informa
healthcare

RESEARCH ARTICLE

Modeling of the growth hormone secretagogue receptor antagonistic activity using chemometric tools

Susheela Sharma[1], B. K. Sharma[2], P. Pilania[2], P. Singh[2], and Y. S. Prabhakar[3]

[1]*Department of Engineering Chemistry, Sobhasaria Engineering College, Sikar-332 021, India, *[2]*Department of Chemistry, S. K. Government College, Sikar-332 001, India, and *[3]*Medicinal and Process Chemistry Division, Central Drug Research Institute, Lucknow-226 001, India*

Abstract

A quantitative structure-activity relationship (QSAR) study has been carried out on growth hormone secreta-gogue receptor antagonistic activity of the derivatives of 2,4-diaminopyrimidine. To obtain significant QSARs, the approaches involving the non-parametric such as Fujita-Ban, and the parametric based on physicochemical and DRAGON descriptors in Hansch type of analysis have been employed. The Fujita-Ban approach, however, was constrained to 18 compounds only due to a limited number of substituents appeared at varying positions. The derived contributions of different substituents and the parent moiety were used to identify the potential congeners. The physicochemical model of Hansch was subsequently used to interpret the type of interactions involved between the receptor sites and varying positions of these compounds. The study, employing DRAGON descriptors in Hansch approach was also carried out on this data set to discuss the prevailing interactions in terms of topological descriptors. The derived highest significant model was discussed to delineate the type of interactions involved and suggestions have been made for different alterations to lead to further potential compounds of the series.

Keywords: *Quantitative structure-activity relationship (QSAR); antagonists of growth hormone secretagogue receptor (GHS-R); binding activity to GHS-R; derivatives of 2, 4-diaminopyrimidine; Fujita-Ban approach; Hansch type of analysis; combinatorial protocol in multiple linear regression (CP-MLR) analysis; physicochemical parameters and DRAGON descriptors*

Introduction

Many chronic disorders such as insulin resistance, diabetes, dyslipidemia, hypertension, coronary heart disease, ischemic stroke, pulmonary disease, gout, osteoarthritis, cancer, neurological diseases, cataracts, gastrointestinal disease, and genitourinary disease (in women) are associated with, a worldwide problem, obesity. It has been shown that the progression of these disorders are reversed or slowed by weight loss [1]. Body weight may be maintained by balancing caloric intake with the energy expenditure. Thus, diet and exercise are crucial to control the weight. However, these may be assisted by appetite reducing oral agents. Ghrelin, produced in stomach and hypothalamus [2], is a 28 amino acid containing orexigenic hormone. It is an endogenous ligand of the growth hormone secretagogue receptor (GHS-R). GHS-R is a G-protein-coupled receptor which is expressed primarily in the pituitary gland and the brain and to a lesser extent in the periphery. The unique post-translational acylation, an octanoyl group at the Ser-3 hydroxyl group, of active ghrelin is essential for orexigenic function and also necessary for its penetration of the blood-brain barrier [2,3]. The intake of nutrient regulates the level of ghrelin in blood. In rats and human subjects, intracerebroventricular and intravenous administration of ghrelin, respectively, increased appetite and food intake [4–8]. It has been observed that the weight gain, induced by ghrelin, is primarily from adipose tissue and not from muscle or bone mass [9]. The inhibition of ghrelin actions with peptidyl GHS-R antagonists [7,9], ghrelin antibody [10], and antisense oligonucleosides [11] resulted in weight loss and food intake decrease in rodents [7, 9–11]. The isoxazole and tetralin carboxamide based GHS-R antagonists [12–15], displaced ^{125}I-ghrelin binding

http://www.informapharmascience.com/enz

RIGHTS LINK

and antagonized ghrelin-induced intracellular Ca^{2+} flux in functional assay. However, these compounds did not exhibit an effect when tested in *in vivo* animal models. More recently, a synthetic study was performed [16] on the derivatives of 2,4-diaminopyrimidine and to identify them as effective GHS-R antagonists. These compounds were first run in the GHS-R binding assays and were further investigated in an intracellular Ca^{+2} flux activity assay.

The variation in the chemical space of these analogues is mainly focused around two phenyl rings, attached to the 2,4-diaminopyrimidine core. In order to investigate the scope of such chemical space around 2, 4-diaminopyrimidine moiety, a high dimensional quantitative structure-activity relationship (QSAR) study has been undertaken on these analogues to rationalize their GHS-R activity profiles. For this, it is necessary to characterize the molecules or their varying structural fragments from different perspectives. Among different methods, graph theoretical approaches provide large number of structural indices characteristic to the molecules and/or their functional units [17–21]. Moreover, when dealing with a large number of descriptors, for the optimum utilization of information content of the generated data sets, it is necessary to follow a typical protocol(s) to identify the best models as well as information rich descriptors corresponding to the phenomenon under investigation. The Combinatorial Protocol in Multiple Linear Regression (CP-MLR), developed recently [22–27], is an approach among many others to address the model evolution in high-dimensional QSAR studies. The aim of present communication is therefore, to establish the QSAR between the reported GHS-R antagonistic activity of 2,4-diaminopyrimidine derivatives and the molecular descriptors, obtained from graph theory.

Material and methods

Data set

In present study 2, 4-diaminopyrimidine derivatives (Table I) have been taken from the literature report [16] along with their antagonistic activity. The activity was expressed as the logarithm of the inverse of inhibitory concentration, pIC_{50}, where IC_{50} represents the molar concentration, required to bring out 50% inhibition of GHS-R. Two different approaches, namely the non-parametric and the parametric, have been used to develop the important QSARs of titled compounds.

The Fujita-Ban methodology [28], based on additivity principle, is a non-parametric approach and requires certain group to occur two or more times at a given varying position in a molecule. The Hansch type of analysis [29–31], on the other hand, is a parametric approach in which physicochemical and/or structural parameters are being used as the correlative parameters. This method is generally used to increase the understanding of the mechanisms of action of a set of congeners and to direct drug design in a congeneric series as well as to attempt to predict biological activities quantitatively. In general, the approach is to set up the equations involving different combinations of the substituents constants, then to allow the correlative methods to aid in the

