	S	N	N	2.69	0.21
68v	SO	CH	CH	2.58	0.21
69v	SO ₂	CH	CH	1.55	0.25

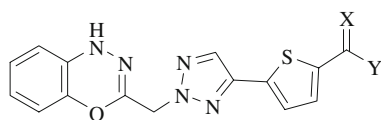
in combination with the 1,4-disubstituted pyrazole led to a structure with excellent predicted activity (44v, Pred. 2.94) but only marginally within the domain of applicability (0.09). The introduction of a benzo-1,2,4-triazine (54v) or benzo-1,3,4-oxadiazine (53v) moiety combined with the 1,3-disubstituted 1,2,3-triazole gave structures of equal or superior predicted activity (54v, Pred. 2.94 and 53v, Pred. 3.02) and comfortably within the domain of applicability (0.14–0.16).

In light of this further structural modifications were performed on the 1,2,3-triazole structure supporting the benzo-1,3,4-oxadiazine 53v. The next modifications focused on the thiophene ring (Table 11).

Replacement of the thiophene by various other five membered heteroarenes gave structures that showed good activity and were within the models domain of applicability, however, none of the new structures surpassed the predicted activity of the thiophene structure 53v. Interestingly introducing the thiophene 1,1-dioxide (69v) led to a dramatic drop in pre-

**Table 12** Predicted activities of dihydro- and tetrahydrothiophenes

ID	$-\log IC_{50}$ predicted	Leverage-limit
70v	2.51	0.25
71v	2.80	0.22
72v	2.67	-0.02
73v	2.84	0.21

**Table 13** Predicted activities of selected benzo-1,3,4-oxadiazines

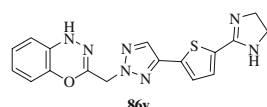
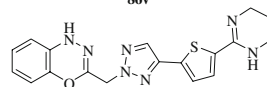
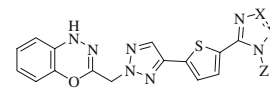
ID	X	Y	$-\log IC_{50}$ predicted	Leverage-limit
74v	O	OH	2.55	0.10
75v	O	OMe	2.67	0.20
76v	O	NH ₂	2.41	0.25
77v	O	NHNH ₂	2.87	0.19
78v	O	NHNHMe	2.49	0.22
79v	O	NHOMe	2.55	0.24
80v	NH	OMe	2.68	0.22
81v	NH	NH ₂	2.85	-0.11
82v	NMe	NHMe	2.91	0.15
83v	NOH	NH ₂	3.21	-0.21
84v	NOH	NHMe	2.88	0.15
85v	S	NH ₂	2.69	0.15

dicted activity (1.55). Since raising the oxidation level of the thiophene reduced the predicted activity it was possible that lowering the oxidation level could favour the activity. In light of this several di- and tetrahydro-thiophene structures were screened. While no one showed superior predicted activity it was interesting to note that the QSAR model, with the exception of the 2,5-dihydro-thiophene **72v**, which was marginally outside of the domain of applicability (-0.02), tolerated these structural modifications (Table 12).

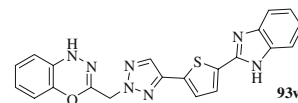
Various modifications of the hydroxamic acid functionality also proved rewarding. The model was able to tolerate a wide range of modifications which included the introduction of both classical and non-classical isosteres (Table 13).

An interesting observation was that the amidine **81v** while affording a good predicted activity was a clear outlier (-0.11); however, the bismethylated amidine **82v** showed equally good predicted activity (2.91) and was clearly within the domain of applicability. Several modifications were made based on this observation that led to dramatic increases in predicted activity (Table 14).

Initially the bismethylated amidine was modified to the cyclic amidines **86v** and **87v** and on doing so the predicted activity increased to 3.53 and 3.49 respectively with both structures marginally within the domain of applicability. Introducing the imidazole moiety **88v** led to even greater predicted activity (3.80) but again the structure remained only marginally within the domain. The predicted activity of this imidazole was reduced on *N*-methylation (**89v**) or *N*-hydroxylation (**90v**) of the free imidazole nitrogen despite a considerable improvement of the domain of applicability. The introduction of additional ring nitrogen atoms to afford

