Journal of Molecular Graphics and Modelling 27 (2008) 309–320

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Selectivity-based QSAR approach for screening and evaluation of TRH analogs for TRH-R1 and TRH-R2 receptors subtypes

Rameshwar U. Kadam [a,1], Archana G. Chavan [a,1], Vikramdeep Monga [b], Navneet Kaur [b], Rahul Jain [b], Nilanjan Roy [a,*]

[a] Centre of Pharmacoinformatics, National Institute of Pharmaceutical Education and Research, Sector 67, S.A.S. Nagar 160062, Punjab, India
[b] Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research, Sector 67, S.A.S. Nagar 160062, Punjab, India

ARTICLE INFO

Article history:
Received 2 April 2008
Received in revised form 20 May 2008
Accepted 20 May 2008
Available online 27 May 2008

Keywords:
TRH
Selectivity-based QSAR
GFA
TRH-R1
TRH-R2

ABSTRACT

Design and development of therapeutically useful CNS selective thyrotropin-releasing hormone (TRH) analogs acting on TRH-R2 receptor subtype, exerting weak or no TRH-R1-mediated TSH-releasing side effects has gained imagination of researchers in the recent past. The present study reports the development and implementation of a selectivity-based QSAR approach for screening selective agonists of TRH-R2 receptor subtype. The statistically significant predictive models were thoroughly validated using an external validation set whose activity was previously unknown. The model was able to predict preference for either of the receptor subtypes successfully.

1. Introduction

Thyrotropin-releasing hormone (TRH, L-pGlu-L-His-L-Pro-NH_2) is synthesized in the hypothalamus [1,2], and operates in the anterior pituitary to control levels of thyroid-stimulating hormone (TSH) and prolactin. It was the first hypothalamic-releasing factor to be characterized; thereby establishing a fundamental proof for the existence of neuroendocrine regulation of pituitary functions by hypothalamus [3]. TRH is also found in many other tissues and is involved in a wide variety of physiological activities [4–8]. It plays a critical role in regulating the pituitary–thyroid axis by stimulating the synthesis and release of TSH. The synthesis and release of prolactin are also controlled by TRH with equal potency. Though the earlier research on this tripeptide was dominated by its neuroendocrine action on the anterior pituitary, it was later discovered that it also affects the central nervous system (CNS). Alteration of neuronal excitability, enhancement of transmitter release and turnover, increase in CNS arousal, increasing blood pressure, body temperature and respiration rate, alteration of body water and food intake, enhancement of locomotor activity and production of antinociception are some of the observed CNS effects of TRH. Interestingly, the broad spectrum of CNS stimulatory actions has attracted greater attention for potential therapeutic applications of TRH instead of its endocrine properties. These CNS-mediated effects provide the rationale for using TRH in treatment of brain and spinal injuries and certain CNS disorders viz. Alzheimer's disease and motor neuron disease. However, in common with other peptide-based drugs the efficacy of TRH is compromised by its instability and hydrophilic nature. The high-dosage regimes thus needed to obtain neuropharmacological effects often result in adverse side effects arising from the endocrine actions of TRH. Although, analogs with potent CNS activity have been synthesized, none is devoid of endocrine TSH-releasing and other nonendocrine CNS side effects [9].

TRH receptors belong to the rhodopsin/β-adrenergic receptor subfamily of seven transmembrane (TM)-spanning, G protein-coupled receptors (GPCRs) and are situated on the cell surface [10]. TRH-R1 was the first TRH receptor to be cloned from mouse pituitary tumor; orthologous receptors were subsequently cloned from other species, including rat, chicken, white sucker and human [11–13]. A second TRH receptor subtype TRH-R2 was recently identified in rat, mouse and white sucker [14–17]. Amino acid sequences of the two-receptor subtypes from the same species

* Corresponding author. Tel.: +91 172 2214682; fax: +91 172 2214692.
E-mail address: nilanjanroy@niper.ac.in (N. Roy).
[1] Authors have contributed equally to work.

Abbreviations: TRH, thyrotropin-releasing hormone; QSAR, quantitative structure activity relationship; GFA, genetic function approximation; CNS, central nervous system.

1093-3263/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2008.05.005

Table 1
Data set of TRH analogs used in the study, with their experimental and predicted activities for TRH-R1 and TRH-R2 receptor subtypes

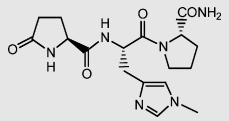
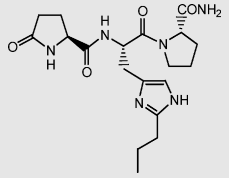
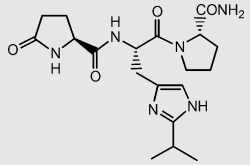
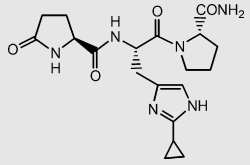
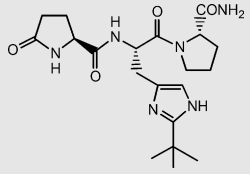
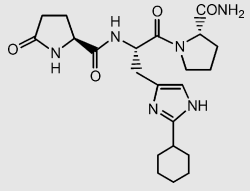
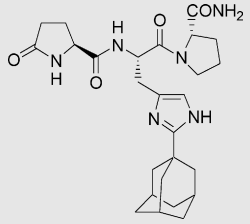
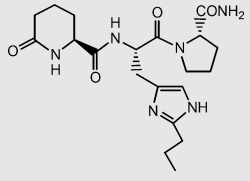
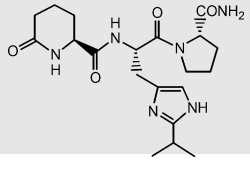
Mol ID	Structure	TRH-R1 receptor Potency (pEC_{50})				TRH-R2 receptor Potency (pEC_{50})				Sel. Cl.[a]
		Exp.	Pre.	Res.	Pot. Cl.[a]	Exp.	Pre.	Res.	Pot. Cl.[a]	
2		9.30	8.72	0.58	TP	9.69	9.73	−0.03	TP	FP
14[b]		6.54	4.85	1.69	FN	7.61	7.08	0.53	TP	TP
15		4.00	4.61	−0.61	TN	5.82	6.08	−0.26	TP	TP
16		4.00	3.25	0.75	TN	6.38	6.15	0.23	TP	TP
17[b]		4.00	4.13	−0.13	TN	5.88	5.97	−0.08	TP	TP
18		4.00	4.41	−0.41	TN	5.95	5.92	0.02	TP	TP
19		4.00	3.80	0.20	TN	4.00	4.13	−0.13	TN	TN
20		6.42	6.29	0.13	TP	7.53	6.87	0.65	TP	FN
21		4.00	3.90	0.10	TN	5.72	5.91	−0.19	TP	TP

