Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 13 (2005) 2331–2338
Bioorganic & Medicinal Chemistry

Comparative molecular similarity indices analysis (CoMSIA) studies of 1,2-naphthoquinone derivatives as PTP1B inhibitors

M. Elizabeth Sobhia* and Prasad V. Bharatam

Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research (NIPER), Sector-67, S.A.S. Nagar (Mohali) 160 062, India

Received 1 August 2004; accepted 17 December 2004
Available online 18 January 2005

Abstract—Protein tyrosine phosphatase-1B (PTP1B) has been demonstrated to play a key role in the negative signalling pathway of insulin. Potent and orally active PTP1B inhibitors are considered to be promising pharmacological agents for the treatment of type-2 diabetes and resistance to weight gain. CoMSIA studies have been preformed on 1,2-naphthoquinone derivatives that are reported to be potential non-peptidic inhibitors of PTP1B. For the selection of dataset to develop the model, the reported molecules were subjected to property filters and segregated into training and test set. As the crystal structure of PTP1B–naphthoquinone derivative is not known, the most active molecule was subjected to simulated annealing dynamics method and the lowest energy conformer was reminimised and considered as the bioactive conformation. Database-inertial alignment was followed for aligning the molecules. Different CoMSIA models were built to get the best related field.

1. Introduction

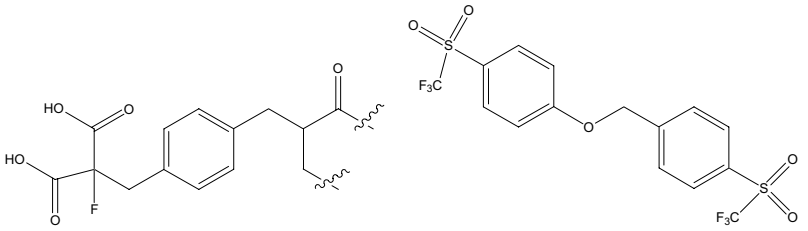
Protein tyrosine phosphatases (PTP-ases) belong to a diverse family of enzymes, which are shown to play an important role in the insulin-signalling pathway. The interaction of insulin with its receptor leads to autophosphorylation of certain tyrosine residues within the receptor, thus activating the receptor kinase. PTPases dephosphorylate the activated insulin receptor and attenuate the tyrosine kinase activity. This decreases the tyrosine phosphorylation activity in insulin responsive tissues like muscle, liver and adipose and causes a reduction in the metabolic action of insulin and therefore, hyperglycaemia. The decrease in tyrosine phosphorylation activity in cells appears to be caused by the increase in protein tyrosine phosphatase activity.[1–3] Among PTPases, PTP1B has been demonstrated to dephosphorylate insulin receptor and act as a negative regulator of insulin signalling. A recent study on PTP1B knockout mice showed that the loss of PTP1B activity enhanced sensitivity towards insulin and resistance to obesity, indicating that a potent, orally active PTP1B inhibitor could be a potential agent for the treatment of Type 2 diabetes and obesity.[4]

Three-dimensional structures of PTP1B-inhibitor complexes, determined by X-ray crystallography have unravelled distinguished features about the different binding sites and inhibitors. The primary binding site, that is, the active site of PTP1B is conserved in most of the PTPases, consisting of eight amino acid residues (His-Cys-Ser-Ala-Gly-Ile-Gly-Arg) that bind to phospho tyrosine moiety and consisting of the catalytic residue Cys (Cys215).[5] Several peptidic systems like D-A-D-E-OMT-L (**1**, OMT = L-O-(2-malonyl)tyrosine) and non-peptidic (**2**) inhibitors, directed to active site have been developed.[6,7] Puius et al. discovered a second binding site adjacent to the active site and inhibitors binding to both active site and the second binding site have been developed (**3**).[8] Inhibitors specific to the residues Arg47 and Asp48 for obtaining selective inhibitors over other PTPases are also found in the literature.[9] Efforts are also being taken to develop selective inhibitors as potential agents over the phosphatase TCPTP (**4**, **5**).[10] The crystal structure of the oxidised PTP1B revealed a sulfonyl amide binding group in the active site of the protein, which is being exploited for designing novel PTP1B inhibitors.[11] Among the PTP1B inhibitors known so far, ertiprotafib is the only drug candidate, developed by American Home Products, has reached the stage of phase 2 clinical trials. However, this molecule was withdrawn from phase 2 clinical trials due to its

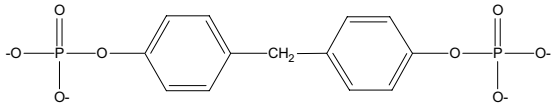
* Corresponding author. E-mail addresses: mesophia@niper.ac.in; pvbharatam@niper.ac.in