**86v****87v**

88v X = Y = CH, Z = H
89v X = Y = CH, Z = Me
90v X = Y = CH, Z = OH
91v X = N, Y = CH, Z = H
92v X = Y = N, Z = H

**93v****Table 14** Predicted activities of selected cyclic amidines and azoles

ID	$-\log IC_{50}$ predicted	Leverage-limit
86v	3.53	0.06
87v	3.49	0.10
88v	3.80	0.03
89v	2.86	0.26
90v	2.75	0.27
91v	3.72	0.07
92v	3.60	-0.02
93v	2.82	0.15

the triazole (**91v**) and tetrazole (**92v**) did not significantly affect the predicted activity nor the domain of applicability. Benzo fusion (**93v**), however, led to a reduction of the predicted activity but improved domain of applicability. While the data from the experimental and virtual studies have been recorded with the same units it must be noted that the predicted activities produced by the virtual model are in some cases significantly higher. The high biological activities predicted are only indicative of which structures should be targeted for synthesis on the basis that they meet or approach the optimal values for the chosen descriptors for the given model.

In summary, novel structural scaffolds were found using QSAR predictive workflow [57] combined with synthetic chemistry knowledge in order to ensure that the novel scaffolds projects are chemistry driven. This in silico screen based on a simple QSAR model [58] clearly achieved its objective in identifying compounds with improved predicted activity while simultaneously identifying structural modifications that were deemed out of the domain of applicability and therefore the scope of the models reliability. The in silico screen thus demonstrates the usefulness of constructing QSAR models which can aid in identifying new synthetic targets for drug discovery. While not within the scope of this article the preparation of several of the most promising virtual targets is now underway in order to experimentally validate the proposed model.

Conclusions

In this study we have identified five descriptors that successfully model the HDAC inhibitory activity. The validation procedures utilized in this work (separation of data into independent training and validation sets, Y-randomization) illustrated the accuracy and robustness of the produced QSAR model not only by calculating its fitness on sets of training data, but also by testing the predictive ability of the model. Based on the produced QSAR model we have designed novel structures that could be further investigated as novel effective HDAC inhibitors. The proposed method, due to the high predictive ability, offers a useful alternative to the costly and time consuming experiments for determining HDAC inhibitory activity.

Acknowledgments This work was supported by funding from the Cyprus Research Promotion Foundation (Grants No. KINH/0505/03 and PLYPH/0506/25).