Table 1 (*Continued*)

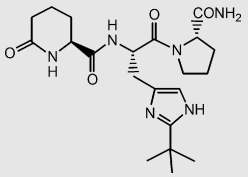
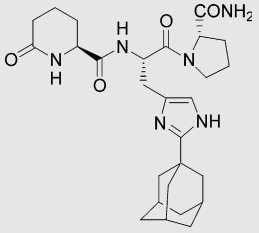
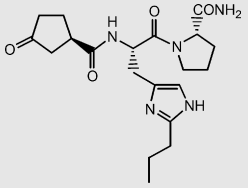
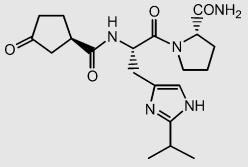
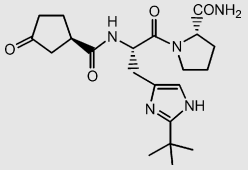
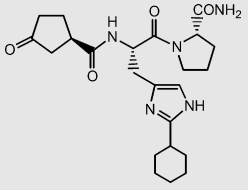
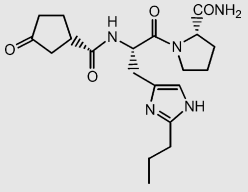
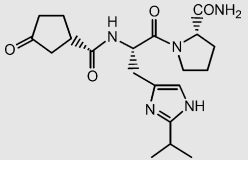
Mol ID	Structure	TRH-R1 receptor Potency (pEC_{50})				TRH-R2 receptor Potency (pEC_{50})				Sel. Cl.[a]
		Exp.	Pre.	Res.	Pot. Cl.[a]	Exp.	Pre.	Res.	Pot. Cl.[a]	
22		4.00	4.19	−0.19	TN	5.48	5.80	−0.32	TP	TP
23		4.00	3.93	0.07	TN	4.00	3.98	0.01	TN	TN
24		4.00	4.22	−0.22	TN	4.00	4.66	−0.66	TN	TN
25[b]		4.00	3.70	0.30	TN	4.00	3.65	0.35	TN	TN
26		4.00	4.02	−0.02	TN	4.00	3.62	0.37	TN	TN
27		4.00	3.87	0.13	TN	4.00	3.71	0.28	TN	TN
28		4.00	4.45	−0.45	TN	4.00	4.66	−0.66	TN	TN
29[b]		4.00	3.93	0.07	TN	4.00	3.65	0.35	TN	TN

Table 1 (*Continued*)

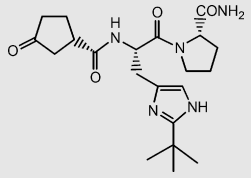
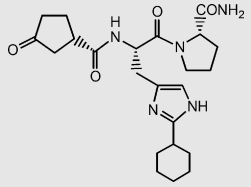
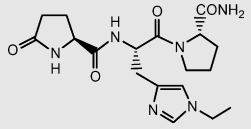
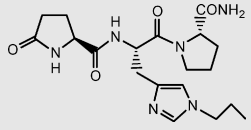
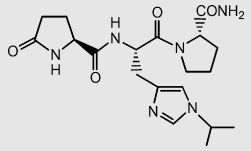
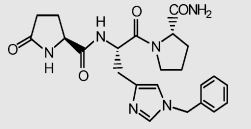
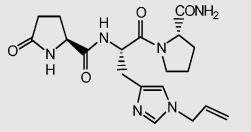
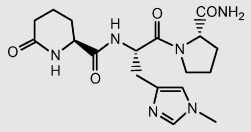
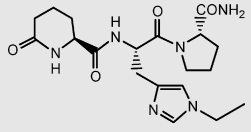
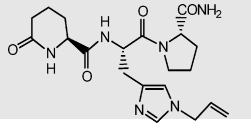
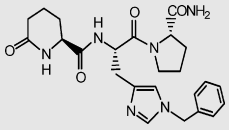
Mol ID	Structure	TRH-R1 receptor Potency (pEC_{50})				TRH-R2 receptor Potency (pEC_{50})				Sel. Cl.[a]
		Exp.	Pre.	Res.	Pot. Cl.[a]	Exp.	Pre.	Res.	Pot. Cl.[a]	
30		4.00	4.00	0.00	TN	4.00	3.62	0.37	TN	TN
31		4.00	4.08	−0.08	TN	4.00	3.71	0.28	TN	TN
32[b]		7.79	7.47	0.32	TP	8.48	8.28	0.20	TP	TN
33		6.23	5.98	0.25	TP	6.92	6.81	0.10	TP	TN
34		5.79	5.67	0.12	TP	7.74	7.53	0.21	TP	TP
35		6.02	6.32	−0.30	TP	6.85	6.51	0.33	TP	TN
36		5.25	5.33	−0.08	TP	6.42	6.69	−0.27	TP	TP
37		8.31	8.46	−0.15	TP	8.61	8.75	−0.14	TP	TN
38		7.85	8.08	−0.23	TP	8.55	8.37	0.18	TP	TN
39[b]		5.29	5.35	−0.06	TP	6.46	6.66	−0.20	TP	TP

Table 1 (*Continued*)

Mol ID	Structure	TRH-R1 receptor Potency (pEC_{50})				TRH-R2 receptor Potency (pEC_{50})				Sel. Cl.[a]
		Exp.	Pre.	Res.	Pot. Cl.[a]	Exp.	Pre.	Res.	Pot. Cl.[a]	
40		4.77	4.32	0.45	TN	5.69	6.40	−0.70	TP	FP

Exp. is experimental value and Pre. is predicted value. Res. refers to the residual calculated as difference in predicted and experimental pEC_{50}. Pot. Cl. is the potency class defined by comparing experimental versus predicted activity classification. Sel. Cl. is the selectivity class defined by comparing experimental versus predicted selectivity.

[a] TP: true positive; TN: true negative; FP: false positive; FN: false negative.

