

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/233753309>

QSAR Comb. Sci. 26, 2007, No. 5, 608 617

DATASET · NOVEMBER 2012

READS

18

4 AUTHORS, INCLUDING:



Seema Bag

University of Massachusetts Boston

28 PUBLICATIONS **196** CITATIONS

SEE PROFILE



Nilesh Ramesh Tawari

Institute of Chemical Technology, Mumbai

23 PUBLICATIONS **180** CITATIONS

SEE PROFILE



Mariam Degani

Institute of Chemical Technology, Mumbai

54 PUBLICATIONS **392** CITATIONS

SEE PROFILE

3-D-QSAR Analysis of 2-(Oxalylamino) benzoic acid Class of Protein Tyrosine Phosphatase 1B Inhibitors by CoMFA and Cerius2.GA

Subbiah Ramar, Seema Bag, Nilesh R. Tawari and Mariam S. Degani*

Department of Pharmaceutical Sciences and Technology, Institute of Chemical Technology, University of Mumbai, Matunga (E), Mumbai 400019, India, E-mail: msdegani@udct.org

Keywords: Antidiabetic activity, Cerius2.GA, Comparative molecular field analysis, PTP1B inhibitors, QSAR.

Received: July 13, 2006; Accepted: October 17, 2006

DOI: 10.1002/qsar.200630090

Abstract

Three-Dimensional Quantitative Structure – Activity Relationship (3-D-QSAR) studies for 2-(oxalylamino) benzoic acids as Protein Tyrosine Phosphatase 1B (PTP1B) inhibitors were performed using the genetic function approximation algorithm (Cerius2.GA) and Comparative Molecular Field Analysis (CoMFA). A five-term equation was developed with crossvalidated r^2 (r^2_{CV}) and conventional correlation coefficient (r^2) of 0.642 and 0.748, respectively, using Cerius2.GA. The CoMFA model from multifit alignment produced good statistical significance with r^2_{CV} and r^2 of 0.671 and 0.900, respectively. The predictive abilities of Cerius2.GA and CoMFA were determined using a test set of ten molecules giving predictive correlation coefficients of 0.666 and 0.643, respectively, indicating good predictive power. Furthermore, the robustness of the models was verified by bootstrapping analysis. Based on the information derived from Cerius2.GA and CoMFA, we have identified some key features that may be used to design new derivatives and predict their PTP affinities prior to synthesis.

1 Introduction

Diabetes mellitus has been recognized as a genetic disease. It is an endocrine disorder caused by the deficiency or absence of insulin or due to the interference in insulin utilization. Type II (NIDDM) is the most common form of diabetes. Insulin resistance in liver and peripheral tissues together with the pancreatic cell defects, are the common causes of type II diabetes mellitus and human obesity [1].

Tyrosine phosphorylation of specific intracellular proteins controlled by the actions of Protein Tyrosine Kinases (PTKs) and Protein Tyrosine Phosphatases (PTPs) is recognized as a key process by which a number of polypeptide hormones and growth factors transduce and coordinate their biological effects *in vivo* [2]. PTPs constitute a diverse family of enzymes, which are responsible for the selective dephosphorylation of tyrosine residues [3].

Szczepankiewicz *et al.* have reported NMR-based screening with iterative structure-based drug design approach to design PTP1B inhibitors [4]. Other studies using X-ray crystallographic structure of enzyme have given insight into PTP1B inhibitors. Zhou and Ji [5] have used docking and performed Three-Dimensional Quantitative

Structure-Activity Relationship (3-D-QSAR) on few molecules from the series reported by Andersen *et al.*

Several papers indicate the usefulness of 3-D-QSAR along with X-ray crystallographic studies, NMR-based and docking studies in the design of new compounds, as this would lead to a better understanding of ligand–receptor interactions and thus to a better design of new molecules [6, 7, 8]. The contour maps obtained from 3-D-QSAR models aid in visualizing the comparative region of differences in the related enzymes and also suggest the flexibility in protein structure, which may not be observed in the X-ray structure analysis [9]. It is also known that as the numbers of active derivatives in a series increase, the formulation of a useful SAR without quantification becomes increasingly difficult. Thus, molecular models developed using 3-D-QSAR can be used to improve interpretation of pharmacological data and predict novel biologically active compounds in the series [10].

In the present work, 3-D-QSAR study of 2-(oxalylamino) benzoic acid analogues of Protein Tyrosine Phosphatase 1B (PTP1B) inhibitors with antidiabetic activity has been carried out using Cerius2.GA and Comparative Molecular Field Analysis (CoMFA). We have used X-ray crystallographic structural data (PDB code: 1C88) [19] to

enhance the predictivity of the QSAR (CoMFA) model. Studies with Cerius2.GA using GFA gave similar outputs, which seem to validate the CoMFA model. The predictability of each model thus generated was evaluated using a test set not included in the model creation.

The contour maps produced from the CoMFA model permit an understanding of the steric, electrostatic requirements for ligand binding. As a consequence, the structural variations in the training set that give rise to variations in the molecular fields at particular regions in space are correlated to biological activity [11]. The 3-D-QSAR models thus generated revealed useful information about the requirements for the development of new PTP1B inhibitors.

2 Methods

2.1 Dataset for Analysis

A set of 58 molecules belonging to 2-(oxalylamino) benzoic acids, reported as nonpeptide classical competitive inhibitors of PTP1B by Andersen *et al.* have been considered in this study (Table 1) [12]. Out of these, ten molecules *i.e.*, molecules 44 and 50–58 for which biological activity was not reported, could not be included in the generation of the model. The remaining 48 molecules were divided into training set (38 molecules) and test set (ten molecules, molecule number 06, 08, 14, 17, 24, 31, 40, 42, 46, and 48) using chemical as well as biological diversity as guidelines for division [13].

2.1.1 Biological Data

Activity of all the compounds used in the QSAR study has been measured by the same assay and was reported as K_i . The compounds were tested against a soluble form of recombinant human PTP1B (amino acid residues 1–321), using *p*-nitrophenyl phosphate (*p*NPP) as substrate, the inhibitory constant values were measured at pH 5.5 [12]. The negative logarithm of the measured K_i , *i.e.*, pK_i was calculated and used in the 3-D-QSAR study, thus correlating the data linear to the free energy change. The log activity difference between the most active and the least active analogues was of the order of approximately four.