References

- Santini V, Gozzini A, Ferrari G (2007) Histone deacetylase inhibitors: molecular and biological activity as a premise to clinical application. *Curr Drug Metab* 8:383–393. doi:10.2174/138920007780655397
- Elaut G, Rogiers V, Vanhaecke T (2007) The pharmaceutical potential of histone deacetylase inhibitors. *Curr Pharm Des* 13:2584–2620. doi:10.2174/138161207781663064
- An W (2007) Histone acetylation and methylation: combinatorial players for transcriptional regulation. *Subcell Biochem* 41:351–369
- de Ruijter AJM, van Gennip AH, Caron HN, Kemp S, van Kuilenburg ABP (2003) Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem J* 370:737–749. doi:10.1042/BJ20021321
- Rasheed WK, Johnstone RW, Prince HM (2007) Histone deacetylase inhibitors in cancer therapy. *Expert Opin Investig Drugs* 16:659–678. doi:10.1517/13543784.16.5.659
- Balakin KV, Ivanenkov YA, Kiselyov AS, Tkachenko SE (2007) Histone deacetylase inhibitors in cancer therapy: latest developments, trends and medicinal chemistry perspective. *Anticancer Agents Med Chem* 7:576–592
- Glozak MA, Seto E (2007) Histone deacetylases and cancer. *Oncogene* 26:5420–5432. doi:10.1038/sj.onc.1210610
- Gallinari P, Di Marco S, Jones P, Pallaoro M, Steinkühler C (2007) HDACs, histone deacetylation and gene transcription: from molecular biology to cancer therapeutics. *Cell Res* 17:195–211
- Andrade CH, Salum LDB, Castilho MS, Pasqualoto KFM, Ferreira EI, Andricopulo AD (2008) Fragment-based and classical quantitative structure-activity relationships for a series of hydrazides as antituberculosis agents. *Mol Divers* 12:47–59
- Yap CW, Li H, Ji ZL, Chen YZ (2007) Regression methods for developing QSAR and QSPR models to predict compounds of specific pharmacodynamic, pharmacokinetic and toxicological properties. *Mini Rev Med Chem* 7:1097–1107. doi:10.2174/138955707782331696
- Castilho MS, Guido RVC, Andricopulo AD (2007) 2D Quantitative structure-activity relationship studies on a series of cholesteryl ester transfer protein inhibitors. *Bioorg Med Chem* 15:6242–6252. doi:10.1016/j.bmc.2007.06.021
- Melagraki G, Afantitis A, Sarimveis H, Igglessi-Markopoulou O, Alexandridis A (2006) A novel RBF neural network training methodology to predict toxicity to *Vibrio fischeri*. *Mol Divers* 10:213–221. doi:10.1007/s11030-005-9008-y
- Afantitis A, Melagraki G, Sarimveis H, Koutentis PA, Markopoulos J, Igglessi-Markopoulou O (2006) A novel QSAR model for evaluating and predicting the inhibition activity of dipeptidyl aspartyl fluoromethylketones. *QSAR Comb Sci* 25:928–935. doi:10.1002/qsar.200530208
- Roy K (2006) Ecotoxicological modeling and risk assessment using chemometric tools. *Mol Divers* 10:93–94. doi:10.1007/s11030-006-9025-5
- Xia B, Ma W, Zheng B, Zhang X, Fan B (2008) Quantitative structure-activity relationship studies of a series of non-benzodiazepine structural ligands binding to benzodiazepine receptor. *Eur J Med Chem* 43:1489–1498. doi:10.1016/j.ejmech.2007.09.004
- Melagraki G, Afantitis A, Makridima K, Sarimveis H, Igglessi-Markopoulou O (2006) Prediction of toxicity using a novel RBF neural network training methodology. *J Mol Model* 12:297–305. doi:10.1007/s00894-005-0032-8
- Tropsha A, Golbraikh A (2007) Predictive QSAR modeling workflow, model applicability domains, and virtual screening. *Curr Pharm Des* 13:3494–3504. doi:10.2174/138161207782794257
- Muegge I, Oloff S (2006) Advances in virtual screening. *Drug Discov Today Technol* 3:405–411
- Afantitis A, Melagraki G, Sarimveis H, Koutentis PA, Markopoulos J, Igglessi-Markopoulou O (2006) Investigation of substituent effect of 1-(3,3-diphenylpropyl)-piperidiny phenylacetamides on CCR5 binding affinity using QSAR and virtual screening techniques. *J Comput Aided Mol Des* 20:83–95. doi:10.1007/s10822-006-9038-2
- Melagraki G, Afantitis A, Sarimveis H, Koutentis PA, Markopoulos J, Igglessi-Markopoulou O (2007) Identification of a series of novel derivatives as potent HCV inhibitors by a ligand-based virtual screening optimized procedure. *Bioorg Med Chem* 15:7237–7247. doi:10.1016/j.bmc.2007.08.036
- Melagraki G, Afantitis A, Sarimveis H, Koutentis PA, Markopoulos J, Igglessi-Markopoulou O (2007) Optimization of biaryl piperidine and 4-amino-2-biarylurea MCH1 receptor antagonists using QSAR modeling, classification techniques and virtual screening. *J Comput Aided Mol Des* 21:251–267. doi:10.1007/s10822-007-9112-4
- Guido RVC, Oliva G, Andricopulo AD (2008) Virtual screening and its integration with modern drug design technologies. *Curr Med Chem* 15:37–46. doi:10.2174/092986708783330683
- Chen H-F, Kang J-h, Li Q, Zeng B-s, Yao X-j, Fan B-t, Yuan S-g, Panay A, Doucet JP (2003) 3D-QSAR study on apicidin inhibit histone deacetylase. *Chin J Chem* 21:1596–1607
- Xie A, Liao C, Li Z, Ning Z, Hu W, Lu X, Shi L, Zhou J (2004) Quantitative structure-activity relationship study of histone deacetylase inhibitors. *Curr Med Chem Anticancer Agents* 4:273–299. doi:10.2174/1568011043352948
- Wang D-F, Wiest O, Helquist P, Lan-Hargest H-Y, Wiech NL (2004) QSAR studies of PC-3 cell line inhibition activity of TSA and SAHA-like hydroxamic acids. *Bioorg Med Chem Lett* 14:707–711. doi:10.1016/j.bmcl.2003.11.062
- Liu B, Lu A-J, Liao C-Z, Liu H-B, Zhou J-J (2005) 3D-QSAR of sulfonamide hydroxamic acid HDAC inhibitors. *Acta Physi-Chim Sin* 21:333–337
- Guo Y, Xiao J, Guo Z, Chu F, Cheng Y, Wu S (2005) Exploration of a binding mode of indole amide analogues as potent histone deacetylase inhibitors and 3D-QSAR analyses. *Bioorg Med Chem* 13:5424–5434. doi:10.1016/j.bmc.2005.05.016