[b] Test set compounds in model-1 (17, 25, 29, and 32) and model-2 (14, 25, 32, and 39).

indicate overall identity of 51% [18]. A comparison of tissue distribution of TRH-R1 and TRH-R2 had been ensued revealing that TRH-R1 is highly expressed in the anterior pituitary and is mainly involved in signaling of TRH within neuroendocrine regions of brain, the autonomic nervous system and the visceral brainstem regions. A very limited mRNA expression for TRH-R1 is exhibited in other regions of the CNS. Whereas, TRH-R2 is highly expressed in rat brain and spinal cord, but is not detectable in the pituitary. The specific expression of TRH-R2 in those areas of brain which are important for the transmission of somatosensory signals and higher CNS functions indicates that in the CNS this receptor subtype may be of major functional importance. It can therefore be speculated that a therapeutically useful CNS selective TRH analog operating *via* TRH-R2 should exert weak or no TSH-releasing effects mediated by TRH-R1. Thus, if analogs with higher selectivity for TRH-R2 receptor subtype were developed, they could probably demonstrate more specific and selective profiles of CNS effects. The present study aimed at such a separation of analogs with higher potency against TRH-R2 than TRH-R1 by development and implementation of a QSAR-based approach.

2. Methods

2.1. Dataset

Recently, 38 TRH analogues [19–24], were synthesized in our laboratory; 28 of which were experimentally tested for potency on both TRH receptors subtypes R1 and R2 while, 10 molecules were with undetermined potency. This dataset of 28 molecules (Table 1) was used for development of QSAR models for TRH receptors subtype, R1 and R2 in an attempt to screen selective analogs of TRH receptors subtype, R2 over R1. 10 molecules with undetermined potency at the outset were kept aside as external validation set to verify the results of the selectivity prediction model.

2.2. Molecule building and optimization

The molecules were built using SYBYL7.1 molecular modeling package [25] installed on a Silicon Graphics Fuel Work Station Running IRIX 6.5 operating system. The basic conformation for TRH was modeled and minimized using PM3 Hamiltonian using MOPAC. In order to generate accurate charge information a single-point energy calculation was also performed using the AM1 Hamiltonian [26] on the PM3 geometry [27–29]. Remaining molecules were built by making required modifications on TRH and were minimized similarly. Mulliken charges [30] were assigned to all the molecules.

2.3. Descriptor generation and feature selection

Three-dimensional representations of molecules were used to calculate their numerical representations with DRAGON 5.0 descriptor generation routines [31]. Approximately 1349 descriptors were calculated and feature selection was performed to reduce the dimensionality of data [32,33]. In this attempt, descriptors with identical or constant values for all the molecules, and those having zero values for more than 5% of molecules were removed. One of the descriptors from a pair having pair wise correlation of more than 0.7 was also eliminated. Finally, 502 descriptors were retained for further analysis. The correlation matrices for descriptors in the final models are shown in Supplementary material; Tables 1 and 2.

2.4. Regression analysis and QSAR model development

The QSAR models for TRH-R1 and TRH-R2 were developed using the genetic function approximation (GFA) [34] approach implemented in Cerius^2 4.10 software package [35]. The dataset of 28 molecules was divided into training and test set of 24 and 4 molecules, respectively. The equation term was set to linear polynomial and the mutation probability was specified as 50%. The population size for each generation was set at 100. Correlation coefficient (r^2) was used to rank the equations; the topmost ranking equations were validated and the one, which proved its goodness in all the tests, was taken up as the final model.

3. Result and discussion

3.1. Development of QSAR models

The main aim of identifying selective TRH-R2 agonists can be split into two short objectives of selecting potent TRH-R2 agonists and determining the stronger potency on TRH-R2 instead of TRH-R1. To fulfill the first objective, a statistically significant QSAR model was developed to enable reliable prediction of TRH-R2 agonistic potency in terms of pEC_{50} (Table 1). The final model, henceforth called model-1 (Eq. (1)) consisted of four descriptors, i.e. atom centered fragment descriptor, functional group counts, burden Eigen value index, 2D-autocorrelation-indices. To enable the comparison and identification of selective analogues, a predictive QSAR model for TRH-R1 agonistic potency was developed using the same dataset (Table 1). The final selected model for TRH-R1 receptor agonistic potency, henceforth called model-2 (Eq. (2)), also consisted of four descriptors, i.e. a Moran descriptors, 2D-autocorrelation index, a WHIM descriptor, H autocorrelation. The common symbols, meanings and values of the descriptors used in both QSAR models development are given

Table 2
Common symbols and definitions of the descriptors used in both QSAR models

Sr. No.	Symbol	Descriptors	Descriptors meaning
1	ACI	MATS1p	Moran autocorrelation-lag 1/weighted by atomic polarizabilites
2	AE	EEIig14x	Eigen value 14 from edge adjacency matrix weighted by edge degrees
3	*W*/*P*(*C*)	G2p	Second component symmetry directional WHIM index/weighted by atomic polarizabilties
4	*H*	H3u	H autocorrelation of lag 3/unweighted
5	*C*	C-038	Al–C(=X)–Al where "X" = any electronegative atom (N, O, S, P, Se, and halogens); "Al" = aliphatic group; "=" = double bond
6	ŃX	nCt	Number of total tertiary C(sp3)
7	*B*	BEHm7	Highest Eigen value n.7 of burden matrix/weighted by atomic masses
8	ACI	GATS4p	Geary autocorrelation-lag 4/weighted by atomic polarizabilites

in Table 2; Supplementary material; Tables 3a and 3b [36–38].

$$pEC_{50} = 40.8522 - 1.35 \times \langle C\text{-}038 \rangle - 0.67 \times \langle nCt \rangle - 12.71 \times \langle BEHm7 \rangle + 5.00 \times \langle GATS4P \rangle \quad (1)$$

$N = 24$, LOF = 0.298, $r^2 = 0.954$, $r^2_{adj} = 0.945$, F-test = 99.265, LSE = 0.132, $q^2 = 0.929$, BS $r^2 = 0.955$, BS error = 0.000, and $r^2_{pred} = 0.887$.

$$pEC_{50} = -22.5786 - 30.2063 \times \langle MATS1p \rangle - 4.72172 \times \langle EEig14x \rangle + 155.38 \times \langle G2p \rangle + 2.97646 \times \langle H3u \rangle \quad (2)$$

$N = 24$, LOF = 0.215, $r^2 = 0.960$, $r^2_{adj} = 0.952$, F-test = 115.220, LSE = 0.995, $q^2 = 0.961$, BS $r^2 = 0.961$, BS error = 0.000, and $r^2_{pred} = 0.729$.