The K_i values were converted into pK_i according to the formula,

$$pK_i = -\log_{10} K_i$$

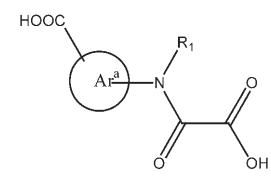
2.2. Molecular Modeling

2.2.1. Cerius2.GA

The molecular modeling studies were performed using the software Cerius2 (version 3.5) installed on a Silicon

Table 1. Dataset used for QSAR studies [12]

(a)



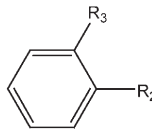
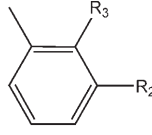
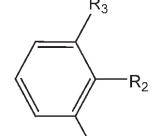
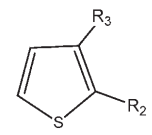
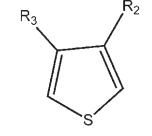
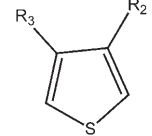
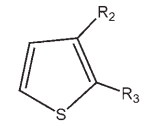
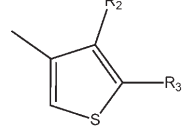
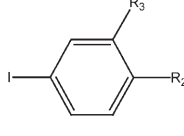
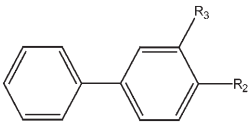
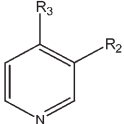
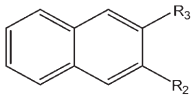
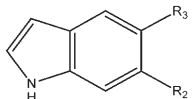
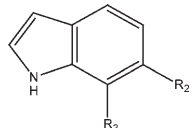
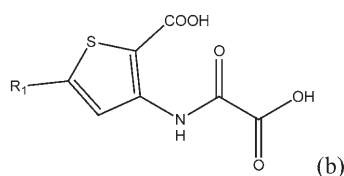
| Serial no. | Ar ^a | R ₁ | Biological activity ^b |
|------------|--|----------------|----------------------------------|
| 1 |  | H | −1.3617 |
| 2 |  | H | −3.2305 |
| 3 |  | H | −3.0414 |
| 4 |  | Me | −3.0414 |
| 5 |  | H | −2.0334 |
| 6 |  | H | −1.8751 |
| 7 |  | H | −1.5682 |
| 8 |  | H | −3.3010 |
| 9 |  | H | −1.1461 |

Table 1. (cont.)

| Serial no. | Ar ^a | R ₁ | Biological activity ^b |
|------------|---|----------------|----------------------------------|
| 10 |  | H | −1.1461 |
| 11 |  | H | −2.2041 |
| 12 |  | H | −0.9956 |
| 13 |  | H | −0.9031 |
| 14 |  | H | −1.2553 |

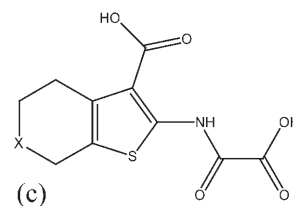
^a The oxalamide is attached at the R₂ position and the carboxy group is attached at the R₃ position.


^b Biological activity is expressed as log (1/K_i).



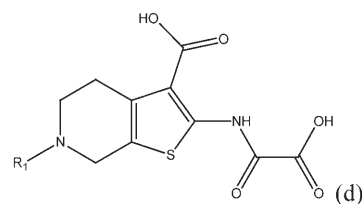
| Serial no. | R ₁ | Biological activity ^a |
|------------|-------------------------------|----------------------------------|
| 15 | Ph | −1.1139 |
| 16 | 3-Thienyl | −1.0414 |
| 17 | 4-(i-Bu)-Ph | −1.1139 |
| 18 | 4-F-Ph | −0.9031 |
| 19 | 4-Cl-Ph | −1.0000 |
| 20 | 4-OH-Ph | −0.6532 |
| 21 | 4-OMe-Ph | −0.9542 |
| 22 | 4-PhO-Ph | −1.4150 |
| 23 | 4-BnO-Ph | −1.2041 |
| 24 | 4-(HOOCCH ₂ -O)-Ph | −0.3979 |
| 25 | 3-NO ₂ -Ph | −0.6721 |
| 26 | 3-NH ₂ -Ph | −0.7853 |
| 27 | 3-OMe-Ph | −1.0792 |
| 28 | 3,4-OMe-Ph | −1.3979 |
| 29 | 3,5-OMe-Ph | −0.9031 |

^a Biological activity expressed as log (1/K_i) against PTP1B.



| Serial no. | X | n | Biological activity ^a |
|------------|---|---|----------------------------------|
| 30 | CH ₂ | 0 | −1.0792 |
| 31 | CH ₂ | 1 | −0.9085 |
| 32 | CH ₂ | 2 | −1.0792 |
| 33 | O | 1 | −1.1461 |
| 34 | S | 1 | −0.9191 |
| 35 | SO | 1 | −0.7076 |
| 36 | SO ₂ | 1 | −0.7853 |
| 37 | C=O | 1 | −0.1139 |
| 38 | CH-OH | 1 | −0.3010 |
| 39 |  | 1 | −0.3424 |

^a Biological activity expressed as log (1/K_i) against PTP1B.



| Compound | R ₁ | Biological activity ^a |
|----------|---|----------------------------------|
| 40 | −H | 0.5376 |
| 41 | −CH ₃ | 0.0000 |
| 42 | −CH ₂ Ph | 0.0000 |
| 43 | −CH ₂ -(3-MeO-Ph) | 0.1549 |
| 44 | −CH ₂ -(2-naphthyl) | ND |
| 45 | −CH ₂ -(2-pyridyl) | 0.0000 |
| 46 | −CH ₂ -(3-Pyridyl) | 0.0458 |
| 47 | −CH ₂ -(4-Pyridyl) | −0.0414 |
| 48 | −CH ₂ -(2-quinolyl) | −0.2553 |
| 49 | −(CH ₂) ₂ Ph | 0.5686 |
| 50 | −(CH ₂) ₂ -(1-naphthyl) | ND ^b |
| 51 | −(CH ₂) ₂ -(1-biaryl) | ND |
| 52 | −(CH ₂) ₂ -(4-BnO-Ph) | ND |
| 53 | −(CH ₂) ₂ -(2-BnO-Ph) | ND |
| 54 | −(CH ₂) ₂ -(3-thienyl) | ND |
| 55 | −(CH ₂) ₂ -(2-pyridyl) | ND |
| 56 | −(CH ₂) ₂ -(4-pyridyl) | ND |
| 57 | −(CH ₂) ₂ -CH ₂ -Ph | ND |
| 58 | −(CH ₂) ₂ -CH-Ph ₂ | ND |

^a Biological activity expressed as log (1/K_i) against PTP1B.

