informa
healthcare

Human protein tyrosine phosphatase 1B inhibitors: QSAR by genetic function approximation

NARSINGH SACHAN, SHIVAJIRAO S. KADAM, & VITHAL M. KULKARNI

Department of Pharmaceutical Chemistry, Poona College of Pharmacy, Bharati Vidyapeeth Deemed University, Pune-411038, India

Abstract

Protein tyrosine phosphatase 1B (PTP 1B), a negative regulator of insulin receptor signaling system, has emerged as a highly validated, attractive target for the treatment of non-insulin dependent diabetes mellitus (NIDDM) and obesity. As a result there is a growing interest in the development of potent and specific inhibitors for this enzyme. This quantitative structure-activity relationship (QSAR) study for a series of formylchromone derivatives as PTP 1B inhibitors was performed using genetic function approximation (GFA) technique. The QSAR models were developed using a training set of 29 compounds and the predictive ability of the QSAR model was evaluated against a test set of 7 compounds. The internal and external consistency of the final QSAR model was 0.766 and 0.785. The statistical quality of QSAR models was assessed by statistical parameters r^2, r^2_{cv} (crossvalidated r^2), r^2_{pred} (predictive r^2) and lack of fit (LOF) measure. The results indicate that PTP 1B inhibitory activity of the formylchromone derivatives is strongly dependent on electronic, thermodynamic and shape related parameters.

Keywords: *Quantitative structure-activity relationship (QSAR), genetic function approximation (GFA), protein tyrosine phosphatase 1B (PTP 1B), non-insulin dependent diabetes mellitus (NIDDM), descriptor, inhibition*

Introduction

Insulin resistance in the liver and peripheral tissues coupled with elevated fasting plasma glucose and impaired glucose tolerance are the hallmarks of non-insulin dependent diabetes mellitus (NIDDM) [1]. Protein tyrosine phosphatases (PTPs) constitute a diverse family of enzymes that function as negative regulators of insulin signaling cascade and have been implicated as novel targets for the therapeutic enhancement of insulin action in insulin resistance states [2].

PTP 1B, a cytosolic PTP, appears to play a major role in insulin sensitivity and the dephosphorylation of the insulin receptor on the basis of many biochemical and cellular studies [3]. A recent pivotal PTP 1B knockout mice study revealed that mice lacking functional PTP 1B exhibited increased sensitivity towards insulin and are resistant to obesity [4]. These results taken together, establish a direct role for PTP 1B in down

regulating the insulin function. Recent insights into the mechanism of insulin actions have demonstrated that reversible tyrosine phosphorylation of the insulin receptor and its cellular substrate proteins play a central role in the mechanism of insulin action.

Among the PTP 1B inhibitors known so far, ertiprotafib the only drug candidate, developed by American Home Products, reached the stage of Phase II clinincal trials, but was withdrawn from clinical trial due to its undesirable side effects. More recently a PTP 1B antisence oligonucleotide ISIS-113715, developed by ISIS Pharmaceuticals inc. is in Phase I clinical trials [5]. As there is no successful molecule available in the market for this potential target, there is an urgent need for developing therapeutically useful PTP 1B inhibitors.

3D-QSAR methodologies such as comparative molecular field analysis (CoMFA), comparative molecular similarity analysis (CoMSIA) and genetic

Correspondence: V. M. Kulkarni, Department of Pharmaceutical Chemistry, Poona College of Pharmacy, Bharati Vidyapeeth Deemed University, Pune-411038, India. Tel: + 91-20-25437237. Fax: +91-20-25439383. E-mails: vmkulkarn.60@gmail.com

RIGHTSLINK()

function approximation (GFA) are used as valuable tools in designing new molecules. In this report we present the results of GFA carried out on formylchromone derivatives, reported as PTP1B inhibitors. We have used the GFA technique to generate different QSAR models from various descriptors generated using Cerius2 molecular modeling software. The GFA technique was used since it generates a population of equations rather than one single equation for correlation between biological activity and physico-chemical properties. GFA developed by Rogers involved combination of Friedman's multivariate adaptive regresssion splines (MARS) algorithm with Holland's genetic algorithm to evolve a population of equations that best fit the training set data [6]. This is done as follows:

(i) An initial population of equations is generated by random choice of descriptors. The fitness of each equation is scored by a lack-of-fit (LOF) measure

$$LOF = LSE/\{1 - [(c + d \times p)/m]\}^2$$

where LSE is least square error, c is the number of basis functions in the models, d is the smoothing parameter which controls the number of terms in the equation, p is the number of features contained in all terms of the models and m is the number of compounds in the training set.

(ii) Pairs from the population of equation are chosen at random and 'crossovers' are performed and progeny equations are generated.

(iii) The fitness of each progeny equation is assessed by LOF measure.

(iv) If the fitness of the new progeny equation is better, then it is preserved.