In the statistical values given along with the equations, N is number of compounds in training set; LOF is lack of fit; r^2 is squared correlation coefficient; r^2_{adj} is square of adjusted correlation coefficient; F-test is a variance related statistic that compares two models differing by one or more variable to see if the more complex model is more reliable than the less complex one, the model is supposed to be good if the F-test is above a threshold value; LSE is least-square error; q^2 is the square of the correlation coefficient of the cross-validation; r^2_{pred} is a r^2 like statistic based on test set predictions and is derived from predicted sum of squared residuals (PRESS), which in turn is the indicator of how well the model performs while predicting new data. Both the QSAR models fulfilled the rule of thumb condition that the ratio of 'number of data points/number of descriptors in equation' should be greater than or equal to 4. The fits of experimental pEC_{50} versus that predicted by the QSAR model for training sets are shown in Figs. 1a and 2a. The small residuals among the predicted and experimental potencies (Figs. 1b and 2b) for the test set along with close r^2 and q^2 values preliminarily indicated good predictability of the models.

3.2. Statistical validation of QSAR models

QSAR model validation is an essential part in developing a statistically valid and predictive model because; the real utility of a

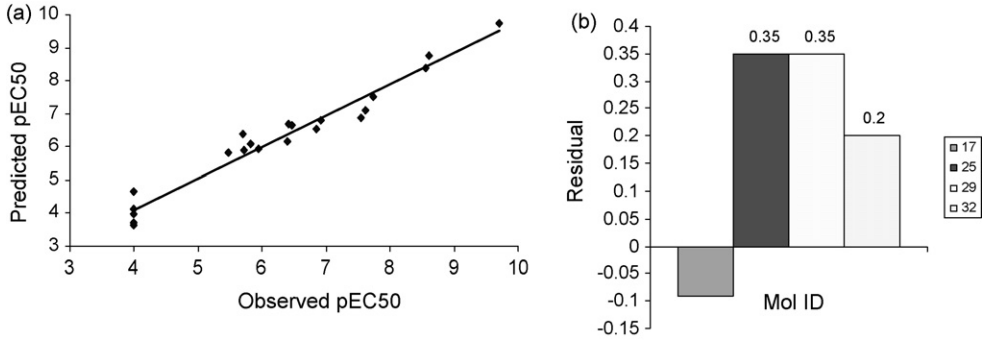


Fig. 1. (a) Scatter-plot of actual versus predicted activity for training set molecules in model-1. (b) Graph of residual values (difference between actual and predicted activities) for test set molecules of for model-1.

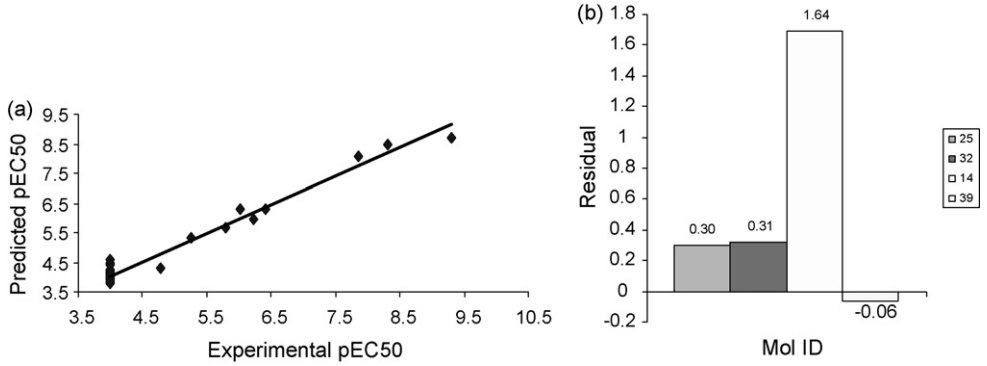


Fig. 2. (a) Scatter-plot of actual versus predicted activity for training set molecules in model-2. (b) Graph of residual values (difference between actual and predicted activities) for test set molecules of for model-2.

QSAR model is in its ability to predict correctly the modeled property for new compounds. In present study a number of QSAR equations were generated, most of which were having good primary statistics represented in terms of high values of r^2, Fischer statistics (*F*-test), etc. Although these values are indicative of the significance of the model, tests like randomization and full cross-validation are required for determining its robustness. In order to select high quality robust models, which could withstand to the violations of statistical assumptions, from among the multiple primary models, randomization and full cross-validation tests were run [39]. Qualitative evaluation of the predictability of the models was done by sensitivity and specificity analysis.

3.2.1. Fischer statistics (F-test)

Fischer statistics (*F*) is the ratio between explained and unexplained variance for a given number of degrees of freedom. The larger the value of *F*, the greater the probability that the QSAR equation is significant. The *F* values for the final models 1 and 2 were 99.265 and 115.220, respectively at 95% confidence level, which suggests that these models are statistically significant.

3.2.2. Randomization test

In this test, the dependent variable is randomly shuffled and a new QSAR model is developed using the original independent variable matrix. The process is repeated several times, and the poor values of r^2 and q^2 for the new models ensure the robustness of the original QSAR models [40–43]. In the present case, 99 random trials were run for each model. None of the random trials could score better than the original models 1 and 2. The regression coefficient (r) for the model from non-random data was of the order of 0.9 against 0.3 for the randomized data in both cases. The results of randomization tests have been presented in Table 3.

3.2.3. Full cross-validation

Following the randomization tests, the models were subjected to full cross-validation. A standard cross-validation in GFA encompasses optimization of regression coefficients and not the optimization of choice of descriptors. That is, the regression model is validated only for the specific subset of descriptors obtained from GFA. In contrast, a full cross-validation encompasses the entire algorithm, including both the choice of descriptors and the optimization of regression coefficients. Each full cross-validation step finds the subset of descriptors for a training set of *N* compounds based on the principles of 'leave-*n*-out' according to which, *n* number of molecules are removed from the training set and a model is generated; this is then used to predict the activities of the removed molecules. The procedure is repeated until each molecule has been removed at least once. The value of *n* was sequentially set to 1, 2, 5, 7 and 10. According to the literature, the predictive QSAR model must have $q^2 > 0.5$ [41,44]. It was very encouraging to observe the high r^2_{cv} (q^2) values of the order of 0.9 for all the runs (Table 4).