Table I. The structures of the 2, 4-diaminopyrimidine analogues included in training set along with their GHS-R binding activity.
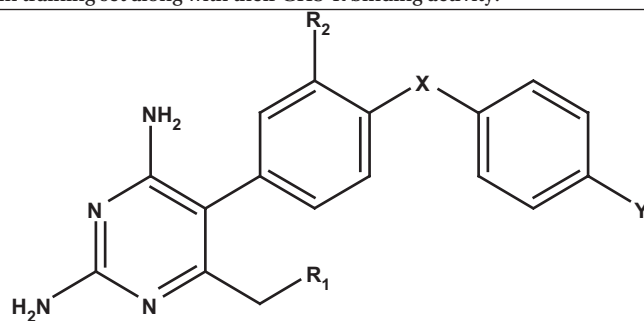
S. No.	R_1	R_2	X	Y	Binding IC_{50} (nM)[a]
1	CH_3	H	$NHCH_2$	Cl	310
2	CH_3	H	$NHCH_2$	SO_2CH_3	7.4
3[b]	CH_3	H	$NHCH_2$	SO_2CH_3	2.4
4	CH_3	H	$NHCH_2$	NO_2	11
5	CH_3	H	$NHCH_2$	$COCH_3$	57
6	CH_3	H	$NHCH_2$	$CHOHCH_3$	13
7	CH_3	H	$NHCH_2$	CN	160
8	CH_3	H	$NHCH_2$	SO_2CF3	21
9	OCH_2Ph	H	$NHCH_2$	Cl	19
10	OCH_2Ph	H	$NHCH_2$	NO_2	0.2
11	OCH_2Ph	H	$NHCH_2$	SO_2CH_3	0.3
12	OCH_2Ph	H	$NHCH_2$	SO_2CF3	1.0
13	OCH_2Ph	H	$NHCH_2$	$COCH_3$	1.2
14	OCH_2Ph	H	$NHCH_2$	CN	2.8
15	OCH_2Ph	H	$NHCH_2$	CF3	2.9
16	OCH_2Ph	H	$NHCH_2$	OCH_3	7.7
17	OCH_2Ph	H	$NHCH_2$	H	50
18	CH_2Ph	H	$NHCH_2$	Cl	3170
19	OCH_2Ph	H	OCH_2	SO_2CH_3	6.9
20	OCH_2Ph	H	OCH_2	CN	83
21	OCH_2Ph	H	CH_2NH	NO_2	0.8
22	CH_3	H	CH_2NH	NO_2	9.2
23	CH_3	H	CH_2NH	SO_2CH_3	9.8
24	CH_3	H	CH_2NH	Cl	300
25	CH_3	Cl	$NHCH_2$	SO_2CH_3	4.2
26	CH_3	Br	$NHCH_2$	SO_2CH_3	5.7
27	CH_3	H	$N(CH_3)CH_2$	SO_2CH_3	3.7

[a]Taken from Ref. [16]. [b]Compound having 3,5-F_2 at terminal phenyl ring.

selection of the 'best equation' justifying it statistically and avoiding the chance correlations. For the present study, the physicochemical parameters were taken from the literature [32]. The indicator variables, representative of the presence or absence of certain structural characteristic, have also been used. Both of these approaches have limitations as far as their use is concern for a new series of compounds. The Fujita-Ban methodology can not be employed if the appearance of certain substituent at a given position is less than two. Further, the study can not be extrapolated beyond the prevailing substituent's pattern in the chemical space. Similarly the Hansch type of approach is found unsuitable, if the desired physicochemical parameters have not been determined for certain substituents. The use of structural descriptors, instead of physicochemical parameters, may

then be proper to address the interaction involved at receptor sites.

In order to widen the scope of the study, the graph theory may be used as an alternative approach to obtain correlative descriptors. The highly informative descriptors, identified through certain mathematical procedure, may then be used to derive significant QSAR models to address the phenomenon under investigation. The DRAGON software [20], developed recently, is able to compute a large number of descriptors belonging to 0D-, 1D-, 2D- and 3-D classes. Prior to the computation of molecular descriptors, the structures of the compounds have been drawn in ChemDraw using the standard procedure. These structures were converted into 3D objects using the default conversion procedure implemented in the CS Chem3D Ultra. These 3D-structures were then ported to DRAGON software and a total number of 478 descriptors, pertaining to 0D-, 1D-, and 2D-classes, have only been computed for the sake of simplicity to interpret them in terms of common physical phenomenon. The descriptor classes along with their definitions and scope in addressing the structural features have been given in Table II. As the total number of descriptors involved in this study is very large, only the names of descriptor classes and the actual descriptor involved in the models have been addressed in the discussion. The QSAR model generation and validation have been done using the combinatorial protocol in multiple linear regression (CP-MLR) analysis.

CP-MLR

The CP-MLR is a 'filter'-based variable selection procedure for the development of statistical models in high dimensional QSAR studies [22–27]. It involves a combinatorial strategy with approximately placed 'filters' interfaced with MLR and extracts diverse models having unique combination of descriptors from the dataset. The filters set the thresholds for the descriptors in terms of inter-parameter correlation cutoff limits in subset regressions (filter-1), *t*-values of the regression coefficients (filter-2), internal explanatory power (filter-3; square root of adjusted multiple correlation coefficient of regression equation, *r*-bar), and the external consistency (filter-4; Q^2 i.e. cross-validated R^2 from the leave-one-out procedure). Throughout this study, the thresholds for the filters-1, 2 and 4 were assigned as 0.3, 2.0, and $0.3 \leq Q^2 \leq 1.0$, respectively while the filter-3 was assigned an initial value of 0.71. In order to collect the descriptors with higher information content, the threshold of filter-3 was successively incremented with increasing the number of descriptors (per model) by considering the r-bar value of the preceding optimum model as the new threshold for the next generation.

Descriptor classification protocol

The Three-stage descriptor classification protocol [23] is implemented with two-descriptor combinations (baseline models), as these are the simplest to understand and to explain the activity. In the first stage of the classification protocol, the correlations of the activity with two descriptor combinations from the individual descriptor classes (DCs)

Table II. Descriptor classes used for modeling the binding activity of the 2,4-diaminopyrimidine analogues and identified categories in modeling the activity.

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

[a]Reference [20]. [b]Descriptor categories identified at the end of second stage; in this the filter values are: filter-1 as 0.3, filter-2 as 2.0, filter-3 as 0.71, and filter-4 as $0.3 \leq Q^2 \leq 1.0$, the number of compounds in each dataset is 27.

of the dataset were used to sort them into four categories. They are primary contributors (category I: a DC forms a model with its constituent descriptors), collective contributors (category II: a DC unable to form a model with its constituent descriptors, but forms model(s) in combination with a descriptor from another such DC), secondary contributors (category III: a DC from a model(s) only in combination with category I) and noncontributors (category IV: a DC unable to form a model(s) in any manner like that of category I, II, and III). The sorted DCs were collated in the second stage to identify all the 3-descriptor models across the categories. In the last stage, the individual descriptors of all three-descriptor models were pooled to discover the higher models for the activity.

All the identified models have been put to the randomization test [24,33] by repeated randomization of the activity

to discover the chance correlations, if any, associated with them. For this every model has been subjected to 100 simulation runs with scrambled activity. The scrambled activity models with regression statistics better than or equal to that of the original activity model have been counted to express the percent chance correlation of the model under scrutiny. The model development procedure has been finally validated externally by creating test set from complete data set.

Results and discussion

Initially, the GHS-R Ca^{+2} flux activity was correlated to the GHS-R binding activity of all the 35 congeners to confer whether the two activities are inter-correlated to each other. The derived correlation between them is as in Equation (1)

$$pIC_{50}(Binding) = 0.774\ (\pm 0.129)\ pIC_{50}(Ca^{+2}\ flux)$$
$$+0.849$$
$$n = 35,\ r = 0.871,\ s = 0.414,$$
$$F = 103.470,\ p < 10^{-4}$$
(1)

Here and in the follow up discussion, n, r, s, F and p are respectively the number of data-points considered in the analysis, the regression coefficient, the standard error of estimate, the Fischer ratio and the significance of the model. Above correlation has divulged significant statistical parameters, which ensures that the GHS-R binding activity and the GHS-R intracellular Ca^{+2} flux activity are inter-related to each other. The subsequent elaboration is, therefore, based on the consideration of only the pIC_{50}(Binding) as the dependent variable.