A distinct feature of GFA is that it produces a population of models. GFA models provide some useful additional information such as relevance of a particular descriptor in the model and activity prediction. GFA has been applied to a set of ellipticine analogues for different types of anticancer activities [7]. Recently, GFA was used for binding affinity predictions of ligands using free energy force field descriptor terms as in the case of inhibitors of glycogen phosphorylase [8] and peptidomimetic renin inhibitors [9,10]. GFA has also been used for the QSAR analysis of steroids, dopamine β-hydroxylase inhibitors [11] and anticancer agents [12]. An interesting application of GFA is in the QSAR studies on acetylcholinesterase inhibitors which has resulted in discovery of a new molecule, E2020, for the treatment of Alzheimer's disease [13]. Our strategy follows the methodology used previously to generate successful 3D-QSAR models and molecular modeling applied for antifungal [14], antibacterial [15], antiHIV-1 [16], antidiabetic [17,18], and antitubercular agents [19].

Materials and methods

Chemical data

Molecules. A series of 36 molecules belonging to formylchromone derivatives as PTP 1B inhibitors were taken from the literature and used [20,21]. The 3D-QSAR models were generated using a training set of 29 molecules. The observed and predicted biological activities of the training set molecules are presented in Table I. Predictive power of the resulting models was evaluated by a test set of 7 molecules with uniformly distributed biological activities. The observed and predicted biological activities of the test set molecules are presented in Table II. Selection of test set molecules was made by considering the fact that test set molecules represent a range of biological activity similar to the training set. The mean of the biological activity of the training and test set was -1.021 and -1.037, respectively. Thus the test set is a true representative of the training set.

Biological activity

The negative logarithm of the measured IC_{50} (μM) against human recombinant PTP 1B (h-PTP 1B) enzyme as pIC_{50} ($pIC_{50} = \log 1/IC_{50}$) was used as dependent variable, thus correlating the data linear to the free energy change. Since some compounds exhibited insignificant/no inhibition, such compounds were excluded from the present study.

Molecular modeling

Software. All molecular modeling studies were carried out using Cerius2 (version 4.10 L) running on Redhat Linux 3.0 [22]. Structures were constructed from the builder module and partial charges were assigned using the charge equilibration method within Cerius2 [23]. Throughout the study, Universal forcefield 1.02 was used [24]. The molecules were subsequently minimized until a root mean square deviation 0.001 kcal/mol Å was achieved and used in the study.

Calculation of descriptors

Different types of descriptors were calculated for each molecule in the study table using default settings within Cerius2. These descriptors include electronic, spatial, structural, thermodynamic and molecular shape analysis (MSA). A complete list of descriptors used in the study is given in given Table III.

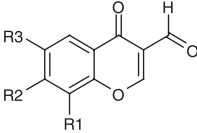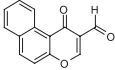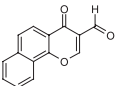Table I. Structures and biological activities of the training set molecules.
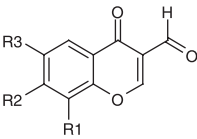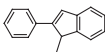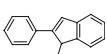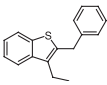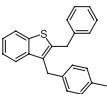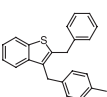
No.	R1	R2	R3	Obsd. act.[a]	Pred. act.[b]
1	H	H	-CH (CH$_3$)$_2$	−1.340	−1.360
2	H	H	Cl	−1.447	−1.396
3	H	H	Br	−1.301	−1.496
4	Br	H	Br	−1.301	−1.375
5	H	H	NO$_2$	−1.959	−2.091
6	H	H	F	−1.740	−1.444
7				−1.146	−1.241
8				−1.398	−1.251
9	H	H	-C$_6$H$_4$ C$_6$H$_5$	−0.634	−0.868
10	H	H		−0.851	−0.758
11	H	H		−0.398	−0.836
12	Br	H		−1.041	−0.715
13	H	H		−0.987	−0.765
14	H	H		−1.041	−0.911
15	Br	H		−1.204	−0.964
16	H	H		−0.887	−1.090
17	Br	H		−0.914	−1.049
18	H	H		−0.881	−0.899

Table I – *continued*

No.	R1	R2	R3	Obsd. act.[a]	Pred. act.[b]
19	Br	H		−1.00	−0.963
20	H	H		−0.792	−0.840
21	Br	H		−0.892	−0.882
22	H	H		−0.778	−1.095
23	H	H		−0.505	−0.677
24	C_6H_5	H	H	−1.204	−1.125
25	H	H		−0.0414	−0.442
26	C_6H_5	H		0.000	0.083
27	H	H	$C_6H_5CH_2$	−1.556	−1.159
28	$C_6H_5CH_2$	H	H	−1.255	−1.139
29	$C_6H_5CH_2$	H	$C_6H_5CH_2$	−1.114	−0.859

[a] Obsd. act = observed biological activity is defined as log 1/ IC50 against human recombinant PTP 1B enzyme (h-PTP 1B) in μM; [b] Pred. act = Predicted biological activity calculated using Equation (7) in Table IV.