Table 3
Randomization test results for both models

Parameters	Model-1	Model-2
r from non-random trial	0.976	0.906
S.D. of non-random trial	4.385	5.751
No. of random trials	99	99
No. of trials with r greater than non-random trial	0	0
No. of trials with r lesser than non-random trial	99	99
Mean r from random trials	0.397	0.327
S.D. of random trials	0.132	0.099

Table 4
Full cross-validation test results for both models

Rule	Model-1			Model-2		
	PRESS	S.D.	CV r^2	PRESS	S.D.	CV r^2
Leave-1-out	3.080	69.603	0.956	2.57	57.878	0.956
Leave-2-out	3.262	69.600	0.953	2.31	57.878	0.961
Leave-5-out	3.035	69.600	0.956	3.27	57.878	0.944
Leave-7-out	5.636	69.600	0.919	3.25	57.878	0.943
Leave-10-out	5.661	69.600	0.919	3.40	57.878	0.941

3.3. Sensitivity and specificity

The sensitivity and specificity of the models should also be established during validation to assess the propensity of the QSAR models for correct qualitative prediction of the dependent variable [45]. A set of compounds, which has been studied experimentally, can be divided into true positives (TP), i.e. active and true negatives (TN), i.e. in actives. The sensitivity of a model expresses the ability of the model to identify TP. It can be illustrated as the probability of predicting a truly positive molecule as positive, and can be calculated as

$$\text{Sensitivity}\,(\%) = 100 \times \frac{\text{TP}-\text{FN}}{\text{TP}}$$

TP–FN represents the number of correct predictions. A sensitivity of 60% means that for 60% of the cases the model will correctly predict truly active compounds to be so. In about 40% of the cases truly active compounds will be predicted as inactive, i.e. false negatives. Thus, the sensitivity can be regarded as a conservative evaluation since all substances modeled as positive, in fact are truly positive.

Specificity on the other hand, expresses the ability of the model to predict if a truly negative substance is negative. The specificity can be illustrated as the probability that the prediction is negative, given that the substance is truly negative and can be calculated as

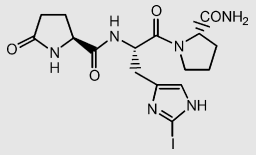
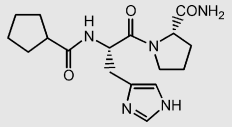
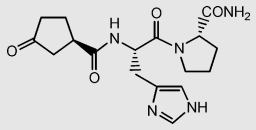
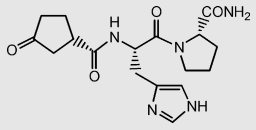
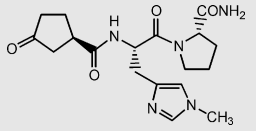
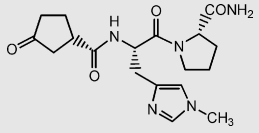
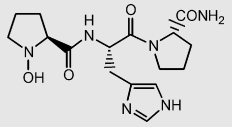
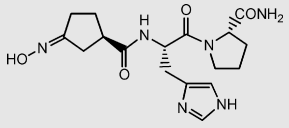
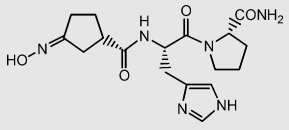
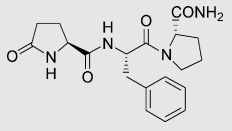
$$\text{Specificity}\,(\%) = 100 \times \frac{\text{TN}-\text{FP}}{\text{TN}}$$

Specificity of a model can be regarded as a non-conservative evaluation since not all truly negative substances are modeled as such. In particular study, a specificity of 60 means that in 60% of the cases the model will correctly predict truly inactive compounds. In about 40% of the cases an inactive substances will be predicted as active, i.e. false positive in the present study. Molecules have been considered as active on a particular receptor subtype if the $pEC_{50} \geq 5$, else they have been classified as inactive. All the 28 molecules (Table 1) were used for calculating the sensitivity and specificity values. While both models presented 100% specificity, the sensitivities of models 1 and 2 were observed to be 100% and 90%, respectively (Table 1). This implies that model-1 would correctly predict the potency profile of the molecules for R2 receptor subtype. However, model-2 would wrongly classify 10% of the inactive molecules as active on R1 subtype. Based on these statistics it can be expected that while searching for molecules, which are active on R2 receptor subtype but inactive on R1 receptor subtype, the chance of correctly selecting a molecule is higher than the chance of correct rejection.

3.4. Selectivity prediction

To confirm the probability of identifying true selective molecules using the two models, the sensitivity and specificity tests were performed on the observed versus predicted selectivity values for the dataset. A molecule was considered as selective if it was at least 10 times more potent for R2 receptor subtype, i.e. (pEC_{50} for R2) − (pEC_{50} for R1) ≥ 1. For the purpose of this study,

Table 5
Predicted and experimental activities for the validation set molecules

Mol ID	Structure	TRH-R1 receptor Potency (pEC_{50})				TRH-R2 receptor Potency (pEC_{50})				Sel. Cl.[a]
		Exp.	Pre.	Res.	Pot. Cl.[a]	Exp.	Pre.	Res.	Pot. Cl.[a]	
3		6.02	5.87	−0.15	TP	6.00	4.90	−1.1	FN	TN
4		3.00	6.98	3.98	FP	6.00	10.20	4.2	TP	TP
5		6.01	6.53	0.52	TP	7.38	8.78	1.4	TP	TP
6		7.00	6.83	−0.17	TP	8.03	8.78	0.75	TP	TP
7		6.6	8.47	1.87	TP	7.99	7.42	−0.57	TP	FN
8		7.79	8.21	0.42	TP	8.63	7.42	−1.21	TP	TN
9		4.15	6.90	2.75	FP	7.22	11.30	4.08	TP	TP
10		4.30	3.74	−0.56	TN	7.15	8.17	1.02	TP	TP
11		5.20	3.32	−1.88	FN	7.10	8.17	1.07	TP	TP
12		6.96	6.29	−0.67	TP	6.82	9.58	2.76	TP	TP

Exp. is experimental value and Pre. is predicted value. Res. refers to the residual calculated as difference in predicted and experimental pEC_{50}. Pot. Cl. is the potency class defined by comparing experimental versus predicted activity classification. Sel. Cl. is the selectivity class defined by comparing experimental versus predicted selectivity.
[a] TP: true positive; TN: true negative; FP: false positive; FN: false negative.

selective R2 agonists were defined as positives and all other molecules were called negatives. The observed values for sensitivity and specificity were 90% and 93%, respectively (Table 1). This implies that the chance of predicting a true selective molecule to be selective is 90% as compared to 7% for predicting a non-selective molecule as selective. The chance of rejecting a true non-selective is 93% as compared to 10% for wrongly rejecting a selective molecule. These results were in accordance with those expected from the earlier examination of probability of predicting the potency profile for the two-receptor subtypes separately.