^b ND: not defined.

Graphics O2 R5000 workstation [14]. The molecules were constructed using a 3-D sketcher in the BUILD option. Partial charges were assigned using charge equilibration method within Cerius2. All the molecules were energy minimized until the root mean square deviation reached $0.01 \text{ kcal} \cdot \text{mol}^{-1} \text{ \AA}^{-1}$.

2.2.1.1. Calculation of Descriptors

Different types of descriptors were calculated for each molecule in the study table using default settings within the Cerius2. These descriptors include electronic, spatial, structural, thermodynamic, and Molecular Shape Analysis (MSA) descriptors. A complete list of different physico-chemical descriptors used for the QSAR study is given in Table 2. As MSA descriptors calculate 3-D properties of ligands, conformational analysis of all molecules was performed using random sampling search with maximum number of conformers limited to 50 [15]. The lowest energy conformer of the molecule with highest biological activ-

ity (compound 49) was used as reference for calculation of MSA descriptors.

2.2.1.2. Generation of QSAR Models

QSAR analysis generates the relationship between physicochemical descriptors including 3-D features and biological activity. The underlying assumption is that the variations in biological activity within a series can be correlated with changes in measured or computed molecular features of the molecules. Application of Cerius2.GA allows the construction of high quality predictive models and provides additional information, which is not the case in standard regression techniques, even for a dataset with many features. GFA was performed using 25000 crossovers and smoothness value set to 1.0 and other default settings were maintained. The number of terms in the equation was fixed to five including a constant. Crossvalidated r^2 (r_{CV}^2) was calculated using crossvalidated test option in the statistical tools supported in Cerius2.

Table 2. Descriptors used for the generation of present QSAR models

| No | Descriptor | Family | Description |
|----|-------------------|----------------|--|
| 1 | Energy | Conformational | Energy of the selected conformation |
| 2 | Charge | Electronic | Sum of partial charges |
| 3 | Fcharge | Electronic | Sum of formal charges |
| 4 | Apol | Electronic | Sum of atomic polarizabilities |
| 5 | Dipole-mag | Electronic | Dipole moment magnitude |
| 6 | Dipole-X | Electronic | Dipole moment-X-component |
| 7 | Dipole-Y | Electronic | Dipole moment-Y-component |
| 8 | Dipole-Z | Electronic | Dipole moment-Z-component |
| 9 | HOMO | Electronic | Highest occupied molecular orbital energy |
| 10 | LUMO | Electronic | Lowest unoccupied molecular orbital energy |
| 11 | Sr | Electronic | Superdelocalizability |
| 12 | DIFFV | MSA | Difference volume |
| 13 | Fo | MSA | Common overlap volume ratio |
| 14 | NCOSV | MSA | Noncommon overlap steric volume |
| 15 | COSV | MSA | Common overlap steric volume |
| 16 | ShapeRMS | MSA | RMS to shape reference |
| 17 | SR Vol | MSA | Volume of shape reference compound |
| 18 | RadOfGyr. | Spatial | Radius of Gyration |
| 19 | Area | Spatial | Molecular surface area |
| 20 | Density | Spatial | Molecular density |
| 21 | PMI-mag | Spatial | Principle moment of inertia |
| 22 | PMI-X | Spatial | Principle moment of inertia-X-component |
| 23 | PMI-Y | Spatial | Principle moment of inertia-Y-component |
| 24 | PMI-Z | Spatial | Principle moment of inertia-Z-component |
| 25 | Vm | Spatial | Molecular volume |
| 26 | MW | Structural | Molecular weight |
| 27 | Rotlbonds | Structural | Number of rotatable bonds |
| 28 | H bond acceptor | Structural | Number of H bond acceptors |
| 29 | H bond donor | Structural | Number of H bond donors |
| 30 | $A \log P$ | Thermodynamic | Logarithm of partition coefficient |
| 31 | Fh ₂ O | Thermodynamic | Desolvation free energy for water |
| 32 | Foct | Thermodynamic | Desolvation free energy for octanol |
| 33 | Hf | Thermodynamic | Heat of formation |
| 34 | Molref | Thermodynamic | Molar refractivity |

2.2.2 CoMFA

CoMFA was performed using SYBYL 6.6 installed on a Silicon Graphics Indy R 5000 workstation [16]. Structural manipulations were performed using standard Tripos force field [17], partial atomic charges of the molecules were calculated using an AM1 model Hamiltonian within MOPAC [18].

As the crystallographic conformation of all the molecules of the series used were not available, the initial conformation of the most active molecule (compound 49) was derived by necessary modifications from compound 40, extracted from X-ray crystallographic structure of PTP1B-compound 40 complex (PDB code: 1C88) [19]. Furthermore, simulated annealing was performed, as it enables the rapid identification of good solutions, ideally the global minimum. The system was heated to 1000 K for 1.0 ps and then annealed to 250 K for 1.5 ps. The annealing function was exponential, 50 such cycles were run and the resulting 50 conformers were optimized using Powell method with a convergence of 0.001 kcal/mol. The default settings for all other minimization options and a distance dependent dielectric of 1.0 were employed throughout the calculation. All the other molecules were constructed using standard geometries and standard bond lengths using the nearest structures obtained from simulated annealing. A constrained minimization was performed initially in which the uncharged part of the new inhibitor molecule was defined as aggregate. The constraints were then removed and minimization was repeated by Powell method till the Root Mean Square Deviation (RMSD) of 0.001 kcal/mol was achieved.