MSA descriptors. MSA descriptors [25] were calculated using the MSA module within Cerius2. As MSA descriptors calculate three-dimensional properties of the ligands, knowledge of active conformer of the molecules under study is essential. The crystallographic conformation of the present series of molecules is not available / deposited at protein data bank. Hence conformational analysis on all molecules was performed using a random sampling search [26] and Universal force field [24], with the maximum number of conformers set equal to 150. The lowest energy conformer of the molecule with the highest biological activity (compound **26**, Table I) was used as reference for calculation of MSA descriptors.

In the present QSAR study no descriptor related to ligand-enzyme interactions (such as binding energy) has been used. Hence we presume that differences in binding orientations would have no effect on the conclusions from the present QSAR analysis.

Generation of QSAR models

QSAR analysis in computational research is responsible for the generation of models to correlate biological activity and physicochemical properties of a series of compounds. The underlying assumption is that the variations of biological activity within a series can be correlated with changes in measured or

Table II. Structures and observed, predicted activities along with residuals for the test set molecules.

No	R1	R2	R3	Obsd. act[a]	Pred. act[b]	Residual
1	H	H	H	-1.863	-1.463	-0.400
2	Cl	H	Cl	-1.398	-1.325	-0.073
3	H	CH_3	Cl	-1.255	-1.349	0.094
4	H	H	C_6H_5	-1.146	-1.162	0.016
5	H	H		-0.778	-1.049	0.271
6	C_6H_5	H	C_6H_5	-0.519	-0.809	0.273
7	C_6H_5	H	$C_6H_5\ C_6H_4$	-0.301	-0.511	0.189

[a] Obsd. act = observed biological activity is defined as log 1/ IC50 against human recombinant PTP 1B enzyme (h-PTP 1B) in μM; [b] Pred. act = Predicted biological activity calculated using Equation (6) in Table IV.

computed molecular features of the molecules. In the present study, the QSAR models were generated by the GFA technique. The application of the GFA algorithm allows the construction of high-quality predictive models and makes available additional information not provided by standard regression techniques, even for data sets with many features. GFA was performed using 100,000 crossovers, smoothness value of 1.0 and other default settings for each combination. The number of terms in the equation was fixed to four including constant in the training set. The set of equations generated were evaluated on the following basis: (a) LOF measure; (b) Variable terms in the equations; (c) Cross validated and non-cross validated r^2; (d) Randomized cross validated r^2; (e) Predictive ability of equation. Cross validated r^2 (r^2 cv), Randomized cross validated r^2, were calculated using the cross validated test option in the statistical tools supported in Cerius2.

The predictive r^2 was based only on molecules not included in the training set and is defined as: $r^2_{pred} = (SD - PRESS)/SD$, where SD is the sum of

Table III. Descriptors used in the present study.

No	Descriptor	Type	Descriptor
1	Vm	Spatial	Molecular volume[d]
2	Area	Spatial	Molecular surface area[d]
3	Density	Spatial	Molecular density[d]
4	RadOfGyr	Spatial	Radius of gyration[d]
5	PMI−X	Spatial	Principle moment of inertia X−component
6	PMI−Y	Spatial	Principle moment of inertia Y−component
7	PMI−Z	Spatial	Principle moment of inertia Z−component
8	PMI-mag	Spatial	Principle moment of inertia[d]
9	MW	Structural	Molecular weight[d]
10	RotlBonds	Structural	Number of rotatable bonds[d]
11	HbondAcc	Structural	Number of hydrogen bond acceptors[d]
12	HbondDon	Structural	Number of hydrogen bond donors[d]
13	AlogP98	Thermodynamic	Logarithm of partition coefficient[d]
14	MolRef	Thermodynamic	Molar refractivity[d]
15	Dipole-mag	Electronic	Dipole moment[d]
16	Dipole−X	Electronic	Diploe moment−X−component
17	Dipole−Y	Electronic	Dipole moment−Y−component
18	Dipole−Z	Electronic	Dipole moment−Z−component
19	Charge	Electronic	Sum of partial charges[d]
20	Apol	Electronic	Sum of atomic polarizabilities[d]
21	HOMO	Electronic	Highest occupied molecular orbital energy
22	LUMO	Electronic	Lowest unoccupied molecular orbital energy
23	Sr	Electronic	Superdelocalizability[d]
24	Foct	Thermodynamic	Desolvation free energy for octanol
25	Fh2o	Thermodynamic	Desolvation free energy for water
26	Hf	Thermodynamic	Heat of formation
27	DIFFV	MSA	Difference volume
28	COSV	MSA	Common overlap steric volume
29	Fo	MSA	Common overlap volume ratio
30	NCOSV	MSA	Non-common overlap steric volume
31	Shape RMS	MSA	RMS to shape reference
32	SR Vol	MSA	Volume of shape reference compound

[d] default descriptor

the squared deviations between the biological activity of molecules in the test set and the mean biological activity of the training set molecules and PRESS is the sum of squared deviations between predicted and actual activity values for every molecule in the test set. Like r_{cv}^2 the predictive r^2 can assume a negative value reflecting a complete lack of predictive ability of the training set for the molecules included in the test set [27,28].