3.5. External validation of QSAR models

The QSAR models were further validated using an external validation set of 10 molecules whose experimental results were yet undetermined. The selectivity and potency predictions were made using the two models, and compared with the results from the subsequent experimentation. The experimental and predicted values along with the residuals for the validation set are given in Table 5. According to our selectivity criteria, five molecules (4, 9, 10, 11 and 12) were predicted to be selective for TRH-R2 receptor over TRH-R1 (Fig. 3), two molecules (5 and 6) would show a higher potency for TRH-R2 receptor and the other three molecules (3, 7 and 8) would show more potency on TRH-R1 (Table 5). On biological testing, molecules 4, 9 and 10 did prove to be TRH-R2 selective and molecule 11 showed higher activities on TRH-R2 though not sufficient to be categorized as selective. Similarly, as predicted, molecules 5 and 6 were more active on TRH-R2. Molecule 3 showed a slight preference towards TRH-R1 among the third group of molecules predicted to be more active on this subtype. The observed preference for either of the receptor subtypes was differing from the predictions only in 3 cases.

Based on the difference in predicted and experimental potency values, three molecules appeared as outliers for model-1 and two for model-2. The r^2_{pred} was calculated at 0.60 and 0.61, respectively after removing the outliers. Thus, the models could predict the potency of the validation set within acceptable limits. The sensitivities for predicting the potency profile of the molecules were 89% and 83%, respectively for models 1 and 2 (Table 5). The specificity for model-2 was observed at 100% because two molecules were false positives as compared to a single TN. For selectivity prediction, the models displayed a sensitivity of 86% and specificity of 100%. These values held true with the previous inferences of the models being more stringent for predicting a molecule to be R2 selective. It was indeed observed that there were no false positives but one false negative for selectivity. Successful validations of the models lead us to believe that the developed QSAR models were statistically significant, reliable for predicting potency and for identification of TRH-R2 selective molecules.

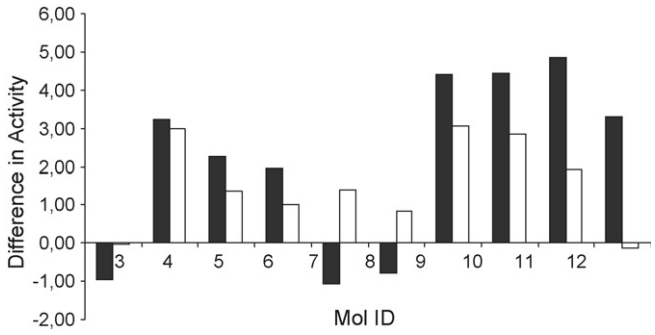


Fig. 3. Plot of difference between predicted and experimental activities, represented by black and white columns, respectively for validation set molecules by TRH-R1 and TRH-R2 models. Difference in activity = pEC_{50} for TRH-R2 − pEC_{50} for TRH-R1.

3.6. Elucidation of SAR

Analysis of the molecular structures in the dataset, in corroboration with the developed QSAR models provided some interesting information regarding the role of different moieties in TRH selectivity and their correlation with molecular descriptors. TRH analogs containing mono-substituted histidine moieties where N^{τ} or C-2 positions of histidine were substituted with various alkyl, aryl and cycloalkyl groups, e.g. molecules 14, 15, 16, 17, 18 and 19 (Table 1) and the molecules in which the whole of the central histidine was replaced with other amino acids containing aliphatic side chain were more potent towards receptor subtype R2 as compared with R1. This suggested that the central histidine ring is important for the endocrine potency of the original peptide and not much important for its CNS-mediated effects. On the other hand, the replacement of the pGlu acid with six membered counterparts like 3-oxocyclopentane-1-carboxylic acid or pyroaminoadipic acid, e.g. molecules 24–31 (Table 1), did not present any noticeable affinity for either of the receptor subtypes. Interestingly simultaneous TRH analogues generated by replacement of pGlu acid with 3-oxocyclopentane-1-carboxylic acid or pyroaminoadipic acid and of the central histidine ring with other amino acids containing an aliphatic side chain, lead to high selectivity for the TRH-R2 subtype and low potency at TRH-R1. In model-1 the potency is negatively correlated to the C-038 (atom centered fragment descriptor) and BEHm7 (burden Eigen value index) descriptors; the correlation being particularly deterministic with BEHm7. The atom centered fragment descriptors represent the electronegative atoms and aliphatic groups in molecule and the Eigen value indices are believed to reflect the topology of whole molecule because they contain contributions from all atoms of the molecule. Together these descriptors very well represent the effects of histidine substitutions on TRH-R2 potency. The functional group count descriptors, which are simplest measures for defining molecular size in terms of total number of atoms in the molecule can account for the observed affects of pGlu acid substitutions. In model-2 the MATS1p descriptor (having Moran coefficient value usually in interval [−1, +1]) shows negative

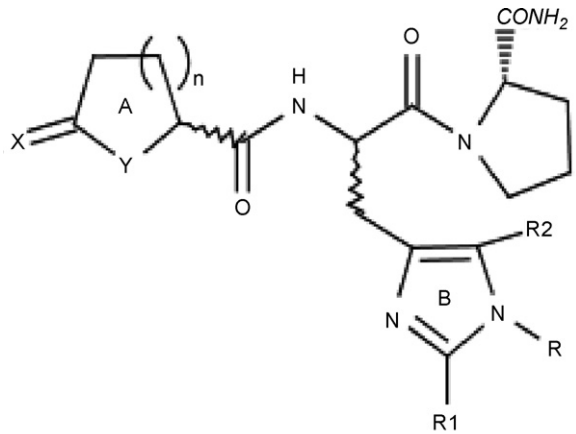
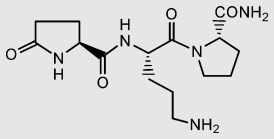
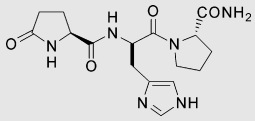
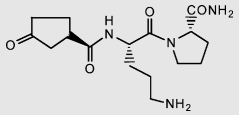
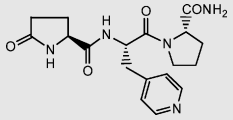
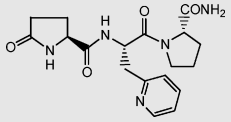
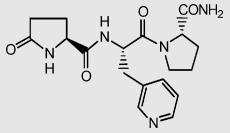
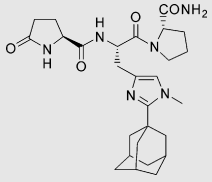
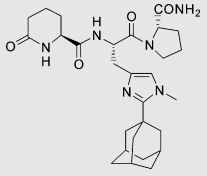
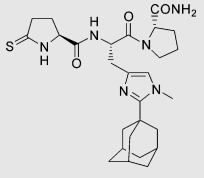
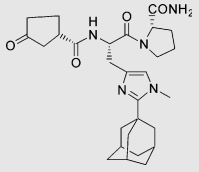


Fig. 4. Main skeleton of the TRH analogs used in this study. X = O and S; Y = CH_2 and NH; n = 1 and 2; R = various primary, secondary and tertiary alkyl groups, aryl group, cycloalkyl groups of various sizes, substituted aryl groups; R_1 = various primary, secondary or tertiary alkyl groups, cycloalkyl groups of various sizes, unsubstituted or substituted aryl groups, halogens F, Cl, Br or I; R_2 = F, Cl, Br or I. The stereochemistry of the ring **A** could be (*R*, *S*), (*R*) or (*S*); the stereochemistry of the ring **B** could be (*R*, *S*), (*R*) or (*S*).