In formulation of the Fujita-Ban matrix, eighteen compounds of Table I were retained in the training set, with compound **1** as the parent congener. However, nine compounds (**3, 6, 15–18, 25–27**) from this Table were not included in this set as certain substituents, at a given position, in these congeners occurred only once. The matrix consisting of 18 compounds (rows) and 9 substituents (including parent contribution) related to varying positions of the parent moiety (columns) is not documented here for the sake of brevity. The matrix was subjected to MRA and the derived statistical parameters of the study corresponding to the binding activity were:

$$n = 18,\ r = 0.991,\ s = 0.180,\ F = 60.753,\ p < 10^{-4}$$

All these statistical parameters tune to the highly significant results as F-value remained significant at 99% level while r-value has accounted for 98% of variance in observed activities. The calculated activity values (Table III) of the compounds considered in the study were in close agreement with the observed ones. The activity contributions of different substituents and parent moiety are given in Table IV. From this Table, the substituents that make higher positive contribution to activity, relative to parent moiety, may easily be identified. Thus, the substituent $OCH_2C_6H_5$ present at R_1 and

the substituents SO_2CH_3, NO_2, SO_2CF_3, $COCH_3$ and CN, in that order, present at position Y are advantageous. The appropriate substituents for varying positions, which make highest positive contribution to parent moiety may be selected for the further design of more active analogues of the series.

It is important to note that the Fujita-Ban approach cannot extrapolate beyond the compounds, considered in the training set whereas the Hansch approach, discussed below can do so. Thus, in the latter approach the data set has been extended to include the compounds which were not considered in Fujita-Ban study. In order to carry out the Hansch type of analysis, a number of physicochemical parameters for the R_1- and Y-substituents were selected and analyzed in a systematic manner for their correlations to GHS-R binding activity. The data-set, consisting of substituent constants such as hydrophobicity, π, hydrogen-bond donor, HD, hydrogen-bond acceptor, HA, electronic, σ (*meta* and *para*), field, F, resonance, R, dipole moment, μ, molar refraction, MR, molecular weight, MW and van der Waals volume V_w for each of the varying positions of the compounds were considered by permuting them to derive all possible models. Due to limited variations at incision X, the indicator variable was only considered. The highest significant correlation that could be emerged from these substituents constants is given by Equation (2)

$$pIC_{50} = 1.473(\pm 0.29)\ HA(R_1) + 2.005(\pm 0.40)\ MR_Y$$
$$-1.216(\pm 0.54)\ I_X$$
$$+1.108(\pm 0.38)\ I_Y + 5.377$$
$$n = 27,\ r = 0.924,\ s = 0.397,\ F = 32.338,$$
$$p < 10^{-4},\ Q^2_{LOO} = 0.791,\ Q^2_{L5O} = 0.812$$
(2)

where the indicator variables I_X and I_Y highlighting the presence of an $-OCH_2-$ at X-incision in between two phenyl rings and NO_2-substituent at Y-position, respectively. The values 1 or 0 for these variables indicate, respectively, the presence or absence of aforesaid structural features. The statistical indices, Q^2_{LOO} and Q^2_{L5O} have been obtained, respectively, by leave-one-out (LOO) and leave-five-out (L5O) procedures [34–36]. The derived statistical parameters have indicated highly significant results and Equation (2) as such reflected the parametric requirement of various substituents present at different positions of said analogues. The r^2-value has accounted for 85% of variance in the observed activity values while F-value remained significant at 99% level. The sufficiently high values of cross-validated indices, Q^2_{LOO} and Q^2_{L5O}, are both in favor of a robust and a highly predictable QSAR model. The calculated activity values using Equation (2) have remained in close agreement with the observed ones (Table III). From Equation (2), it appeared that a R_1-substituent having hydrogen-bond acceptor character and a Y-substituent being more bulky/polar are advantageous in improving the binding activity of a compound. Additionally, the presence of NO_2-substituent at Y and absence of $-OCH_2-$ at X are also essential.

Table III. Observed, calculated and predicted (LOO) pIC_{50} values of training set compounds.

S. No.	Obsd.	Calc. F.B.	Eq. (2)		Eq. (12)		Eq. (13)		Eq. (14)		Eq. (15)		Eq. (16)	
			Calc.	LOO	Calc.	LOO	Calc.	LOO	Calc.	LOO	Calc.	LOO	Calc.	LOO
1	6.51	6.48	6.59	6.59	6.47	6.47	6.33	6.28	6.64	6.66	6.68	6.71	6.50	6.50
2	8.13	8.10	8.08	8.08	7.67	7.64	8.16	8.16	8.14	8.14	8.18	8.18	7.96	7.94
3	8.62	–[b]	8.08	8.08	9.16	9.68	8.02	7.96	8.14	8.10	8.18	8.14	8.85	8.92
4	7.96	8.03	7.96	7.96	7.89	7.88	8.17	8.24	8.14	8.16	8.18	8.20	8.17	8.25
5	7.24	7.37	7.62	7.62	7.04	7.02	7.50	7.52	7.39	7.40	7.43	7.44	7.84	7.89
6	7.89	–[b]	7.55	6.65	7.04	6.96	7.53	7.51	8.04	8.12	7.43	7.40	7.86	7.84
7	6.80	6.97	6.65	7.96	7.24	7.36	6.81	6.82	6.64	6.61	6.68	6.67	6.96	6.98
8	7.68	7.62	7.96	8.06	8.14	8.19	8.03	8.07	8.14	8.18	8.18	8.23	8.01	8.04
9	7.72	7.90	8.06	9.43	7.79	7.79	7.53	7.50	7.77	7.78	7.80	7.81	7.70	7.70
10	9.70	9.46	9.44	9.56	9.23	9.10	9.39	9.30	9.44	9.39	9.45	9.41	9.24	9.05
11	9.52	9.53	9.55	9.43	8.98	8.90	9.05	8.99	9.28	9.24	9.29	9.26	9.02	8.95
12	9.00	9.05	9.43	9.09	9.43	9.52	9.02	9.02	9.28	9.31	9.29	9.33	9.08	9.10
13	8.92	8.80	9.09	8.12	8.38	8.33	8.41	8.35	8.52	8.49	8.54	8.51	8.91	8.91
14	8.55	8.40	8.12	7.86	8.61	8.63	8.16	8.02	7.94	7.86	7.97	7.88	8.01	7.95
15	8.54	–[b]	7.86	8.43	8.98	9.04	7.23	7.09	7.77	7.64	7.80	7.66	9.09	9.17
16	8.11	–[b]	8.43	7.06	7.76	7.72	8.40	8.46	8.47	8.49	8.49	8.52	7.71	7.63
17	7.30	–[b]	7.06	6.59	7.32	7.32	7.50	7.53	7.77	7.85	7.80	7.88	7.70	7.78
18	5.50	–[b]	6.59	8.34	6.70	6.92	6.72	6.92	5.80	5.95	5.87	6.05	5.60	5.66
19	8.16	8.19	8.34	6.90	8.12	8.11	8.54	8.74	8.72	8.94	8.65	8.93	8.55	8.71
20	7.08	7.06	6.90	9.43	7.76	7.83	7.86	7.91	7.63	7.87	7.57	7.88	7.54	7.76
21	9.10	9.36	9.44	7.96	9.23	9.27	9.60	9.77	9.43	9.48	9.44	9.50	9.23	9.28
22	8.04	7.93	7.96	8.08	7.89	7.86	8.40	8.48	8.12	8.12	8.16	8.17	8.16	8.21
23	8.01	8.00	8.08	6.59	7.67	7.65	8.10	8.11	8.12	8.13	8.16	8.17	7.95	7.95
24	6.52	6.37	6.59	8.08	6.47	6.46	6.57	6.58	6.64	6.65	6.68	6.71	6.50	6.50
25	8.38	–[b]	8.08	8.08	8.11	8.08	8.35	8.35	8.12	8.09	8.16	8.14	7.97	7.93
26	8.24	–[b]	8.08	8.08	8.11	8.09	8.48	8.50	8.12	8.11	8.16	8.15	7.98	7.95
27	8.43	–[b]	8.08	7.54	8.43	8.43	7.76	7.61	7.48	7.03	7.45	6.82	7.54	7.10