2.2.2.1 Alignment Rules

Superimposition of molecules was carried out by the following alignments:

(1) Atom-based alignment: This alignment involved the RMS fitting (atom-based fitting) of the heavy atoms of the ligands. The compounds were fitted on the template molecule (compound 49), making use of the heavy atoms of the common functionality present in all compounds of this series.

(2) Multifit alignment: In this case, alignment of the molecules was carried out by flexible fitting (multifit) of the atoms of the ligands to the template molecule, compound 49. This involved energy calculation and fitting on the template molecule using force (force constant 20 kcal·mol⁻¹ Å⁻¹) followed by energy minimization. The heavy atoms of the template compound 49 were considered for this alignment (Figure 1). All aligned molecules are shown in Figure 2.

(3) Field fit alignment: This was carried out using SYBYL QSAR rigid body field fit command within SYBYL and using compound 49 as template molecule. Field fit adjusts the geometry of the ligands such that the fields match with that of the template molecule.

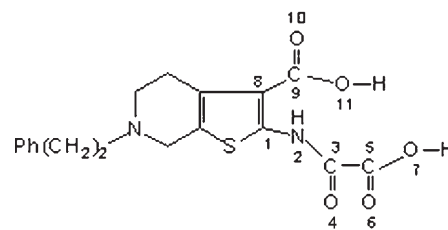


Figure 1. Compound 49 with atoms used for superimposition are marked.

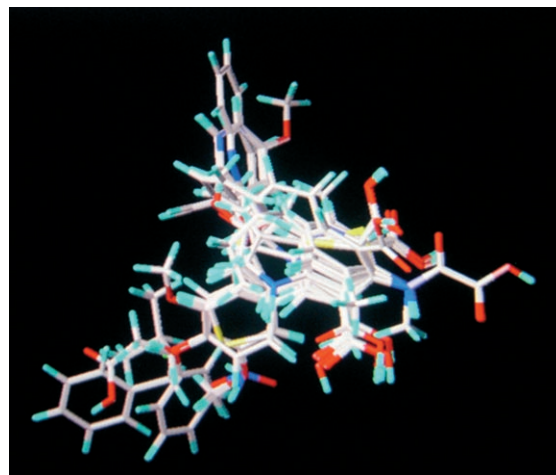


Figure 2. Alignment of all molecules in the training set and test set.

2.2.2.2 CoMFA Interaction Energies

The steric and electrostatic CoMFA fields were calculated at each lattice intersection of a regularly spaced grid of 2.0 Å in all three dimensions within defined region. The van der Waals potential and columbic terms representing the steric and electrostatic fields respectively were calculated using standard Tripos force fields. A distance-dependent dielectric constant of 1.00 and a sp³ carbon atom with +1.00 charge was used as a probe atom. The steric and electrostatic fields were truncated at +30.00 kcal/mol, and the electrostatic fields were ignored at the lattice points with maximal steric interactions.

2.2.2.3 Calculation of Additional Descriptors

Various physicochemical descriptors used in the CoMFA study were calculated using Cerius2 molecular modeling software version 3.5 running on Silicon Graphics, O2 R5000 workstation. These include *A* log *P*, HOMO, and LUMO. These descriptors were directly used as additional regressors in the PLS analysis.

2.2.2.4 Partial Least Squares (PLS) Analysis

PLS method was used to linearly correlate the CoMFA fields with the inhibitory activity values. The crossvalidation analysis was performed using the Leave-One-Out (LOO) method in which one compound is removed from the dataset and its activity is predicted using the model derived from the rest of the dataset. The r_{CV}^2 that resulted in the optimum number of components and lowest standard error of prediction were considered for further analysis. Equal weights were assigned to steric and electrostatic fields using COMFA_STD scaling option. To speed up the analysis and reduce noise, a minimum filter value of 2.00 kcal/mol was used. Final analysis was performed to calculate conventional r^2 using the optimum number of components. To further assess the robustness and statistical confidence of the derived models, bootstrapping analysis for 100 runs was performed. Bootstrapping involves the generation of many new datasets from the original dataset and is obtained by randomly choosing samples from the original dataset. The statistical calculation is performed on each of these bootstrapping samplings. The difference between the parameters calculated from the original dataset and the average of the parameters calculated from the many bootstrapping samplings is a measure of the bias of the original calculations. The entire crossvalidated results were analyzed with the assumption that a value of r_{CV}^2 above 0.3 indicates that probability of chance correlation is less than 5% [16].

2.2.2.5 Predictive r^2 Values

The predictive ability of each 3-D-QSAR model was determined by a set of ten compounds not included in the training set. These molecules were aligned, and their activities were predicted. The predictive correlation coefficient ($r_{pred.}^2$), based on molecules of the test set, is defined as

$$r_{pred.}^2 = (\text{SD-PRESS})/\text{SD}$$

where SD is the sum of the squared deviations between the biological activities of the test set and mean activities of the training set molecules and PRESS is the sum of

squared deviation between predicted and actual activity values for every molecule in a test set.

3. Results and Discussion

3.1 Cerius2.GA

Different QSAR equations were generated using the genetic function algorithm in Cerius2. A total of 38 compounds were used for the QSAR model generation (Table 1). Ten compounds were used as the test set. The same test set was used both for Cerius2.GA and CoMFA. A preliminary study was undertaken to study the effect of a number of crossovers and the value of the smoothing parameter “ d ”. GFA crossover of 25000 and d value of 1.0 gave reasonable convergence. Hence, the number of crossovers was set as 25000 for all models and the length of equation was fixed to five terms. Different sets of equations (models) were generated using combination of different descriptors. The selection of the best model was based on the values of r^2 (square of the correlation coefficient for the training set of compounds), r_{CV}^2 , LOF (Friedman's Lack of Fit), $r_{pred.}^2$ (predictive r^2 for the test set of compounds).