28. Ragno R, Simeoni S, Valente S, Massa S, Mai A (2006) 3-D QSAR studies on histone deacetylase inhibitors. A GOLPE/GRID approach on different series of compounds. *J Chem Inf Model* 46:1420–1430. doi:10.1021/ci050556b
29. Ragno R, Simeoni S, Rotili D, Caroli A, Botta G, Brosch G, Massa S, Mai A (2008) Class II-selective histone deacetylase inhibitors. Part 2: alignment-independent GRIND 3-D QSAR, homology and docking studies. *Eur J Med Chem* 43:621–632. doi:10.1016/j.ejmech.2007.05.004
30. Juvale DC, Kulkarni VV, Deokar HS, Wagh NK, Padhye SB, Kulkarni VM (2006) 3D-QSAR of histone deacetylase inhibitors: hydroxamate analogues. *Org Biomol Chem* 4:2858–2868. doi:10.1039/b606365a
31. Wagh NK, Deokar HS, Juvale DC, Kadam SS, Kulkarni VM (2006) 3D-QSAR of histone deacetylase inhibitors as anticancer agents by genetic function approximation. *Indian J Biochem Biophys* 43:360–371
32. Jaiswal D, Karthikeyan C, Shrivastava SK, Trivedi P (2006) QSAR modeling of sulfonamide inhibitors of histone deacetylase. *Internet Electron J Mol Des* 5:345–354
33. Vadivelan S, Sinha BN, Rambabu G, Boppana K, Jagarlapudi SARP (2008) Pharmacophore modeling and virtual screening studies to design some potential histone deacetylase inhibitors as new leads. *J Mol Graph Model* 26:935–946. doi:10.1016/j.jmgm.2007.07.002
34. Chen Y-D, Jiang Y-J, Zhou J-W, Yu Q-S, You Q-D (2008) Identification of ligand features essential for HDACs inhibitors by pharmacophore modelling. *J Mol Graph Model* 26:1160–1168. doi:10.1016/j.jmgm.2007.10.007
35. Price S, Bordogna W, Bull RJ, Clark DE, Crackett PH, Dyke HJ, Gill M, Harris NV, Gorski J, Lloyd J, Lockey PM, Mullett J, Roach AG, Roussel F, White AB (2007) Identification and optimisation of a series of substituted 5-(1H-pyrazol-3-yl)-thiophene-2-hydroxamic acids as potent histone deacetylase (HDAC) inhibitors. *Bioorg Med Chem Lett* 17:370–375. doi:10.1016/j.bmcl.2006.10.048
36. Price S, Bordogna W, Braganza R, Bull RJ, Dyke HJ, Gardan S, Gill M, Harris NV, Heald RA, van den Heuvel M, Lockey PM, Lloyd J, Molina AG, Roach AG, Roussel F, Sutton JM, White AB (2007) Identification and optimisation of a series of substituted 5-pyridin-2-yl-thiophene-2-hydroxamic acids as potent histone deacetylase (HDAC) inhibitors. *Bioorg Med Chem Lett* 17:363–369. doi:10.1016/j.bmcl.2006.10.045
37. Stewart JJP (2007) Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J Mol Model* 13:1173–1213. doi:10.1007/s00894-007-0233-4
38. CambridgeSoft Corporation ChemOffice. <http://www.cambridge-soft.com>
39. MOPAC2007 Stewart JJP, Stewart Computational Chemistry, Version 7.295W web:<http://OpenMOPAC.net>
40. Svozil D, Lohninger H TOPIX. <http://www.lohninger.com/topix.html>
41. Kennard RW, Stone LA (1969) Computer aided design of experiments. *Technometrics* 11:137–148. doi:10.2307/1266770
42. Melagraki G, Afantitis A, Makridima K, Sarimveis H, Igglessi-Markopoulou O (2006) Prediction of toxicity using a novel RBF neural network training methodology. *J Mol Model* 12:297–305. doi:10.1007/s00894-005-0032-8
43. Afantitis A, Melagraki G, Sarimveis H, Koutentis PA, Markopoulos J, Igglessi-Markopoulou O (2006) A novel QSAR model for predicting induction of apoptosis by 4-aryl-4H-chromenes. *Bioorg Med Chem* 14:6686–6694. doi:10.1016/j.bmc.2006.05.061