[a]pIC_{50} expressed as negative logarithm on molar basis, taken from Ref. [16]. [b]Not included in the Fujita-Ban study.

To widen the scope of present study, a high dimensional QSAR analysis has also been performed on these congeners. Twenty seven analogues, considered above, have been investigated to correlate their the GHS-R binding activity with a variety of 0D-, 1D- and 2D-descriptors computed from DRAGON software. The significant correlation, derived in such descriptors, may then be used to interpret a topological model in relation to commonly employed physicochemical properties. The emerged topological model(s), in this way, will have the wider perspective to describe GHS-R binding activity in relation with DRAGON descriptors. A total number of ten classes pertaining to 0D-, 1D, and 2D-descriptors have been implemented in this software. For present study, however, only seven classes were found important. A large number of descriptors from these classes have yielded significant correlations. These descriptors were identified as the primary contributors (category I) in modeling the inhibitory activity of these compounds. Nine significant models (Equations 3-11), identified in these descriptors and their statistical parameters, have been listed in Table V. From listed models in this Table, the inhibition activities of the compounds have been best explained by CONS, TOPO, BCUT, GVZ and 2DAUTO descriptor classes involving one- and two-descriptors. The CONS descriptors appeared in Equation (4), favors higher number of nitrogen and oxygen atoms (evinced, respectively, through nN and nO descriptors) in a molecular structure for enhanced activity. The TOPO class descriptor, S3K (3-path Kier alpha-modified shape index; information about centrality of branching) and T(N..O) (sum of topological distances between N and O) in Equation (6), favor the higher value of Kier alpha-modified shape index and topological distance between nitrogen and oxygen atom for improving the activity. These descriptors, collectively, have explained 70 percent of variance in the activity. The appeared descriptors from BCUT class in

Table IV. The Fujita-Ban contributions of substituents and parent moiety to the GHS-R binding activity of the 2,4-diaminopyrimidine analogues.

Position	Substituent	Contribution to pIC_{50}
Parent moiety	μ	6.476(±0.22)
R_1	$OCH_2C_6H_5$	1.429(±0.18)
X	CH_2NH	−0.103(±0.22)
	OCH_2	−1.346(±0.30)
Y	CN	0.497(±0.30)
	$COCH_3$	0.892(±0.31)
	NO_2	1.559(±0.26)
	SO_2CF3	1.149(±0.31)
	SO_2CH_3	1.628(±0.27)

Table V. The selected QSAR models, emerged in one and two descriptors from different descriptor classes belonging to category-I.

Descriptor's Class	Constant	Descriptor-1	Descriptor-2	r[a]	s	F	Q^2	Eq.
CONS	6.742	0.722nO	—	0.769	0.638	34.731	0.472	(3)
	3.860	0.544nN	0.755nO	0.815	0.590	22.822	0.525	(4)
TOPO	6.841	0.012T(N..O)	—	0.756	0.653	32.028	0.487	(5)
	4.828	0.396S3K	0.012T(N..O)	0.835	0.561	26.388	0.601	(6)
BCUT	−72.914	8.789BEHv8	29.021BELp2	0.791	0.624	19.181	0.475	(7)
GVZ	3.369	8.441GGI8	22.027JGI3	0.795	0.618	19.801	0.525	(8)
2D-AUTO	10.902	−10.23MATS7v	−3.771GATS1e	0.821	0.582	23.743	0.578	(9)
ACF	8.223	−1.128C-002	0.484O-058	0.754	0.669	15.160	0.449	(10)
PROP	2.291	0.031MR	0.032PSA	0.767	0.653	16.457	0.439	(11)

[a], in all the number of compounds (n) are 27, r is the correlation coefficient, Q^2 is cross-validated index from leave-one-out (LOO) procedure, s is the standard error of the estimate and F is the F-ratio between the variances of calculated and observed activities. **CONS**: nO, Number of oxygen atoms; nN, Number of nitrogen atoms; **TOPO**: T(N..O), Sum of topological distances between N..O; S3K, 3-Path Kier alpha-modified shape index; **BCUT**: BEHv8, Highest eigenvalue n.8 of Burden Marix/weighted by atomic van der Waals volumes; BELp2, Lowest eigenvalue n.2 of Burden Marix/weighted by atomic polarizabilities, **GVZ**: GGI8, Topological charge index of order 8; JGI3, Mean topological charge index of order 3; **2D-AUTO**: MATS7v, Moran autocorrelation of lag 7/weighted by atomic van der Waals volumes; GATS1e, Geary autocorrelation of lag 1/weighted by atomic Sanderson electronegativities; **ACF**: C-002, CH2R2; O-058, O =; **PROP**: MR, Ghose-Crippen molar refractivity; PSA, Fragment based polar surface area.

Equation (7) have revealed the importance of BEHv8, (highest eigenvalue n. 8 of Burden Marix/weighted by atomic van der Waals volumes) and BELp2 (lowest eigenvalue n. 2 of Burden Marix/weighted by atomic polarizabilities). The higher values of such descriptors have been found advantageous in improving the activity of a compound. In Equation (8), the descriptors, GGI8 (topological charge index of order 8) and JGI3 (mean topological charge index of order 3), both from GVZ class, have explained the involvement of charge indices of order 8 and 3 to augment the activity. Similarly the 2D-AUTO descriptors in Equation (9), the MATS7v (Moran autocorrelation of lag 7/weighted by atomic van der Waals volumes) and GATS1e (Geary autocorrelation of lag 1/weighted by atomic Sanderson electronegativities) have suggested the importance of lags 7 and 1 weighted by aforesaid properties. The significance of various emerged models (Table V) may be ascertained through the statistical parameters, r, s, and F. The cross-validated index Q^2, obtained from LOO procedure, may further assist in identifying the robustness of these models. The physical interpretation of resultant descriptors has been described briefly in the footnotes under Table V.