Results

In the present study, QSAR models were generated using a training set of 29 molecules (Table I). A test set of 7 molecules (Table II) with regularly distributed biological activities was used to assess the predictive ability of the generated QSAR models. Biological activity was expressed in terms of pIC_{50}, the negative logarithm of measured IC_{50} (μM) against human recombinant PTP 1B (h-PTP 1B) enzyme. The conformational space of the rotatable bonds in the molecules was explored using a random sampling technique in order to obtain sterically accessible conformations within optimum computational time. Conformational search was performed during the molecular shape analysis (MSA) technique and the lowest energy conformer of each molecule was used for alignment using the MSA technique. All the molecules were superimposed on the lowest energy conformer of the molecule with highest biological activity (compound **26**). Pharmacophore alignment of the molecules used in the present study is shown in Figure 1. The GFA technique was used to generate the QSAR models. It was observed that in each case 100,000 crossovers and smoothing factor d = 1.0 resulted in optimum internal and external predictivity.
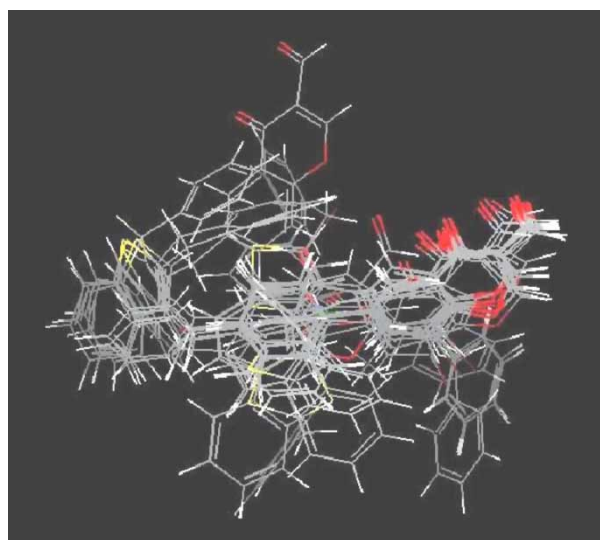

Figure 1.   Pharmacophore alignment of PTP 1B inhibitors used in the present QSAR study.

Hence the number of crossovers has been set to 100,000 for all other models.

Significance of molecular descriptors

The Cerius2 QSAR generates different descriptors belonging to different categories such as conformational, electronic, shape, spatial, thermodynamic, *etc*. Interpretation of QSAR models with more terms becomes difficult for drug design. Moreover all the terms may not be relevant. To obtain stable and consistent results from GFA and also to determine relevant descriptors, we have used a procedure to select a subset of descriptors, from a much large pool of descriptors. GFA was run several times by using molecular descriptors in several combinations to generate different QSAR models containing not more than four terms per equation.

Three models were generated using combination of different descriptors: Model A: Using 20 default descriptors (Table III, from 1–20 descriptors); Model B: Default + Descriptors from 21–26 in Table III (three electronic and three thermodynamic); Model C: Descriptors of Model A + Model B + six MSA descriptors (Table III). We have checked the "inter-variable correlation matrix" (option available within Cerius2) for the equations in all the models (Models A, B, C). This parameter was used to filter off the equations that were showing intercorrelation among the descriptors, even though those equations showed good statistical data. All the statistically significant equations for each QSAR model have been represented in Table IV. The term BA in the equations represents biological activity expressed as pIC_{50} values.

Model A

QSAR equations using GFA were generated using 20 default descriptors (Table III). The resultant equations were evaluated for their predictive power. The best equation from the set of equations was selected on the basis of predictivity, LOF and other statistical parameters such as F value. Equations (1), (2) and (3) (Table IV) showed better internal predictivity and also resulted in better predictions for the test set of molecules. The variable terms in the equations show low correlation among themselves indicating low probability of chance correlation. Equation (1) with better predictive r^2 value is proposed as the best QSAR equation for model A with 20 default descriptors for the present series of molecules.

Model B

This model was built by combination of default, three thermodynamic and three electronic descriptors. The generated set of QSAR equations were evaluated on

Table IV. Summary of the best equations selected from different GFA models. Equation (7) from Model C was selected to explain the observed PTP 1B inhibitory activity of formylchromone derivatives.