0968-0896/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.12.039

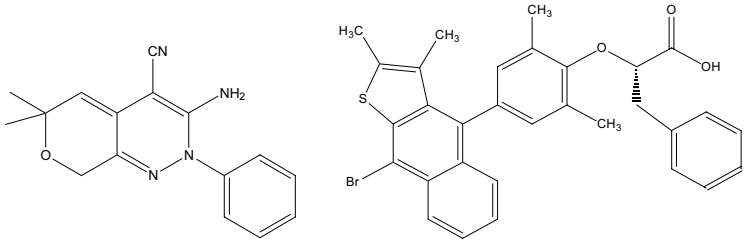
undesirable side effects. Very recently, a PTP1B antisense oligonucleotide ISIS-113715, developed by ISIS Pharmaceuticals Inc. has reached phase 1 clinical trials.[12] As there is no successful molecule available in the market for this potential target, there is an urgent need for developing therapeutically useful PTP1B inhibitors. A large number of patents and leading research articles have appeared in the recent past.[13–15]



1 2



3



4 5

As highlighted above there is a wealth of information available in the literature for understanding the different binding sites and the mode of interaction of inhibitors. However, the crystal structure of PTP1B–naphthoquinone complex is not known. In the absence of crystallographic data, the application of 3D-QSAR methodologies such as CoMFA and CoMSIA are used as valuable tools for designing new molecules.[16–19] CoMSIA is a recently developed approach to perform 3D-QSAR analysis based on molecular similarity indices[20] of various properties such as steric, electrostatic, hydrophobic, hydrogen bonding. The resulting contribution maps are interpreted in a simpler way than CoMFA. Kulkarni et al. have carried out a 3D-QSAR study on PTP1B inhibitors based benzofuran/benzothiophene biphenyls derivatives using CoMFA and CoMSIA methodologies.[21] In the current paper, we present the results of the CoMSIA study carried out on 1,2 naphthoquinones, reported as non-peptidic PTP1B inhibitors. The aim of the study is to develop a robust 3D-QSAR model and to get a correlation between the structural features of this class of compounds and their biological activity. As research in this area, specifically in the development of non-peptidic PTP1B inhibitors is novice, results of the study may provide some useful hints to design more promising chemical scaffolds.

2. Computational details

2.1. Biological data and selection of dataset

Biological data of the molecules reported by Ahn et al.[22] has been used in this study. The compounds were evaluated for their in vitro inhibitory activity against recombinant human PTP1B using fluorescin diphosphate (FDP) as the substrate. Enzyme activity was assayed by measuring the fluorescence of the product, fluorescin monophosphate at 485 nm.[22] Total number of 54 compounds reported were subjected to property filters viz. molecular weight < 500, logP −5 to +5, number of rotatable bonds < 8, number of hydrogen bond donor < 5 and acceptor atoms < 10. The molecules for which the exact biological activity was not given in numerical form but mentioned as 'na - not active' were excluded from the present study. The molecules reported with sodium salt were also not considered for analysis. The 34 compounds fulfilled the above criteria were chosen for the CoMSIA study and were randomly segregated into training and test sets comprising 25 and 9 molecules each, respectively. The reported IC_{50} values converted into pIC_{50} values according to the formula

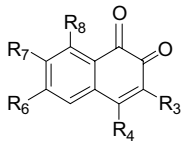
$$pIC_{50} = -\log IC_{50}$$

The dataset chosen for study along with their physicochemical parameters is shown in Table 1.

2.2. Molecular modelling

Molecular modelling studies were carried out using SYBYL 6.8,[23] implemented on an SGI Octane 2.

Table 1. The dataset chosen for CoMSIA study and their physicochemical parameters and pIC_{50} values