In order to have statistical significant correlations, the higher models, involving the descriptors from two different pools, have been derived further. The first pool comprise of 14 descriptors (Table V) which have been identified earlier as primary contributors under the category analysis. These descriptors were subjected to CP-MLR and the resulting two highly significant models, in three descriptors, are shown in Equation (12) and (13)

$$\text{pIC}_{50} = 0.853(0.201)\text{nN} + 9.332(1.066)\text{GGI8} + 27.530(6.007)\text{JGI3} - 1.752$$
$$n = 27, r = 0.886, s = 0.472, F = 28.006,$$
$$p < 10^{-4}, Q^2_{\text{LOO}} = 0.686, Q^2_{\text{L5O}} = 0.712 \tag{12}$$

$$\text{pIC}_{50} = 0.462(0.221)\text{nN} + 0.011(0.002)\text{T(N..O)} - 6.944(2.065)\text{MATS7v} + 4.510$$
$$n = 27, r = 0.853, s = 0.532, F = 20.451, p < 10^{-4},$$
$$Q^2_{\text{LOO}} = 0.633, Q^2_{\text{L5O}} = 0.591 \tag{13}$$

In above models, the descriptor nN is accounting for the number of nitrogen atoms in the structure under consideration. The positive regression coefficient of this descriptor recommends for more number of nitrogen atoms in a compound. Likewise, the higher values of the topological charge index of order 8, GGI8 and mean topological charge index of order 3, JGI3, are also essential for the improvement of activity of a compound. A higher value of sum of topological distances between N and O atoms [T(N..O)] and a lower positive value of MATS7v (Moran autocorrelation of lag 7/weighted by atomic van der Waals volumes) would enhance the binding activity.

The second pool of descriptors was formulated from all 10 DRAGON classes considered collectively. A total number of 478 descriptors have been analyzed, through CP-MLR analysis, to generate 87 models. The participating descriptors (the number being 44), in these models along with their average regression coefficients, and total incidences are listed in Table 6. From 87 such models, the highly significant two models are shown in Equation (14) and (15)

$$\text{pIC}_{50} = 0.012(0.002)\text{T(N..O)} + 0.655(0.221)\text{nHDon} - 0.835(0.149)\text{C} - 002 + 4.195$$
$$n = 27, r = 0.911, s = 0.419, F = 37.643, p < 10^{-4},$$
$$Q^2_{\text{LOO}} = 0.724, Q^2_{\text{L5O}} = 0.709 \tag{14}$$

Table VI. Descriptors identified for modeling the binding affinities of 2,4-diaminopyrimidine derivatives at the growth hormone secretagogue receptors along with the average regression coefficients and the total incidences.

Descriptor's Class	Descriptor[a]	Avg reg coeff (incidence)[b]	Descriptor[a]	Avg reg coeff (incidence)[b]	Descriptor[a]	Avg reg coeff (incidence)[b]
CONS	Mv	−40.938(3)	RBN	0.642(3)	RBF	28.436(1)
	nN	0.811(23)	—	—	—	—
TOPO	HNar	−5.473(1)	GNar	−8.763(1)	MAXDP	0.636(43)
	X0A	36.292(1)	XMOD	0.067(1)	IC2	2.446(8)
	TIC2	0.033(6)	TIC3	0.017(1)	T(N..N)	0.011(10)
	T(N..O)	0.012(9)	—	—	—	—
BCUT	BELm5	5.952(1)	BELm8	10.419(5)	BEHv8	8.628(6)
	BELv2	16.592(3)	BELv8	5.489(2)	BEHe2	11.367(3)
	BELe8	8.072(7)	BEHp8	9.263(6)	BELp8	5.262(2)
GVZ	GGI5	4.981(2)	GGI8	8.710(15)	JGI3	27.530(1)
2D-AUTO	MATS2v	11.081(7)	MATS7v	−10.027(1)	MATS8v	6.125(1)
	MATS5e	7.223(5)	GATS8v	−5.759(1)	GATS1e	−3.136(2)
	GATS8e	−2.814(2)	—	—	—	—
FUNC	nNHRPh	0.741(2)	nNO2Ph	1.012(14)	nROR	0.974(5)
	nRORPh	−0.910(1)	nHDon	0.649(7)	—	—
ACF	C-002	−0.980(16)	C-025	0.642(1)	H-047	0.120(1)
	O-061	0.506(1)	N-076	1.012(14)	—	—
PROP	MLOGP	−0.601(1)	—	—	—	—

[a]The descriptors are identified from the three parameter model emerged from CP-MLR protocol with filter-1 as 0.30; filter-2 as 2.0; filter-3 as 0.911; filter-4 as $0.3 \leq Q^2 \leq 1.0$; number of compounds in the study are 27; **CONS**: Mv, Mean atomic van der Waals volume (scaled on Carbon atom); RBN, Number of rotatable bonds; RBF, Rotatable bond fraction; nN, Number of nitrogen atoms; **TOPO**: HNar, Narumi harmonic topological index; GNar, Narumi geometric topological index; MAXDP, Maximal electrotopological positive variation; X0A, Average connectivity index of 0 order; XMOD, Modofied Randic chi-1 index; IC2, Information content index (neighbourhood symmetry of 2-order); TIC2 and TIC3, Total information content index (neighbourhood symmetry of 2 and 3-order, respectively); T(N..N) and T(N..O), Sum of topological distances between N..N and N..O, respectively; **BCUT**: BELm5 and BELm8, Lowest eigenvalue n.5 and 8, respectively, of Burden Marix/weighted by atomic masses; BEHv8, Highest eigenvalue n.8 of Burden Marix/weighted by atomic van der Waals volumes; BELv2 and BELv8, Lowest eigenvalue n.2 and 8, respectively, of Burden Marix/weighted by atomic van der Waals volumes; BEHe2, Highest eigenvalue n. 2 of Burden Marix/weighted by atomic Sanderson electronegativities; BELe8, Lowest eigenvalue n.8 of Burden Marix/weighted by atomic Sanderson electronegativities; BEHp8, Highest eigenvalue n.8 of Buden Marix/weighted by atomic polarizabilities; BELp8, Lowest eigenvalue n.8 of Burden Marix/weighted by atomic polarizabilities; **GVZ**: GGI5 and GGI8, Topological charge index of order 5 and 8, respectively; JGI3, Mean topological charge index of order 3; **2D-AUTO**: MATS2v, MATS7v and MATS8v Moran autocorrelation -lag 2, 7 and 8, respectively/weighted by atomic van der waals volumes; MATS5e, Moran autocorrelation-lag 5/weighted by atomic Sanderson electronegativities; GATS8v, Geary autocorrelation-lag 8/weighted by atomic van der waals volumes; GATS1e and GATS8e, Moran autocorrelation-lag 1 and 8, respectively/weighted by atomic Sanderson electronegativities; **FUNC**: nNHRPh, Number of secondary aromatic amines; nNO2Ph, Number of nitro groups (aromatic); nROR, Number of ethers (aliphatic); nRORPh, Number of ethers (aromatic); nHDon, Number of donor atoms for H-bonds (with N and O); **ACF**: C-002, CH2R2; C-025, R-CR-R; H-047 H attached to C1(sp3)/C0(sp2); O-061, O–; N-076, Ar-NO2/R-N(-R)-O/RO-NO2; **PROP**: MLOGP, Moriguchi octanol-water partition coefficient (logP). [b]The average regression coefficient of the descriptor corresponding to all models and the total number of its incidences; the arithmetic sign of the coefficient represents the actual sign of the regression coefficient in the models.