No.	Equation	LOF	r^2	r^2_{cv}	F- value	r^2_{pred}
Model A						
1	$\log(1/IC_{50}) = -2.172220 + 0.000752$ MR $- 0.033806$ Hbond acceptor $+ 0.003790$ Area	0.077	0.740	0.652	23.733	0.745
2	$\log(1/IC_{50}) = -1.958960 + 0.004640$ Vm $- 0.067095$ Hbond acceptor $- 0.010514$ Dipole-mag	0.075	0.747	0.636	24.661	0.718
3	$\log(1/IC_{50}) = -2.144630 + 0.001768$ MR $+ 0.000081$ Apol $- 0.04886$ Hbond acceptor	0.076	0.746	0.654	24.414	0.723
Model B						
4	$\log(1/IC_{50}) = -1.171580 - 0.062104$ Sr $- 0.112848$ Hbond acceptor $+ 0.000569$ PMI-mag	0.073	0.754	0.663	25.511	0.724
5	$\log(1/IC_{50}) = -2.112700 + 0.005140$ MR $- 0.056313$ Foct $- 0.000367$ HOMO	0.068	0.771	0.700	28.100	0.665
6	$\log(1/IC_{50}) = -1.260340 + 0.000047$ Hf $- 0.099213$ Hbond acceptor $+ 0.000516$ PMI-mag	0.074	0.751	0.659	25.113	0.671
Model C						
7	$\log(1/IC_{50}) = -2.056111 + 0.001121$ COSV $- 0.007113$ Sr $- 0.055721$ Foct	0.070	0.766	0.689	27.245	0.785
8	$\log(1/IC_{50}) = -2.057600 - 0.055482$ Foct $+ 0.00392$ Dipole-mag $+ 0.001127$ COSV	0.070	0.766	0.662	27.233	0.778
9	$\log(1/IC_{50}) = -2.068400 - 0.054748$ Foct $+ 0.001144$ COSV $+ 0.026906$ Shape RMS	0.069	0.767	0.677	27.371	0.785
10	$\log(1/IC_{50}) = -2.296100 + 0.001377$ COSV $- 0.054131$ Foct $- 0.057383$ Hbond acceptor	0.067	0.774	0.685	28.602	0.772
11	$\log(1/IC_{50}) = -2.033400 + 0.000035$ Hf $+ 0.000988$ COSV $- 0.053871$ Foct	0.069	0.768	0.691	27.600	0.751

the basis of cross validated r^2, non-cross validated r^2, LOF and frequency of variables used for model generation. Three equations were analyzed for their predictive power. Equation (4) with highest external predictivity was selected as the best QSAR equation for Model B (Table IV). Addition of six descriptors to the QSAR table increased the internal predictivity of the model moderately.

Model C

Deviation of biological activity for a series of molecules can be explained on the basis of differences in the physico-chemical descriptors. Hence, we considered using shape related descriptors in the generation of QSAR models. Six MSA descriptors were calculated using the MSA module added to the QSAR table, and model C was generated with all thirty-two descriptors (Table IV). These equations were analyzed on the basis of cross validated and non-cross validated r^2, LOF, F value and variable terms in the equation. The equations and variable terms in the equations clearly indicate the importance of electronic, shape and thermodynamic based factors in governing the biological activity of these compounds. Detailed statistical analysis of equations resulted in the identifications of five Equations (7)−(11) (Table IV) for Model C. Equation (7) (Table IV) was selected as a single best equation with proper balance of statistical terms for Model C. Inclusion of MSA based parameters clearly shows the improvement in the internal and external predictivity of model C. The internal predictivity of Equation (7) from Model C is more than Model A and B equations while external predictivity of Equation (7) from model C is more than model B and comparable to model A. This equation also has low LOF and higher F value than Model A and B. Therefore, this equation clearly shows the importance of shape related descriptors. Figures 2 and 3 show the graphs of actual and predicted activities of the training set molecules and graph of actual and predicted activities of test set molecules using Equation (7) from Model C respectively.
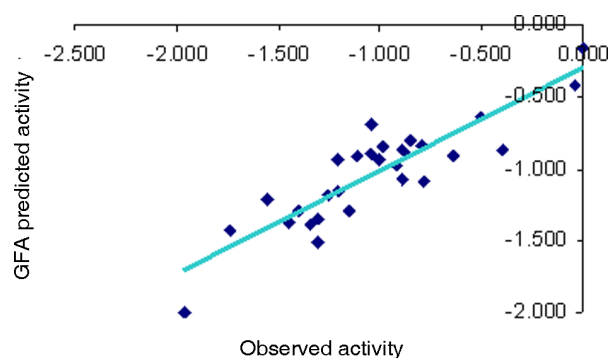

Figure 2. A graph of actual versus predicted activities of training set molecules using Equation (7) of Model C.
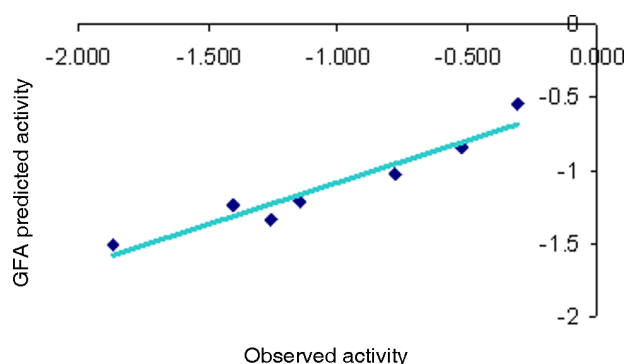
Figure 3. A graph of actual versus predicted activities of test set molecules using Equation (7) of Model C.