S. No.	Compd.[a]	R_3	R_4	R_6	R_7	R_8	MW	logP	NRB	HBD	HBA	pIC_{50}
1	**1**	H	NH_2	H	H	H	173	0.39	1	1	2	−1.39
2	**2**	H	$N(CH_3)C_6H_5$	H	H	H	263	2.86	3	0	2	−1.54
3	**4**	H	OCH_3	H	H	H	188	1.11	2	0	3	−1.46
4	**5**	H	$OCH_2CH_2CH_2OH$	H	H	H	232	0.72	5	1	4	−1.56
5	**7**	H	C_6H_5	H	H	H	234	3.47	1	0	2	0.06
6	**8**	H	C_6H_5-2,5-Cl_2	H	H	H	303	4.51	1	0	2	−0.70
7	**9**	H	$C_6H_5OCH_3$	H	H	H	264	3.22	3	0	3	−0.72
8	**10**	H	C_6H_3-2,5-F_2	H	H	H	270	3.75	1	0	4	0.30
9	**11**	H	$C_6H_4COOCH_3$	H	H	H	292	3.20	4	0	4	−0.19
10	**12**	H	C_6H_4-2-OCH_2CO_2Et	H	H	H	336	3.14	7	0	5	−0.33
11	**13**	H	C_6H_4-4-OCH_2CO_2Et	H	H	H	336	3.14	7	0	5	−0.03
12	**14**	H	C_6H_4-4-OH	H	H	H	250	3.19	2	1	3	0.36
13	**15**	H	C_6H_4-2-OH	H	H	H	250	3.19	2	1	3	−0.20
14	**17**	H	C_6H_4-2-NO_2	H	H	H	279	−0.44	2	0	4	−0.07
15	**18**	H	1-Naphthyl	H	H	H	284	4.48	1	0	2	−0.33
16	**19**	H	3-Indole	H	H	H	273	2.38	1	1	2	0.05
17	**20**	H	3-Indole-5-carboxylic acid	H	H	H	317	2.08	3	2	4	−0.48
18	**21**	H	3-Indole-6-carboxylic acid	H	H	H	317	2.08	3	2	4	−0.66
19	**24**	H	Cyclohexyl	H	H	H	240	3.67	1	0	2	0.49
20	**25**	H	Benzyl	H	H	H	248	3.87	2	0	2	−0.15
21	**26**	H	$(CH_2)_5CH_3$	H	H	H	242	4.24	6	0	2	−0.52
22	**27**	H	Cyclopentyl	H	H	H	226	3.27	1	0	2	−0.62
23	**29**	H	Isopropyl	H	H	H	200	2.98	3	0	2	0.09
24	**36d**	$O(CH_2)_4CH_3$	C_6H_5	H	OMe	H	264	3.22	3	0	3	−0.31
25	**36f**	H	Indole	H	OCH(Bn)-CO_2H	H	437	3.89	7	2	5	−0.31
26	**39a**	H	C_6H_5	H	$NHCO_2Bn$	H	383	4.72	6	1	4	−0.35
27	**39b**	H	C_6H_5	H	H	$NHCO_2Bn$	383	4.72	6	1	4	−0.13
28	**39c**	H	C_6H_5	H	H	$NHCO_2Et$	280	3.29	6	1	4	0.19
29	**42a**	H	C_6H_5	$(CH_2)_2CO_2Me$	H	H	320	3.53	6	0	4	0.48
30	**42b**	H	Indole	$(CH_2)_2CO_2Me$	H	H	359	2.44	6	0	4	0.04
31	**48a**	$(CH_2)_2CO_2Et$	C_6H_5	H	H	H	334	3.69	7	0	4	0.37
32	**48b**	$(CH_2)_2CO_2Et$	Indole	H	H	H	373	2.59	7	1	4	0.57
33	**55c**	CH_2Ph-4-O-CH_2CO_2H	C_6H_5	$(CH_2)_2CO_2Me$	H	H	398	4.65	7	1	5	−0.19
34	**55d**	CH_2Ph-4-O-CH_2CO_2H	Indole	H	H	H	437	3.56	7	2	5	−0.004

[a] The compound numbers are taken from the original paper reporting the SAR, Ref. 22. MW: molecular weight; NRB: number of rotatable bonds; HBD: number of hydrogen bond donor centres; HBA: number of hydrogen bond acceptor centres.

Molecular weight, number of rotatable bonds and the number of hydrogen bond donors and acceptors of the molecules were calculated by the options available in Sybyl. The logP values were calculated using the software Hyperchem 6.0.[24] The most active molecule, (**48b**) of the series was subjected to simulated annealing. For simulated annealing, the molecule was heated up to 2000 K for 1000 fs and cooled down to 200 K in 1000 fs. The lowest energy conformer was selected and was reminimised using Tripos force field parameters, Powell[25] method. All the other molecules were built from the template molecule. A gradient change of 0.05 kcal/molÅ and minimum energy difference of 0.001 kcal/mol was set as convergence criteria. Partial atomic charges were calculated using Gasteiger–Hückel charges.[26]