$$pIC_{50} = 0.011 \ (0.001) \ T(N..O) + 0.731(0.269) \ nNHRPh$$
$$- 0.812(0.151) \ C - 002 + 6.766$$
$$n = 27, \ r = 0.907, \ s = 0.429, \ F = 35.624, \ p < 10^{-4}, \quad (15)$$
$$Q^2_{LOO} = 0.684, \ Q^2_{L5O} = 0.724$$

where Q^2_{LOO} have addressed to a robust model and r-value accounted for 83 and 82 % of variances, respectively, in the observed activities. The newly appeared descriptors in above equations, nHDon (number of donor atoms for H-bonds with N and O) and nNHRPh (number of secondary aromatic amines) belong to FUNC class of descriptors. A higher value of these descriptors in addition to the higher value of topological distances between N and O atoms would increase the binding activity. The other emerged descriptor, from ACF class, C-002, is representative of the fragment

'CH2R2' in which two valences of a carbon atom are satisfied by hydrogen atoms and other two by alkyl groups. The negative regression coefficient of this descriptor advocates the absence of such fragments in a compound to have its improved activity. The descriptors of Table VI hold scope for even higher models. The best model, involving four descriptors, has been selected out of 41 models evolved by CP-MLR and is given in Equation (16)

$$pIC_{50} = 0.591 \ (0.074) \ MAXDP + 0.700 \ (0.217) \ nNO2Ph$$
$$+ 0.461 \ (0.209) \ nHDon - 1.013 \ (0.140) \ C$$
$$- 002 + 3.994 \quad (16)$$
$$n = 27, \ r = 0.925, \ s = 0.397, \ F = 32.397, \ p < 10^{-4},$$
$$Q^2_{LOO} = 0.761, \ Q^2_{L5O} = 0.694$$

The statistical parameters of Equation (16) have now improved over to that of Equation (14) and (15), justifying

Table VII. Cross-correlation matrix[a] amongst predictor variables in models.

	Eq. (2)					Eq. (12)		
	$HA(R_1)$	MR_Y	I_X	I_Y		nN	GGI8	JGI3
$HA(R_1)$	1.000	0.295	0.316	0.047	nN	1.000	0.205	0.157
MR_Y		1.000	0.042	0.235	GGI8		1.000	0.257
I_X			1.000	0.118	JGI3			1.000
I_Y				1.000				
	Eq. (13)					Eq. (14)		
	nN	T(N..O)	MATS7v			T(N..O)	nHDon	C-002
nN	1.000	0.022	0.197		T(N..O)	1.000	0.197	0.199
T(N..O)		1.000	0.250		nHDon		1.000	0.207
MATS7v			1.000		C-002			1.000
	Eq. (15)					Eq. (16)		
	T(N..O)	nNHRPh	C-002			MAXDP	nNO2Ph	nHDon	C-002
T(N..O)	1.000	0.183	0.199		MAXDP	1.000	0.008	0.094	0.000
nNHRPh		1.000	0.163		nNO2Ph		1.000	0.082	0.069
C-002			1.000		nHDon			1.000	0.207
					C-002				1.000
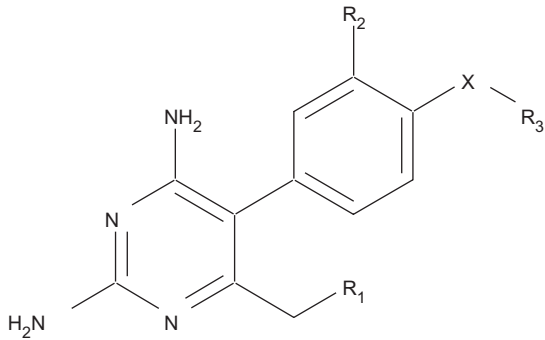
[a]Matrix elements are the *r*-values.
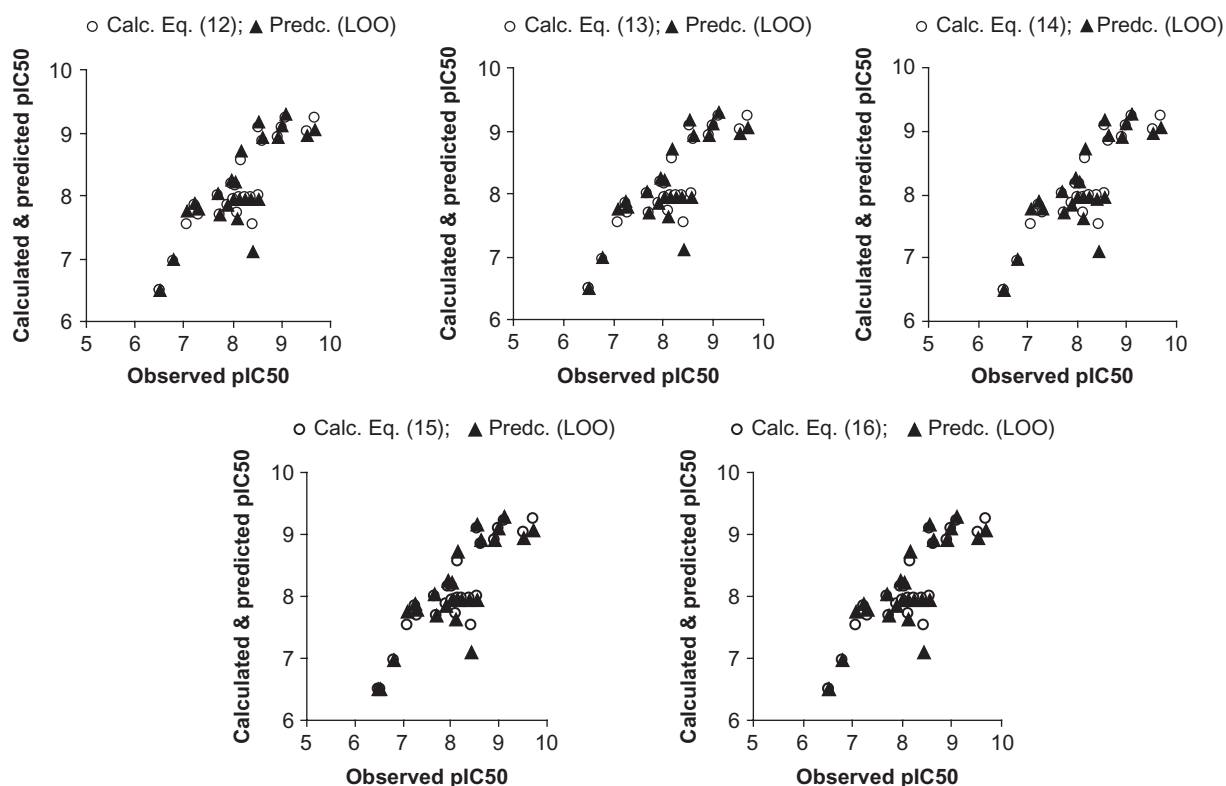
Figure 1. Plot of observed versus calculated and predicted pIC_{50} values.
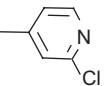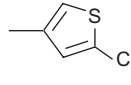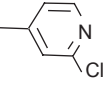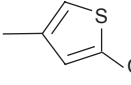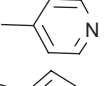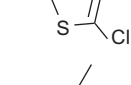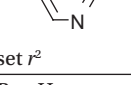
its superiority over to that of previously derived models. The TOPO class descriptor, MAXDP (the maximal electro-topological positive variation), which contributes positively to the activity and hence recommends maximum positive field effects for improvement of activity. The other FUNC class descriptor, nNO2Ph (number of aromatic nitro groups) recommends nitro substituted aromatic ring to augment the activity. A more number of donor atoms for H-bonds (nHDon) and absence of 'CH2R2' fragment (C-002) in a structure are essential to improve the binding activity of titled compounds.