Equation (7) from Model C with proper balance of all statistical terms was selected as the best equation to explain the variance in the biological activity of formylchromones as PTP 1B enzyme inhibitors. The observed and predicted biological activities of training and test set molecules are given in Tables I and II respectively.

Randomization tests

To determine the model's reliability and significance, the randomization procedure was performed at 95% (19 trials) and 98% (49 trials) confidence levels. The randomization was carried out by repeatedly permuting the dependent variable set. If the score of the original QSAR model proved better than those from the permuted data sets, the model would be considered statistically significant. The results of randomization tests are shown in Table V. The correlation coefficient r^2 for the nonrandomized QSAR model was 0.766, better than those obtained from the randomized data. None of the permuted data sets produced an r^2 comparable to 0.766; hence, the value obtained from the original GFA model is significant.

Discussion

The Cerius2 QSAR module provides different descriptors divided into categories such as spatial, structural, electronic, conformational, thermodynamic and shape related descriptors. Among those some descriptors constitute a default set. Using this default set we have obtained a reasonably well predicted model (Model A) with cross validated r^2 (r^2_{cv}) of 0.652. Therefore in order to optimize the internal and external predictivity, the default descriptor set was extended in two different ways by including, (*a*) three electronic and three thermodynamic (Model B), (*b*) Descriptors of Model B + six MSA descriptors (Model C), available in the Cerius2 QSAR module to generate different models using GFA. With these additions the models were greatly improved in terms of internal and external consistency.

Interpretation of models

Model A. The equation describing biological activity for this model is Equation (1) (Table IV) containing Area–Spatial descriptor, MR–thermodynamic descriptor and HbondAcc–structural descriptor.

Area, a spatial descriptor is the molecular surface area that describes the van der Waals area of a molecule. The molecular surface area determines the extent to which a molecule exposes itself to the external environment. This descriptor is related to binding, transport and solubility. Area is positively correlated with the biological activity. The shape reference compound **26** possesses maximum area while the less active compounds **8** and **9** possess smaller surface area. It indicates that the active site is large and favors larger molecules.

MR is also a thermodynamic descriptor. Molecular refractivity index of a substituent is a combined measure of its size and polarizability. MR represents dispersion forces aiding the binding of an inhibitor to the enzyme, so positive correlation favors this binding. As MR is an approximation of molecular volume and its positive correlation with activity indicates that larger molecules would be more active than the smaller ones, then, compounds **25** and **26** being larger show better activity than compounds **5** and **6**.

The structural descriptor, HbondAcc strength is also significantly correlated with log $1/IC_{50}$, but the statistical criteria indicate a lower significance as it is negatively correlated. This equation showed low internal as well as external predicitivity. This indicates that other physicochemical parameters may be responsible for the variance in the biological activity of the present set of compounds.

Table V. Results of randomized r^2 for Equation (7) (Model C).

Confidence level	Trials	$r^2_{nonrandom}$	r^2_{random}	SD[a]	SD[b]	$r^2 <$ [c]	$r^2 <$ [d]
95%	19	0.766	0.098	5.336	0.208	19	0
98%	49	0.766	0.261	3.975	0.204	49	0

[a] Number of standard deviations of the mean value of r^2 of all random trials to the non-random r^2 value; [b] Standard deviation of the r^2 values of all random trials from the mean value of r^2; [c] Number of r^2 values from random trials that are less than the r^2 value for the non-random trial; [d] Number of r^2 values from random trials that are greater than the r^2 value for the non-random trial.

Model B. Among statistically significant QSAR Equations (4), (5) and (6) (Table IV) in Model B, Equation (4) with better predictive r^2 of 0.729 was selected as the representative for this class. The variable terms contributing to the biological activity in Equation (4) from Model B include one electronic descriptor–Sr, one spatial descriptor–PMI-mag and one structural descriptor–HbondAcc.

Sr is an electronic parameter. It is an index of reactivity in aromatic hydrocarbons (AH). This term was negatively correlated and indicates that the compounds having higher superdelocalizability show less activity.