44. Chakraborti AK, Gopalakrishnan B, Sobhia ME, Malde A (2003) 3D-QSAR studies of indole derivatives as phosphodiesterase IV inhibitors. *Eur J Med Chem* 38:975–982. doi:10.1016/j.ejmech.2003.09.001
45. Afantitis A, Melagraki G, Sarimveis H, Koutentis PA, Markopoulos J, Igglessi-Markopoulou O (2006) A novel QSAR model for evaluating and predicting the inhibition of dipeptidyl aspartyl fluoromethylketones. *QSAR Comb Sci* 25:928–935. doi:10.1002/qsar.200530208
46. Ghosh P, Thanadath M, Bagchi MC (2006) On an aspect of calculated molecular descriptors in QSAR studies of quinolone antibacterials. *Mol Divers* 10:415–427. doi:10.1007/s11030-006-9018-4
47. Wu W, Walczak B, Massart DL, Heuerding S, Erni F, Last IR, Prebble KA (1996) Artificial neural networks in classification of NIR spectral data: design of the training set. *Chemom Intell Lab Syst* 33:35–46. doi:10.1016/0169-7439(95)00077-1
48. Todeschini R, Consonni V, Mannhold R (2000) In: Kubinyi H, Timmerman H (Series Ed.) Handbook of molecular descriptors. Wiley-VCH, Weinheim
49. Tropsha A, Gramatica P, Gombar VK (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb Sci* 22:69–77. doi:10.1002/qsar.200390007
50. Golbraikh A, Tropsha A (2002) Beware of q²! *J Mol Graph Model* 20:269–276. doi:10.1016/S1093-3263(01)00123-1
51. Melagraki G, Afantitis A, Sarimveis H, Igglessi-Markopoulou O, Alexandridis A (2006) A novel RBF neural network training methodology to predict toxicity to *Vibrio fischeri*. *Mol Divers* 10:213–221. doi:10.1007/s11030-005-9008-y
52. Jalali-Heravi M, Kyani A (2007) Application of genetic algorithm-kernel partial least square as a novel nonlinear feature selection method: activity of carbonic anhydrase II inhibitors. *Eur J Med Chem* 42:649–659. doi:10.1016/j.ejmech.2006.12.020
53. Afantitis A, Melagraki G, Sarimveis H, Koutentis PA, Markopoulos J, Igglessi-Markopoulou O (2006) A novel simple QSAR model for the prediction of anti-HIV activity using multiple linear regression analysis. *Mol Divers* 10:405–414. doi:10.1007/s11030-005-9012-2
54. Petitjean M (1992) Applications of the radius-diameter diagram to the classification of topological and geometrical shapes of chemical compounds. *J Chem Inf Comput Sci* 32:331–337. doi:10.1021/ci00008a012
55. Melagraki G, Afantitis A, Sarimveis H, Igglessi-Markopoulou O, Supuran CT (2006) QSAR study on para-substituted aromatic sulfonamides as carbonic anhydrase II inhibitors using topological information indices. *Bioorg Med Chem* 14:1108–1114. doi:10.1016/j.bmc.2005.09.038
56. Balaban AT (1982) Highly discriminating distance-based topological index. *Chem Phys Lett* 89:399–404. doi:10.1016/0009-2614(82)80009-2
57. Tropsha A, Golbraikh A (2007) Predictive QSAR modeling workflow, model applicability domains, and virtual screening. *Curr Pharm Des* 13:3494–3504. doi:10.2174/138161207782794257
58. Hewitt M, Cronin MTD, Madden JC, Rowe PH, Johnson C, Obi A, Enoch SJ (2007) Consensus QSAR models: do the benefits outweigh the complexity? *J Chem Inf Model* 47:1460–1468. doi:10.1021/ci700016d