The following four models were generated using a combination of different descriptors:

- (1) Model A: Using default descriptors.
- (2) Model B: Default + Thermodynamic descriptors.
- (3) Model C: Default + MSA descriptors.
- (4) Model D: Combination of all descriptors.

The statistically significant equations from different QSAR models are given in Table 3.

3.1.1. Validation of the QSAR Model

The derived QSAR models were validated for their reliability and significance by two methods: (1) Randomization and (2) LOO crossvalidation.

(1) Randomization: The best Cerius2.GA model was selected for validation by randomization method by keeping 95% (19 trials) and 98% confidence (49 trials). The randomization was done by repeatedly permuting the depen-

Table 3. Summary of the best QSAR equation selected from different GFA model

| Model | Equation | LOF | r^2 | r_{CV}^2 | F -test | $r_{pred.}^2$ | Descriptors |
|-------|---|-------|-------|------------|-----------|---------------|-------------------------|
| A | BA = -8.65871 - 0.007187*PMI _{imag} + 0.653434*H bond donor + 0.025082*MW + 0.008207*PMI-Y | 0.394 | 0.646 | 0.552 | 19.64 | 0.552 | Default |
| B | BA = -5.19768 + 0.015612*H bond donor - 0.469018*Rotlbonds + 0.133278*Dipole-Z | 0.357 | 0.669 | 0.527 | 16.71 | 0.526 | Default + Thermodynamic |
| C | BA = -6.73191 - 0.007722*NCOSV + 0.018086*MW + 0.475943*H bond donor - 0.000738*PMI-Y | 0.278 | 0.743 | 0.649 | 23.81 | 0.649 | Default + MSA |
| D* | BA = -3.38147 + 0.007167*COSV + 0.574957*H bond donor + 0.12875*Dipole-Y - 1.0031*LUMO | 0.272 | 0.748 | 0.642 | 24.53 | 0.666 | All |

D* = This model was found to be statistically the most significant.

dent variable set. If the score of the original QSAR model proved to be better than those from the permuted datasets, the model would be considered statistically significant, better than those obtained from the permuted dataset. The randomization test results from 19 trials and 49 trials are shown in Table 4. The correlation coefficient r^2 for the nonrandom QSAR model was 0.748, significantly better than those obtained from randomized data. None of the permuted sets produced an r^2 comparable with 0.748; hence the value obtained for the original Cerius2.GA model is significant.

(2) LOO crossvalidation: Crossvalidation of 0.748 by LOO method indicated that the results obtained for the best QSAR equation from model D were not by chance correlation.

QSAR model D obtained from a combination of all descriptors showed the best internal and external predictivity among all the generated models. Summary of the ten best

GFA equations of model D is given in Table 5. The nonvalidated r^2 and r_{CV}^2 values obtained were 0.748 and 0.642, respectively. The predictive correlation coefficient r_{pred}^2 value was found to be 0.666. The goodness of fit is also corroborated by the LOF value of 0.272. The plot of predicted (GFA) versus actual activity for training set molecules is shown in Figure 3. The test set residuals of GFA analyses are shown in Figure 4. The best QSAR equation generated by the GFA run is given below.

QSAR Equation:

$$BA = -3.38147 + 0.007167 * \text{COSV} + 0.544957 * \text{H bond donor} + 0.12875 * \text{Dipole-Y} - 1.0031 * \text{LUMO}$$

According to this, the observed PTP1B inhibitory activity for 2-(oxalylamino) benzoic acid class of compounds is influenced by H bond donor, COSV, Dipole-Y, and LUMO. The descriptors H bond donor, COSV, and Dipole-Y positively correlate with biological activity and LUMO negatively correlate with PTP1B inhibitory activity.

Table 4. Results of randomization tests

| Confidence level | Trials | $r_{\text{nonrandom}}^2$ | r_{random}^2 (mean) | SD ^a | SD ^b | $r^2 <^c$ | $r^2 >^d$ |
|------------------|--------|--------------------------|------------------------------|-----------------|-----------------|-----------|-----------|
| 95% | 19 | 0.748 | 0.328 | 3.775 | 0.077 | 19 | 0 |
| 98% | 49 | 0.748 | 0.0912 | 5.398 | 0.104 | 49 | 0 |

^aNumber of standard deviations of the mean value of r^2 of all random trials to the nonrandom r^2 value.

^bStandard deviation of the r^2 value of all random trials from the mean value of r^2 .

^cNumber of r^2 values from random trials that are less than the r^2 value for the nonrandom trial.

^dNumber of r^2 values from random trials that are greater than the r^2 value for the nonrandom trial.

3.2 CoMFA

The first step in CoMFA was choosing a suitable alignment method. Analysis A (Table 6) shows results obtained from the three different alignments. Comparing the predictive ability of the three CoMFA models from analysis A, the model generated from multifit alignment exhibited best overall statistical results with r_{CV}^2 of 0.646, external predictive ability r_{pred}^2 of 0.645, and high bootstrapped r^2 of 0.911. The steric and electrostatic contributions of this model were 66.10% and 33.90%, respectively. Thus, the QSAR