Principal moment of inertia for any body describes the total mass of the body and its distribution. In the case of molecules, PMI-mag is a spatial descriptor describing orientation of the molecules and groups. In this case the spatial orientation of the molecule (in other words, conformation of the side chain) appears to be determining the activity profile. Thus for all the molecules in this class, similar conformation of the side chain is essential for activity. Requirement of conformational rigidity is important since structurally rigid elements are absent in these molecules. The importance of HbondAcc, a structural descriptor which appeared in the equations is the same as explained above.

Model C. QSAR equations for model C were generated using thirty-two descriptors (Table III). Equation (7) (Table IV) with better internal and external predictivity was selected as the representative equation for model C. The variable terms in this equation are Sr, Foct and COSV.

Structure-activity relationship of PTP 1B inhibitors with representative QSAR equation

Equation (7) (Model C, Table IV) with good internal and external predicitivity was selected as representative equation to explain the variance in the biological activity of PTP 1B inhibitors from the QSAR Models A, B, and C. This equation includes one electronic parameters–Sr, a shape parameter–COSV and a thermodynamic parameter–Foct contributing to the biological activity.

Sr. Superdelocalizability, an electronic parameter, is an index of reactivity of occupied and unoccupied orbitals. Sr is a second-order perturbation term indicating both the comparative chemical reactivities of different molecules and the stabilization energy in the formation of a complex with another molecule. Superdelocalizability can be described in three ways, one for each kind of attack; nucleophilic, electrophilic and radical. This term was negatively correlated

and indicates that the compounds with higher superdelocalizability show less activity as in the case of compound **1** while compound **23** with minimum Sr within the series shows better activity.

COSV. A shape descriptor, COSV is the common volume between shape reference and the other analogues. Because the shape reference molecule (compound **26**) is the one with highest biological activity and COSV is positively correlated, molecules that are structurally/conformationally similar to the most active molecule are expected to exhibit higher activity. Compounds **23** and **25** which are structurally and conformationally similar to reference molecule show better activity than the other compounds in the series such as **27**, **28** and **29**.

Foct. Foct, a thermodynamic descriptor which represents the octanol desolvation free energy of the molecule is correlated negatively with biological activity. Negative correlation of this term indicates that increase in the Foct value would result in decrease in activity as in the case of compound **5** which with highest foct value shows least activity and compounds **25** and **26** which with lowest foct values exhibit maximum activity.

Since most of the PTP 1B inhibitors are polar molecules containing groups, such as CO, CHO and O and act by forming H-bonds, steric and electrostatic interactions, there is hardly any contribution from desolvation entropy during binding. Hence, there is a negative correlation with biological activity.

Conclusions

A series of formylchromone derivatives was recently reported as potent, selective PTP 1B inhibitors with very good inhibitory activity against human recombinant PTP 1B enzyme. QSAR analysis was performed using statistical technique GFA, coupled with the use of combinations of different classes of descriptors. The generated models were analyzed for their statistical significance. A randomization test and intervariable correlations matrix were used to check the possibility of "chance correlation" for the generated equations. The models were also validated for their external prediction power.

GFA handled the physico chemical descriptors effectively in the generation of QSAR models with significant statistical terms including external predictivity. For the current series of molecules the descriptors COSV, Sr, and Foct appear to contribute significantly to the observed biological activity. Thus current QSAR analysis reveals that shape, electronic and thermodynamic descriptors contribute significantly to the biological activity of PTP 1B inhibitors.

The selected model has good internal and external predictivity. This equation explains about 77% ($r^2 = 0.766$) variance in the biological activity. As the results from the present QSAR analysis agree with the other previously reported series these studies may be used together for the design of newer compounds with better PTP 1B inhibitory activity.

Acknowledgements

References

[1] Defronzo RA. Pathogenesis of type 2 diabetes: Metabolic and molecular implications for identifying diabetic genes. Diabet Rev 1997;5:177–267.

[2] Blaskovich MA, Kim HO. Recent discovery and development of protein tyrosine phosphatase inhibitors. Exp Opin Ther Patent 2002;12:871–905.

[3] Byon JCH, Kusari J, Kusari AB. Protein–tyrosine phosphatase-1B acts as a negative regulator of Insulin signal transduction. Mol Cell Biochem 1998;182:101–108.

[4] Elchelby M, Payette P, Michalisyn E, Cromlish W, Collinschan CC, Ramachandran C, Gresser MJ, Tremblay ML, Kennedy BP. Structure of protein tyrosine phosphatase 1B in complex with inhibitors bearing two phosphotyrosine mimetics. Science 1999;283:1544–1548.

[5] Elizabeth AH, Nigel L. Protein tyrosine phosphatase 1B inhibitors for the treatment of type 2 diabetes and obesity; recent advances. Cur Opin Invest Drugs 2003;4:1179–1189.

[6] Rogers D, Hopfinger AJ. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. J Chem Inf Comput Sci. 1994;34:854–866.