All the descriptors appeared in above models had no mutual correlation is shown in Table VII. The calculated (using models) and predicted activities (through LOO procedure) have been found in close agreement to the observed ones (Table III). The graphical representation of the same is given in Figure 1 for the sake of clarity and goodness of fit.

Above model Equations (12–16) were further subjected to randomization process, where 100 simulations per model were carried out but none of the identified models has shown any chance correlation. Additionally, the above model Equations have been concomitantly validated with

Table VIII. The structures[a] of the 2, 4-diaminopyrimidine analogues included in test set and the predictions with corresponding test set r^2.

S.No.	R₁	R₃	Obsd.	Eq. (12)	Eq. (13)	Eq. (14)	Eq. (15)	Eq. (16)
				pIC₅₀ (M) Calculated
28	CH_3	(4-methyl-2-chloropyridin-5-yl)	6.96	7.26	6.54	6.64	6.68	6.50
29	CH_3	(4-methyl-2-chlorothiophen-5-yl)	6.54	6.38	5.82	6.64	6.68	6.50
30	OCH_2Ph	(4-methyl-2-chloropyridin-5-yl)	8.82	8.54	7.98	7.92	7.94	7.70
31	OCH_2Ph	(4-methyl-2-chlorothiophen-5-yl)	8.20	7.58	7.26	7.77	7.80	7.70
32	OCH_2Ph	(4-methylpyridin-3-yl)	7.68	8.17	8.24	7.92	7.94	7.70
33	OCH_2Ph	(5-methyl-2-chlorothiophen-3-yl)	7.37	7.58	7.45	7.77	7.80	7.70
34	OCH_2Ph	(tert-butyl)	7.17	7.78	7.44	7.77	7.80	7.70
35	OCH_2Ph	(3-methylquinolin-... yl)	6.89	8.36	7.57	7.91	7.93	7.72
		Test set r^2		0.456	0.497	0.563	0.558	0.552

[a]In all the structures: X =NHCH₂, R₂ = H.

the external test set of structurally dissimilar (bearing heteroaryl group at R₃) eight compounds listed in Table VIII. They have shown the test set r^2 values in the range of 0.456 to 0.563. The predictions of the test set compounds obtained with the models (12-16) and the corresponding predictive r^2 have also been given in the same Table.

Conclusions

Fujita-Ban and Hansch type of analyses, for 27 congeners of Table 1, have revealed the results which are complementary to each other. For example, the Hansch type of study has shown the importance of a hydrogen-bond acceptor substituent at R₁; the Fujita-Ban study, in conformity with it, has assigned highest substituent's contribution to a similar substituent such as $OCH_2C_6H_5$. Similarly, in Hansch type of analysis the substituent such as NO_2 having higher refraction parameter, accounting for bulk and polar effect, at Y

has been identified advantageous. In agreement to this, the Fujita-Ban study has emphasized the importance of substituents such as SO_2CH_3, NO_2, SO_2CF_3, and $COCH_3$. Both these studies have favored a –CH_2NH- type of substitution at incision X and H at R₂.

Since the descriptors in DRAGON software have been computed for the structure of a molecule as a whole, therefore, it may be improper to rationalize the individual variations in terms of evolved descriptors. Thus, for the data set in Table I, the descriptors appeared in highest significant model have advocated the importance of maximum positive field effects (MAXDP), more nitro substituted phenyl ring (nNO2Ph) in addition to a larger number of donor atoms for H-bonds containing substituents (nHDon). Likewise, the descriptor C-002 favored the absence of a particular fragment 'CH2R2' in a structure to enhance the GHS-R binding activity of titled compounds.

The present study may, therefore, provide further scope for substitutional modifications of 2, 4-diaminopyrimidine scaffold to bring out high GHS-R binding potential compounds. The substitutions having increased chain lengths, such as OCH_2CH_2Ph, at R_1 position and hydrogen donor, such as $N(OH)CH_2$, at X may be explored.

Acknowledgments

We are thankful to our Institutions for providing necessary facilities to complete this work.

References

1. Klein S, Romijn JP. Chapter 33-Obesity. In: *Williams Textbook of Endocrinology*. 10th ed.: Larsen PR, Kronenberg HM, Melmed S, Polonsky KS (Eds.), WB Saunders Co., St. Louis: MO; 2003. p 1619–1635.

2. Kojima M, Hosoda H, Date Y, Nakazato H, Matsuo H, Kangawa K. Ghrelin is a growth-hormone-releasing acylated peptide from stomach. *Nature* 1999;402:656–660.

3. Banks WA, Tschöp M, Robinson SM, Heiman ML. Extent and direction of ghrelin transport across the blood-brain barrier is determined by its unique primary structure. *J Pharmacol Exp Ther*. 2002;302:822–827.

4. Murikami N, Hayashida T, Kuroiwa T, Nakahara K, Ida T, Mondal MS, M. Nakazato M, Kojima M, Kangawa K. Role for central ghrelin in food intake and secretion profile of stomach ghrelin in rats. *J Endocrinol* 2002;174:283–288.

5. Wren AM, Seal LJ, Cohen MA, Brynes AE, Frost GS, Murphy KG, Dhillo WS, Ghatei MA, Bloom SR. Ghrelin enhances appetite and increases food intake in humans. *J Clin Endocrinol Metab* 2001;86:5992–5995.

6. Wren AM, Small CJ, Abbott CR, Dhillo WS, Seal LJ, Cohen MA, Batterham RL, Taheri S, Stanley SA, Ghatei MA, Bloom SR. Ghrelin causes hyperphagia and obesity in rats. *Diabetes* 2001;50:2540–2547.

7. Nakazato M, Murakami N, Date Y, Kojima M, Matsuo H, Kangawa K, Matsukura S. A role for ghrelin feeding in the central regulation of feeding. *Nature* 2001;409:194–198.