Table 5. Summary of good GFA equations of model D

| Equation no. | GFA equation | LOF | r^2 | r_{CV}^2 | BS r^2 | F-Test | r_{pred}^2 |
|--------------|--|-------|-------|------------|----------|--------|--------------|
| 1 | BA = -3.38147 + 0.007167 * COSV + 0.544957 * H bond donor + 0.12875 * Dipole-Y - 1.0031 * LUMO | 0.272 | 0.748 | 0.642 | 0.749 | 24.525 | 0.666 |
| 2 | BA = -2.93752 + 0.422641 * H bond donor + 0.11739 * COSV + 0.177586 * PMI-X | 0.277 | 0.744 | 0.105 | 0.744 | 23.930 | 0.604 |
| 3 | BA = -6.73191 - 0.007722 * NCOSV + 0.475943 * H bond donor + 0.018086 * MW - 0.00738 * PMI-Y | 0.278 | 0.743 | 0.649 | 0.743 | 23.813 | 0.623 |
| 4 | BA = -6.7657 - 0.007422 * NCOSV + 0.018143 * MW + 0.488689 * H bond donor - 0.00052 * PMI-mag | 0.280 | 0.741 | 0.648 | 0.741 | 23.559 | 0.648 |
| 5 | BA = -5.00793 - 1.66038 * Fo + 0.016066 * MW - 0.014988 * NCOSV + 0.518696 * H bond donor | 0.281 | 0.740 | 0.649 | 0.740 | 23.447 | 0.614 |
| 6 | BA = -6.74422 + 0.01793 * MW - 0.007297 * NCOSV + 0.497263 * H bond donor - 0.000683 * PMI-Z | 0.282 | 0.739 | 0.632 | 0.734 | 23.304 | 0.610 |
| 7 | BA = -5.81525 - 0.002987 * Energy + 0.01394 * MW - 0.007183 * NCOSV + 0.501672 * H bond donor | 0.285 | 0.736 | 0.643 | 0.737 | 23.012 | 0.639 |
| 8 | BA = -5.49137 + 0.506002 * H bond donor - 0.158542 * LUMO + 0.012699 * MW - 0.00822 * NCOSV | 0.285 | 0.736 | 0.620 | 0.737 | 23.00 | 0.670 |
| 9 | BA = -5.81603 - 0.00297 * LowEne - 0.007187 * NCOSV + 0.013939 * MW + 0.501696 * H bond donor | 0.285 | 0.736 | 0.643 | 0.737 | 23.00 | 0.639 |
| 10 | BA = -5.83397 - 0.007336 * H bond donor + 0.015642 * MW - 0.125802 * Rotlbonds | 0.285 | 0.736 | 0.640 | 0.740 | 22.971 | 0.617 |

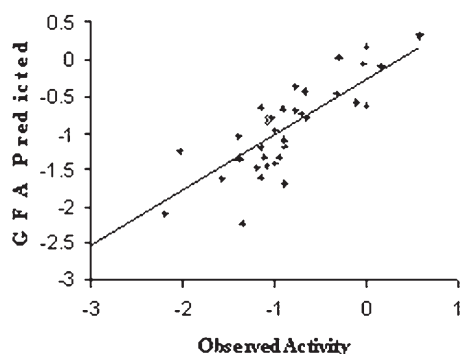


Figure 3. Observed and GFA predicted biological activities of training set (model D).

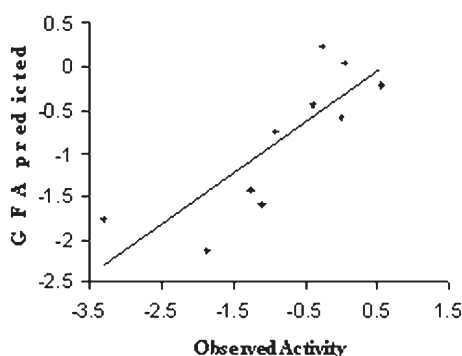


Figure 4. Observed and GFA predicted biological activities of test set (model D).

Table 6. Summary of CoMFA results obtained from different alignments (analysis A)

| Alignment | 1 ^a | 2 ^b | 3 ^c |
|---------------------------------|----------------|----------------|----------------|
| r_{CV}^2 ^d | 0.63 | 0.646 | 0.562 |
| Components | 2 | 3 | 1 |
| SEP ^e | 0.520 | 0.516 | 0.557 |
| r_{nCV}^2 ^f | 0.775 | 0.874 | 0.690 |
| SEE ^g | 0.405 | 0.308 | 0.469 |
| F Value | 60.182 | 78.577 | 79.978 |
| $Pr^2=0$ | 0.00 | 0.00 | 0.00 |
| Contrib. Steric | 0.678 | 0.661 | 0.475 |
| Electrostatic | 0.322 | 0.339 | 0.525 |
| r_{pred}^2 ^h | 0.660 | 0.645 | 0.145 |
| r_{bs}^2 ⁱ | 0.820 | 0.911 | 0.713 |
| Standard deviation ^j | 0.060 | 0.033 | 0.066 |

^a Alignment by RMS fit.

^b Alignment by Multifit.

^c Alignment by Fieldfit.

^d Crossvalidated r^2 .

^e Standard error of prediction.

^f Noncrossvalidated r^2 .

^g Standard error of estimate.

^h Predictive r^2 .

ⁱ From 100 bootstrapping runs.

model based on multifit alignment was considered for further studies.

In addition to steric and electronic parameters, other properties could influence the QSAR. Thus additional parameters like $A \log P$, HOMO, and LUMO in various combinations were used to create different models. These descriptors were calculated using Cerius2 version 3.5 molecular modeling software and were directly used as additional descriptors in the PLS analysis. Analysis B (Table 7) depicts the results obtained from the CoMFA model after the inclusion of additional descriptors.

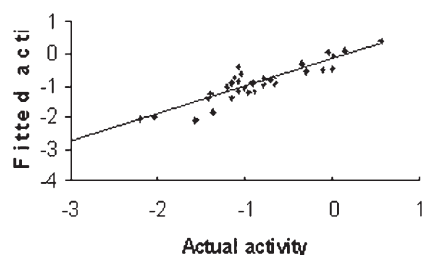
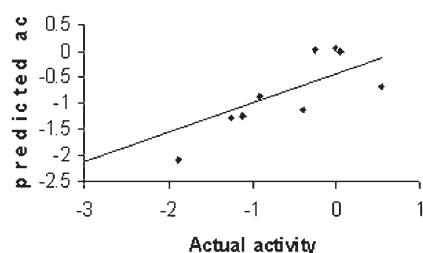
It was found that the CoMFA model based on multifit alignment having steric and electrostatic fields, with LUMO as additional descriptor, showed good internal and external prediction with good statistical significance. This is also similar to the best equation obtained in Cerius2.GA analysis (Table 5, Equation 1). Hence, the CoMFA contours were analyzed using this model. Figure 1 depicts the heavy atoms used for the multifit alignment. Figure 2 shows the alignment of all molecules in training set and test set. The plot of predicted (CoMFA) versus actual activity for training set molecules is shown in Figure 5. The test set residuals of CoMFA analyses are shown in Figure 6. The steric and electrostatic maps of CoMFA model with multifit alignment (Analysis B) are shown in Figures 7 and 8.