Table 6
Screened compounds, predicted to be highly potent and selective for TRH-R2 receptor

Mol ID	Structure	Predicted potency (pEC_{50})		
		TRH-R1	TRH-R2	Difference (pEC_{50}(R2) − pEC_{50}(R1))
44		7.23	11.62	4.39
166		7.4	11.13	3.73
46		6.93	10.67	3.74
43		5.59	10.17	4.58
41		5.73	9.66	3.93
42		5.35	9.59	4.24
53		0.97	7.77	6.8
60		2.32	7.59	5.27
81		1.18	7	5.83
121		1.99	7	5.01

correlation with potency at TRH-R1 subtype suggesting that replacement of pGlu acid with more polar moieties should show low TRH-R1 affinity and consequently selectivity for TRH-R2, as observed in molecules 9, 10 and 11 (Table 5).

3.7. Virtual library design and screening

A virtual library of 138 molecules was subsequently designed using similar methodology as employed for designing the 38 TRH analogues [19–24], while taking into consideration their experimental potency profile and the inferences from the QSAR models (Supplementary material, Table 4). The focus was maintained on modifying pGlu acid and His residues on the TRH tripeptide backbone (Fig. 4) because these two residues together are known to almost completely account for the observed binding energy of the peptide. The library was diversified by modifying pGlu acid and His residues with the non-proteinogenic amino acids while keeping ProNH2 residue intact. The two amino acids were either replaced simultaneously or alone in a series. The above reported thoroughly validated models were then used to screen this in-house virtual library. The top ranking molecules (Table 6) in terms of TRH-R2 selectivity were selected and are presently being synthesized for biological analysis.

4. Conclusion

A selectivity-based QSAR approach was successfully implemented for TRH-R1 and TRH-R2 receptor subtypes, which should prove to be useful in screening therapeutically useful CNS selective thyrotropin-releasing hormone (TRH) analogues acting *via* TRH-R2. Such a guide will be of immense help to medicinal chemist before venturing into costly synthesis and biological testing. The ligand-based strategy employed here could be also be appealing for screening numerous other ligands acting on the GPCR family for which no data on the protein structures is available.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2008.05.005.