8. Bagnasco M, Tulipano G, Melis MR, Argiolas A, Cocchi D, Muller EE. Endogenous ghrelin is an orexigenic peptide acting in the arcuate nucleus in response to fasting. *Regul Pept* 2003;111:161–167.

9. Asakawa A, Inui A, Kaga T, Katsuura G, Fujimiya M, Fujino MA, Kasuga M. Antagonism of ghrelin receptor reduces food intake and body weight gain in mice. *Gut* 2003;52:947–952.

10. Sun Y, Wang P, Zheng H, Smith RG. Ghrelin stimulation of growth hormone release and appetite is mediated through the growth hormone secretagouge receptor. *Proc Natl Acad Sci USA* 2004;101:4679–4684.

11. Shuto Y, Shibasaki T, Otagiri A, Kuriyama H, Ohata H, Tamura H, Kamegai J, Sugihara H, Oikawa S. Hypothalamic growth hormone secretagouge receptor regulates growth hormone secretion. *J Clin Invest* 2002;109:1429–1436.

12. Liu B, Liu G, Xin Z, Serby M, Zhao H, Schaefer V, Falls D, Kaszubska W, Collins C, Sham H. Novel isoxazole carboxamides as growth hormone secretagouge receptor antagonists. *Bioorg Med Chem Lett* 2004;14:5223–5226.

13. Xin Z, Zhao H, Serby M, Liu B, Schaefer V, Falls D, Kaszubska W, Collins C, Sham H, Liu G. Synthesis and structure-activity relationships of isoxazole carboxamides as growth hormone secretagouge receptor antagonists. *Bioorg Med Chem Lett* 2005;15:1201–1204.

14. Zhao H, Xin Z, Liu G, Schaefer V, Falls D, Kaszubska W, Collins C, Sham H. Discovery of tetralin carboxamide growth hormone secretagouge receptor antagonists via scaffold manipulation. *J Med Chem* 2004;47:6655–6657.

15. Zhao H, Xin Z, Patel JR, Nelson LTJ, Liu B, Szczepankiewicz BG, Schaefer V, Falls D, Kaszubska W, Collins C, Sham H, Liu G. Structure-activity relationships on tetralin carboxamide growth hormone secretagouge receptor antagonists. *Bioorg Med Chem Lett* 2005;15:1825–1828.

16. Serby MD, Zhao H, Szczepankiewicz BG, Kosogof C, Xin Z, Liu B, Liu M, Nelson LTJ, Kaszubska W, Falls HD, Schaefer V, Bush EN, Shapiro R, Droz BA, Knourek-Segel VE, Fay TA, Brune ME, Beno DWA, Turner TM, Collins CA, Jacobson PB, Sham HL, Liu G. 2,4-Diaminopyrimidine derivatives as potent growth hormone secretagouge receptor antagonists. *J Med Chem* 2006;49:2568–2578.

17. Basak SC, Harriss DK, Magnuson VR. POLLY, University of Minnesota, Duluth, MN, 1988.

18. Molconn-Z, ver. 2.07, eduSoft Lc, a Virginia Corporation, Ashland, VA 23005 USA. www.edusoft-lc.com.

19. (a) Katritzky AR, Lobnov V, Karelson M. CODESSA (Comprehensive descriptors for structural and statistical analysis), University of Florida, Gainesville, FL, 1994. (b) Katritzky AR, Perumal S, Petrukhin R, Kleinpeter E. CODESSA-based theoretical QSPR model for hydantoin HPLC-RT lipophilicities. *J Chem Inf Comput Sci* 2001;41:569–574.

20. DRAGON software version 3.0-2003. By Todeschini R, Consonni V, Mauri A, Pavan M. Milano, Italy. http//disat.unimib.it/chm/Dragon.htm.

21. Gonzalez MP, Helguera AM. TOPS-MODE verces DRAGON descriptors to predict permeability coefficients through low-density polymer. *J Comput-Aided Mol Des* 2003;17:665–672.

22. Prabhakar YS. A combinatorial approach to the variable selection in multiple linear regression: Analysis of Selwood et al data set – A case study. *QSAR Comb Sci* 2003;22: 583–595.

23. Gupta MK, Prabhakar YS. Topological descriptors in modeling the antimalarial activity of 4-(3′,5′-disubstituted aniline)quinolines. *J Chem Inf Model*. 2006;46:93–102.

24. Prabhakar YS, Solomon VR, Rawal RK, Gupta MK, Katti SB. CP-MLR/PLS directed structure-activity modeling of the HIV-1 RT inhibitory activity of 2,3-diaryl-1,3-thiazolidin-4-ones. *QSAR Comb Sci* 2004;23:234–244.

25. Prabhakar YS. A combinatorial protocol in multiple linear regression to model gas chromatographic response factor of organophosphonate esters. *Internet Electron J Mol Des* 2004;3:150–162. http://www.biochempress.com.

26. Gupta MK, Sagar R, Shaw AK, Prabhakar YS. CP-MLR directed QSAR studies on the antimycobacterial activity of functionalized alkenols-topological descriptors in modeling the activity. *Bioorg Med Chem* 2005;13:343–351.

27. Saquib M, Gupta MK, Sagar R, Prabhakar YS, Shaw AK, Kumar R, Maulik PR, Gaikwad, AN, Sinha S, Srivastava AK, Chaturvedi V, Srivastava R, Srivastava BS. C-3 alkyl/arylalkyl-2,3-dideoxy hex-2-enopyranosides as antitubercular agents: Synthesis, biological evaluation and QSAR study. *J Med Chem* 2007;50:2942–2950.

28. Fujita T, Ban T. Structure-activity relation. 3. Structure-activity study of phenethylamines as substrates of biosynthetic enzymes of sympathetic transmitters. *J Med Chem* 1971;14:148–152.

29. Hansch C, Fujita T. p-s-p analysis. A method for the correlation of biological activity and chemical structure. *J Am Chem Soc* 1964;86:1616–1626.

30. Hansch C. Quantitative approach to biochemical structure-activity relationships. *Acc Chem Res* 1969;2:232–239.

31. Hansch C. In: Drug Design, Vol. I, Ariens EJ (Ed.) New York, N Y-Academic Press, 1971. p 271.

32. Hansch C, Leo A. Substituents constants for correlation analysis in chemistry and biology, New York, John Wiley, 1979.

33. So SS, Karplus M. Three-dimensional quantitative structure-activity relationship from molecular similarity matrices and genetic neural networks. 2. Applications. *J Med Chem* 1997;40:4347–4359.

34. Wold S, Eriksson L. Statistical validation of QSAR results. In: *Chemometrics methods in molecular design*, H. Van de Waterbeemed (Ed.), Wiley-VCH, Weinheim, 1995.

35. Efron B. Estimating the error rate of a prediction rule: Some improvement on cross-validation. *J Am Stat Assoc* 1983;78:316–331.

36. Efroymson MA. Multiple regression analysis. In: *Mathematical methods for digital computers*, Ralston A, Wilf HS (Eds.), Wiley-NewYork, 1960.

37. SYSTAT, Version 7.0 SPSS Inc., 444 North Michigan Avenue, Chicago, IL, 60611.