Figure 7 presents the steric contour plot. The green contour represents regions of high steric tolerance (80% contribution), while the yellow contour represents regions of unfavorable steric effect (20% contribution). Sterically favorable green contours were observed around the region of ethylene side chain of the most active compound, 49. The sterically unfavorable yellow contours were observed at various regions like *o*-carboxylic acid, oxalyl amide, and the region just below the heterocyclic ring nitrogen substituted side chain. The lipophilic fragment of the molecule is surrounded by sterically favorable green contours. The phenethyl side chain is in the vicinity of sterically favorable green contours. Compounds 01 and 08 have no groups to extend into this sterically favorable green contoured region and hence showed less activity. In compounds 09 and 10, bulky iodo, and phenyl groups extend into this region and show comparatively better activity. Compounds having methylene as side chain (linker) and bulkier aromatic groups as in compounds 42 and 48 showed less activity.

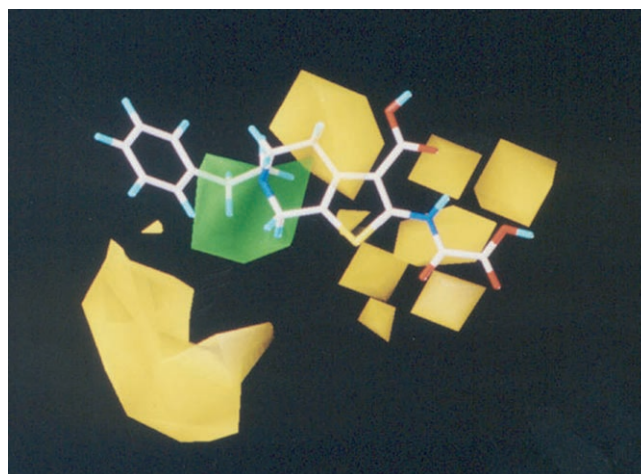
In fused amino-thiophene series, compounds 15–19 showed moderate activity since their substituents enter the sterically favorable region. The polar group containing compounds 20 and 24 showed better activities since the substituted phenyl is in the region of sterically favorable green contours. This may also be due to the additional van der Waals interaction. Substitution of hydrophobic groups like methyl, phenyl, and benzyl in place of the hydroxyl group of compound 20 led to marked decrease in activity because of the bulkier substituents oriented into sterically unfavorable yellow contours. Among the polar substitu-

Table 7. Summary of CoMFA results using multifit alignment and different additional regressors used with standard CoMFA fields (analysis B)

| CoMFA ^a | CoMFA ^a <i>A log P</i> | CoMFA ^a HOMO | CoMFA ^a <i>A log P</i> HOMO | CoMFA ^a LUMO | CoMFA ^a <i>A log P</i> LUMO | |
|----------------------------|-----------------------------------|-------------------------|---|-------------------------|---|--------|
| r_{CV}^2 ^b | 0.646 | 0.631 | 0.547 | 0.444 | 0.671 | 0.661 |
| Components | 3 | 4 | 3 | 4 | 4 | 5 |
| SEP ^c | 0.516 | 0.535 | 0.583 | 0.656 | 0.505 | 0.520 |
| r_{nCV}^2 ^d | 0.874 | 0.879 | 0.791 | 0.802 | 0.900 | 0.903 |
| SEE ^e | 0.308 | 0.306 | 0.396 | 0.391 | 0.278 | 0.278 |
| F Value | 78.577 | 59.764 | 43.007 | 33.462 | 74.548 | 58.878 |
| $Pr^2=0$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Contrib. Steric | 0.661 | 0.639 | 0.677 | 0.670 | 0.605 | 0.601 |
| Electrostatic | 0.339 | 0.346 | 0.275 | 0.274 | 0.295 | 0.296 |
| <i>A log P</i> | – | 0.015 | – | 0.007 | – | 0.004 |
| HOMO | – | – | 0.049 | 0.049 | – | – |
| LUMO | – | – | – | – | 0.100 | 0.099 |
| $r_{pred.}^2$ ^f | 0.645 | 0.618 | 0.675 | 0.671 | 0.643 | 0.647 |
| r_{bs}^2 ^g | 0.911 | 0.830 | 0.870 | 0.862 | 0.936 | 0.939 |
| SD ^h | 0.033 | 0.062 | 0.049 | 0.067 | 0.025 | 0.025 |

^a Alignment by multifit (analysis A).^b Crossvalidated r^2 .^c Standard error of prediction.^d Noncrossvalidated r^2 .^e Standard error of estimate.^f Predictive r^2 .^g From 100 bootstrapping runs.**Figure 5.** A Graph of actual *versus* fitted activities of the training set molecules from multifit alignment (analysis B).**Figure 6.** A graph of actual *versus* fitted activities of the test set molecules from multifit alignment (analysis B).

ents containing compounds 25–29, 3-nitro and 3-amino produced increase in activity because they are away from the sterically unfavorable region. Among methoxy-substituted compounds 27–29, compound 28 showed less potency, due to the *para*-substituted methoxy group entering the

**Figure 7.** CoMFA steric STDEV*COEFF contour plots from multifit alignment (analysis B). Sterically favored areas (contribution level 80%) are represented by green polyhedra. Sterically disfavored areas (contribution level 20%) are represented by yellow polyhedra. The most active compound 49 is displayed as a capped stick model.

sterically unfavorable yellow region. Compounds 27 and 29 showed comparatively more potency than compound 28 since the *meta*-methoxy groups are away from the sterically unfavorable area.