References

[1] R. Böler, F. Enzman, K. Folkers, C. Bowers, A.V. Schally, The identity of chemical and hormonal properties of the thyrotropin-releasing hormone and pyroglutamyl-histidyl-prolinamide, Biochem. Biophys. Res. Commun. 37 (1969) 705–710.
[2] R. Burgus, T. Dunn, D. Desiderio, R. Guillemin, Derives Polypeptidiques de Syntheses Doues d'activite Hypophysiotropic TRF, C. R. Acad. Sci. 269 (1969) 1870–1873.
[3] A.V. Schally, C.Y. Bowers, T.W. Redding, J.F. Barrett, Isolation of thyrotropin releasing factor (TRF) from porcine hypothalamus, Biochem. Biophys. Res. Commun. 25 (1966) 165–169.
[4] H.H. Keller, G. Bartholini, A. Plestscher, Enhancement of cerebral noradrenaline turnover by thyrotropin-releasing hormone, Nature 248 (1974) 528–529.
[5] G.P. Muller, H.J. Chen, J. Meites, In vivo stimulation of prolactin release in the rat by synthetic TRH, Proc. Soc. Exp. Biol. Med. 144 (1973) 613–615.
[6] G. Metcalf, TRH: a possible mediator of thermoregulation, Nature 252 (1974) 310–311.
[7] A.I. Faden, I. Sacksen, L.J. Noble, Structure activity relationships of TRH analogs in experimental spinal injury, Brain Res. 448 (1988) 287–293.
[8] S.K. Salzman, E. Hirofugi, P.B. Knight, C. Llados-Eckman, A.L. Beckman, A. Winokur, Treatment of experimental spinal trauma with thyrotropin-releasing hormone: central serotonergic and vascular mechanisms of action, Cent. Nerv. Syst. Trauma 4 (1987) 181–196.
[9] L. Prokai, Central nervous system effects of thyrotropin-releasing hormone and its analogues: opportunities and perspectives for drug discovery and development, Prog. Drug Res. 59 (2002) 134–169.
[10] R.E. Straub, G.C. French, R.H. Joho, M.C. Gershengorn, Expression cloning of a cDNA encoding the mouse pituitary thyrotropin-releasing hormone receptor, Proc. Natl. Acad. Sci. U.S.A. 87 (1990) 9514–9518.
[11] P. De la Pena, L.M. Delgado, D. Del Camino, F. Barros, Two isoforms of the thyrotropin-releasing hormone receptor generated by alternative splicing have indistinguishable functional properties, J. Biol. Chem. 267 (1992) 25703–25708.
[12] S.M. Duthie, P.L. Taylor, L. Anderson, J. Cook, K.A. Eidne, Cloning and functional characterization of the human TRH receptor, Mol. Cell. Endocrinol. 95 (1993) R11–R15.
[13] Y.M. Sun, R.P. Millar, H. Ho, M.C. Gershengorn, N. Illing, Cloning and characterization of the chicken thyrotropin-releasing hormone receptor, Endocrinology 139 (1998) 3390–3398.
[14] J. Cao, D. O'Donnell, H. Vu, K. Payza, C. Pou, C. Godbout, A. Jakob, M. Pelletier, P. Lembo, S. Ahmad, P. Walker, Cloning and characterization of a cDNA encoding a novel subtype of rat thyrotropin-releasing hormone receptor, J. Biol. Chem. 273 (1998) 32281–32287.
[15] S. Harder, X. Lu, W. Wang, F. Buck, M.C. Gershengorn, T.O. Bruhn, Regulator of G protein signaling 4 suppresses basal and thyrotropin-releasing hormone (TRH)-stimulated signaling by two mouse TRH receptors, TRH-R(1) and TRH-R(2), Endocrinology 142 (2001) 1188–1194.
[16] H. Itadani, T. Nakamura, J. Itoh, H. Iwaasa, A. Kanatani, J. Borkowski, M. Ihara, M. Ohta, Cloning and characterization of a new subtype of thyrotropin-releasing hormone receptors, Biochem. Biophys. Res. Commun. 250 (1998) 68–71.
[17] B.F. O'Dowd, D.K. Lee, W. Huang, T. Nguyen, R. Cheng, Y. Liu, B. Wang, M.C. Gershengorn, S.R. George, TRH-R2 exhibits similar binding and acute signaling but distinct regulation and anatomic distribution compared with TRH-R1, Mol. Endocrinol. 14 (2000) 183–193.
[18] Y. Sun, X. Lu, M.C. Gershengorn, Thyrotropin-releasing hormone receptors-similarities and differences, J. Mol. Endocrinol. 30 (2003) 87–97.
[19] N. Kaur, V. Monga, X. Lu, M.C. Gershengorn, R. Jain, Modifications of the pyroglutamic acid and histidine residues in thyrotropin-releasing hormone (TRH) yield analogs with selectivity for TRH receptor type 2 over type 1, Bioorg. Med. Chem. 15 (2007) 433–443.
[20] N. Kaur, V. Monga, J.S. Josan, X. Lu, M.C. Gershengorn, R. Jain, Synthesis, receptor binding and activation studies of *N*(1)-alkyl-L-histidine containing thyrotropin-releasing hormone (TRH) analogues, Bioorg. Med. Chem. 14 (2006) 5980–5987.
[21] S. Engel, S. Neumann, N. Kaur, V. Monga, R. Jain, J. Northup, M.C. Gershengorn, Low affinity analogs of thyrotropin-releasing hormone are super-agonists, J. Biol. Chem. 281 (2006) 13103–13109.
[22] N. Kaur, X. Lu, M.C. Gershengorn, R. Jain, Thyrotropin-releasing hormone (TRH) analogues that exhibit selectivity to TRH receptor subtype 2, J. Med. Chem. 48 (2005) 6162–6165.
[23] R. Jain, J. Singh, J.H. Perlman, M.C. Gershengorn, Synthesis and biology of new thyrotropin-releasing hormone (TRH) analogues, Bioorg. Med. Chem. 10 (2002) 189–194.
[24] J.H. Perlman, A.O. Colson, R. Jain, B. Czyzewski, L.A. Cohen, R. Osman, M.C. Gershengorn, Role of the extracellular loops of the thyrotropin-releasing hormone receptor: evidence for an initial interaction with thyrotropin-releasing hormone, Biochemistry 36 (1997) 15670–15676.
[25] SYBYL, 7.1, Tripos Associates Inc., 1699, S. Hanley Road, St. Louis, MO631444, USA.
[26] J.S. Dewar, E.G. Zoebisch, E.F. Healy, AM1: a new general purpose quantum mechanical molecular model, J. Am. Chem. Soc. 107 (1985) 3902–3909.
[27] H. Linnan, P.C. Jurs, Assessing the reliability of a QSAR model's predictions, J. Mol. Graph. Model. 23 (2005) 503–523.
[28] R.U. Kadam, D. Garg, A. Chavan, N. Roy, Evaluation of *Pseudomonas aeruginosa* deacetylase LpxC inhibitory activity of dual PDE4-TNF-alpha inhibitors: a multi-screening approach, J. Chem. Inf. Model. 47 (2007) 1188–1195.
[29] R.U. Kadam, N. Roy, Cluster analysis and two-dimensional quantitative structure–activity relationship (2D-QSAR) of *Pseudomonas aeruginosa* deacetylase LpxC inhibitors, Bioorg. Med. Chem. Lett. 19 (2006) 5136–5143.
[30] R.S. Mulliken, Electronic population analysis on LCAO-MO molecular wave function I, J. Chem. Phys. 23 (1955) 1833–1846.
[31] Dragon, 5.3, Milano Chemometrics and QSAR Research Groups Inc., 2002.
[32] R. Buigni, Quantitative structure activity relationships (QSAR) models of mutagens and carcinogens, pp. 1–40.
[33] S.C. Basak, R. Natarajan, D. Mills, D.M. Hawkins, J.J. Kraker, Quantitative structure–activity relationship modeling of juvenile hormone mimetic compounds for *Culex pipiens* larvae, with a discussion of descriptor-thinning methods, J. Chem. Inf. Model. 46 (2006) 65–77.
[34] D. Rogers, A.J. Hopfinger, Application of genetic function approximation to quantitative structure–activity relationships and quantitative structure–property relationships, J. Chem. Inf. Comput. Sci. 34 (1994) 854–866.
[35] Cerius^2, Version 4.9, Accelrys Inc., 6985 Scranton Road, San Diego, CA, USA.
[36] A.R. Leach, V.J. Gillet, Molecular Descriptors, Kluwer Academic Publishers, 2003, pp. 53–75.
[37] J. Gasteiger, Chemoinformatics, A Textbook, Wiley–VCH, GmbH and Co., pp. 415–428.
[38] R. Todeschini, V. Consoni, Handbook of Molecular Descriptors. In Method and Principals of Medicinal Chemistry, Wiley–VCH, pp. 492–513.
[39] C. Hansch, R.P. Verma, 20-(*S*)-Camptothecin analogues as DNA topoisomerase I inhibitors: a QSAR study, Chem. Med. Chem. 2 (2007) 1807–1813.
[40] C.D. Selassie, S. Kapur, R.P. Verma, M. Rosario, Cellular apoptosis and cytotoxicity of phenolic compounds: a quantitative structure–activity relationship study, J. Med. Chem. 48 (2005) 7234–7242.
[41] A. Tropsha, P. Gramatica, V.K. Gombar, The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models, QSAR Comb. Sci. 22 (2003) 69–77.

[42] S. Wold, L. Eriksson, in: H. van de Waterbeemd (Ed.), Chemometrics Methods in Molecular Design, Wiley–VCH, Weinheim, 1995, pp. 309–318.
[43] M.F. Melzig, G.D. Tran, K. Henke, C.D. Selassie, R.P. Verma, Inhibition of neutrophil elastase and thrombin activity by caffeic acid esters, Pharmazie 60 (2005) 869–873.
[44] A. Golbraikh, A. Tropsha, Beware of q^2! J. Mol. Graph. Model. 20 (2002) 269–276.
[45] J.D. Walkera, L. Carlsenb, J. Jaworskac, Improving opportunities for regulatory acceptance of QSARs: the importance of model domain, uncertainty, validity and predictability, QSAR Comb. Sci. 22 (2003) 346–350.