[7] Shi LM, Fan Y, Myers TG, O'Connor PM, Paull KD, Friend SH, Weinstein JN. Mining the NCI anticancer drug discovery databases: Genetic function approximation for the QSAR study of anticancer ellipticine analogues. J Chem Inf Comput Sci 1998;38:189–199.

[8] Venkatarangan P, Hopfinger AJ. Prediction of ligand-receptor binding thermodynamics by free energy force field three-dimensional quantitative structure-activity relationship analysis: Applications to a set of glucose analogue inhibitors of glycogen phosphorylase. J Med Chem 1999;42:2169–2179.

[9] Tokarski JS, Hopfinger AJ. Constructing protein models for ligand-receptor binding thermodynamic simulations: An application to a set of peptidometic renin inhibitors. J Chem Inf Comput Sci 1997;37:779–791.

[10] Tokarski JS, Hopfinger AJ. Prediction of ligand-receptor binding thermodynamics by free energy force field (FEFF) 3D-QSAR analysis: Application to a set of peptidometic renin inhibitors. J Chem Inf Comput Sci 1997;37:792–811.

[11] Hahn M, Rogers D. Receptor surface models. 2 application to quantitative structure-activity relationships studies. J Med Chem 1995;38:2091–2102.

[12] Hopfinger AJ, Kawayami Y. QSAR analysis of a set of benzothiopyranoindazole anti-cancer analogs based upon their DNA intercalation properties as determined by molecular dynamics simulation. Anticancer Drug Des 1992;7:203–217.

[13] Kawakami Y, Inoue A, Kawai T, Wakita M, Sugimoto H, Hopfinger AJ. The rationale for E2020 as a potent acetylcholinesterase inhibitor. Bioorgan Med Chem 1996;4:1429–1449.

[14] Gokhale VM, Kulkarni VM. Understanding the antifungal activity of terbinafine analogues using quantitative structure–activity relationship (QSAR) models. Bioorgan Med Chem 2000;8:2487–2499.

[15] Karki RG, Kulkarni VM. Three-dimensional quantitative structure–activity relationship (3D-QSAR) of 3-aryloxazolidin-2-one antibacterials. Bioorgan Med Chem 2001;9:3153–3160.

[16] Makhija MT, Kulkarni VM. QSAR of HIV-1 integrase inhibitors by genetic function approximation method. Bioorgan Med Chem 2002;10:1483–1497.

[17] Vadlamudi SM, Kulkarni VM. 3D-QSAR of protein tyrosine phosphatase 1B inhibitors by genetic function approximation. Internet Electron. J Mol Des 2003;3:586–609.

[18] Murthy VS, Kulkarni VM. Molecular modeling of protein tyrosine phosphatase 1B (PTP 1B) inhibitors. Bioorg Med Chem 2002;10:897–906.

[19] Kharkar PS, Desai B, Varu B, Loriya R, Naliyapara Y, Gaveria H, Shah A, Kulkarni VM. Three-dimensional quantitative structure-activity relationship of 1, 4-dihydropyridines as antitubercular agents. J Med Chem 2002;45:4858–4867.

[20] Shim YS, Kim KC, Chi DY, Lee KH, Cho H. Formylchromone derivatives as a novel class of protein tyrosine phosphatase 1B inhibitors. Bioorgan Med Chem Lett 2003;13:2561–2563.

[21] Shim YS, Kim KC, Lee KA, Shrestha S, Lee KH, Kim CK, Cho H. Formylchromone derivatives as irreversible and selective inhibitors of human protein tyrosine phosphatase 1B. Kinetic and modeling studies. Bioorgan Med Chem 2005;13:1325–1332.

[22] Cerius2 Version 4.10 is available from Molecular Simulations Inc.: 9685, Onton Road, San Diego CA, 92121, USA.

[23] Rappe AK, Goddard WA, III Charge equilibration for molecular dynamics simulations. J Phys Chem 1991;95:3358–3363.

[24] Rappe AK, Casewit CJ, Colwell KS, Goddard GA, Skiff WM. Application of a universal force field to organic molecule. J Am Chem Soc 1992;114:10024–10035.

[25] Hopfinger AJ. A QSAR investigation fo dihydrofolate reductase inhibition by Baker triazines based upon molecular shape analysis. J Am Chem Soc 1980;102:7196–7206.

[26] Mayer D, Naylor CI, Motoc I, Marhsall G. A unique geometry of the active site of angiotensin converting enzyme consistent with structure-activity studies. J Comput Aided Mol Design 1987;1:3–16.

[27] Waller CL, Opera TL, Giollitti A, Marshal GR. Three-dimensional QSAR of human immunodeficiency virus (1) protease inhibitors. 1. A CoMFA study employing experimentally-determined alignment rules. J Med Chem 1993;36:4152–4160.

[28] Crammer RD, III, Bunce JD, Patterson DE. Cross-validation, boot strapping and partial Least squares compared with multiple regressions in conventional QSAR studies. Quant Struct Act Relat 1998;7:18–25.