CoMFA analysis of fused amino-thiophene compounds 30–34, showed the same potency as compound 12, since the second hydrophobic ring in both the cases enter the

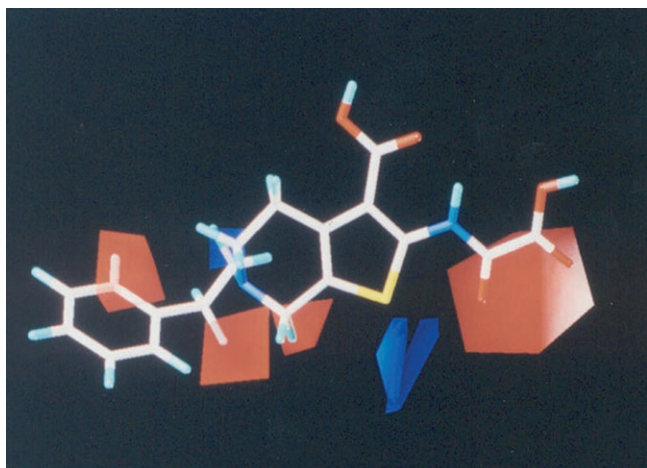


Figure 8. CoMFA electrostatic STDEV*COEFF contour plots from multfit alignment (analysis B). Positive charge favored areas (contribution level 80%) are represented by blue polyhedra. Negative charge favored areas (contribution level 20%) are represented by red polyhedra. The most active compound 49 is displayed as a capped stick model.

sterically favorable green contours. Compounds 35–39 showed better activity than 30–34 because introductions of electronegative atoms make favorable electrostatic interaction in addition to the favorable steric interactions.

Figure 8 depicts the electrostatic contour plot of compound 49. Positive charge favorable blue contours were observed in the vicinity of oxalyl amide nitrogen atom and heterocyclic ring nitrogen atom. This shows that these two nitrogen atoms may be involved in bond formation with positive charge containing residues of PTP1B. Negative charge favorable red contours were observed near the oxygen atoms of oxalyl carbonyl groups. The red contours were also observed beneath the heterocyclic nitrogen atom and substituted phenyl group of compound 49. The negative charge favorable red contours in the sterically favorable green region suggest that bulky substituents with high electron density are desirable for activity.

4 Conclusions

3-D-QSAR analyses by Cerius2.GA and CoMFA were performed with the aim of deriving structural requirements for better PTP1B inhibitory activity. Analysis of this dataset resulted in CoMFA models with good conventional and crossvalidated correlation coefficient by multfit method. CoMFA and Cerius2.GA show that both steric and electrostatic fields contribute significantly towards PTP1B inhibitory activity, while LUMO also has some contribution. Derived Cerius2.GA and CoMFA models were able to predict the activity of the test set molecules with good predictive correlation coefficient. The structural

requirement information derived from this QSAR study could be useful in further designing of potent PTP1B inhibitors.

Acknowledgement

Subbiah Ramar, Seema Bag, and Nilesh R. Tawari are thankful to University Grant Commission (UGC), India for financial support.

References

- [1] R. A. DeFronzo, R. C. Bonadonna, E. Ferrannini, *Diabetes Care* **1992**, *15*, 318–368.
- [2] P. G. Drake, B. I. Posner, *Mol. Cell. Biochem.* **1998**, *182*, 79–89.
- [3] J. L. Evans, B. Jallal, *Exp. Opin. Invest. Drugs* **1999**, *8*, 139–160.
- [4] B. G. Szczepankiewicz, G. Liu, P. J. Hajduk, C. Abad-Zapatero, Z. Pei, Z. Xin, T. H. Lubben, J. M. Trevillyan, M. A. Stashko, S. J. Ballaron, H. Liang, F. Huang, C. W. Hutchins, S. W. Fesik, M. R. Jirousek, *J. Am. Chem. Soc.* **2003**, *125*, 4087–4096.
- [5] M. Zhou, M. Ji, *Bio. Med. Chem. Lett.* **2005**, *15*(24), 5521–5525.
- [6] M. Whitlow, A. J. Howard, D. Stewart, K. D. Hardman, J. H. Chan, D. P. Baccanari, R. L. Tansik, J. S. Hong, L. F. Kuyper, *J. Med. Chem.* **2001**, *44*, 2928–2932.
- [7] A. A. Soderholm, P. T. Lehtovuori, T. H. Nyronen, *J. Med. Chem.* **2006**, *49*, 4261–4268.
- [8] C. Sheng, W. Zhang, H. Ji, M. Zhang, Y. Song, H. Xu, J. Zhu, Z. Miao, Q. Jiang, J. Yao, Y. Zhou, J. Zhu, J. Lu, *J. Med. Chem.* **2006**, *49*, 2512–2525.
- [9] B. A. Bhongade, V. V. Gouripur, A. K. Gadad, *Bio. Med. Chem.* **2005**, *13*, 2773–2782.
- [10] A. Ganjee, X. Lin, *J. Med. Chem.* **2005**, *48*, 1448–1469.
- [11] R. D. Cramer III, D. E. Patterson, J. D. Bunce, *J. Am. Chem. Soc.* **1998**, *110*, 5959–5967.
- [12] H. S. Andersen, O. H. Olsen, L. F. Iversen, A. L. Sorensen, S. B. Mortensen, M. S. Christensen, S. Branner, T. K. Hansen, J. F. Lau, L. Jeppesen, E. J. Moran, J. Su, F. Bakir, L. Judge, M. Shahbaz, T. Collins, T. Vo, M. J. Newman, W. C. Ripka, N. P. Moller, *J. Med. Chem.* **2002**, *45*, 4443–4459.
- [13] A. Tropsha, P. Gramatica, V. K. Gombar, *QSAR Comb. Sci.* **2003**, *2*(1), 69–77.
- [14] Cerius2 version 3.5 is available from Molecular Simulations Inc., 9685, Scranton Road, San Diego, CA, USA 92121.
- [15] J. Hopfinger, *J. Am. Chem. Soc.* **1980**, *102*, 7196–7206.
- [16] SYBYL Molecular Modeling System, Version 6.6; Tripos, Inc., St. Louis, MO, USA 63144-2913.
- [17] M. Clark, R. D. Cramer III, N. Van Opdenbosh, *J. Comput. Chem.* **1989**, *10*, 982–1012.
- [18] MOPAC 6.0; Quantum Chemistry Program Exchange; Indiana University.
- [19] L. F. Iversen, H. S. Andersen, S. Branner, S. B. Mortensen, G. H. Peters, K. Norris, O. H. Olsen, C. B. Jeppesen, B. F. Lundt, W. Ripka, K. B. Moller, N. P. Moller, *J. Biol. Chem.* **2000**, *275*, 10300–10307.