2.3. CoMSIA interaction energy

The reported CoMSIA method is based on molecular similarity indices;[20] this method overcomes some of the drawbacks arising from the functional form of the Lennard–Jones and Coulomb potentials used in CoMFA. Molecular similarity is expressed in terms of five different properties viz., steric, electrostatic, hydrophobic, H-bond donors and acceptors which were calculated[20] using a C^+ probe atom with a radius of 1 Å placed at a regular grid spacing of 2 Å. CoMSIA similarity indices (A_F) for molecule j with atoms i at a grid point q are calculated by the equation

$$A^q_{F,k}(j) = -\sum w_{probe,k} w_{ik}\, e^{-\alpha r} iq^2$$

where k represents the following physicochemical properties: steric electrostatic, hydrophobic, H-bond donor and H-bond acceptor. A Gaussian type distance dependence was used between grid point q and each atom i of the molecule. The default value of 0.3 was used as the attenuation factor (α). The steric indices are related to the third power of the atomic radii, electrostatic descriptors are derived from atomic partial charges, hydrophobic fields are derived from atom-based parameters[27] and H-bond donor and acceptor indices are obtained by a rule based method based on experimental results.[28]

2.4. Alignment and the influence of orientation of the aligned molecules

Database-inertial alignment was used for aligning the molecules. In this method of alignment, all the molecules in the database are aligned with the template in the database and a common structure is provided to evaluate the best fit. In the present case, molecule **48b** was used as the template and naphthoquinone unit was used as the common structure. For studying the influence of the orientation of the aligned molecules,[29] the aligned molecules (Fig. 1a) were systematically rotated globally for 180° in three directions X (Fig. 1b), Y (Fig. 1c) and Z (Fig. 1d) and the PLS analysis was carried out and the effect was noted.

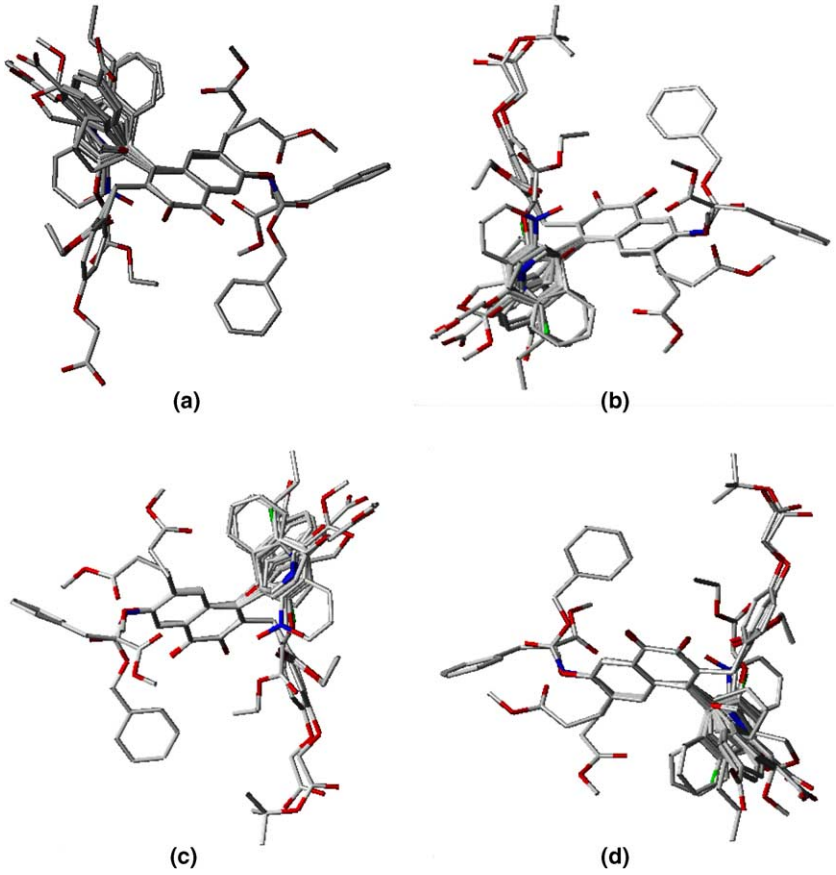


Figure 1. Alignment of the molecules. (a) The alignment used for the development of final model. (b) $X = 180°$, (c) $Y = 180°$ and (d) $Z = 180°$ rotation.

2.5. Partial least square (PLS) analysis

PLS method[30] was used to linearly correlate the CoMSIA fields to biological activity values. The cross-validation analysis was performed using 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 cross-validated r^2_{cv} that resulted in optimum number of components and lowest standard error of prediction were taken. To speed up the analysis and reduce noise, a minimum column filter value (σ) of 2.00 kcal/mol was used for the cross-validation. Final analysis (non-cross-validation) was performed to calculate conventional r^2 (r^2_{ncv}) using the optimum number of components.

2.6. Predictive correlation coefficient (r^2_{pred})

The predictive ability of the 3D-QSAR model was determined from a set of nine compounds that were not included in the training set. These molecules were aligned using the same method of alignment as described above, and their activities were predicted using the model generated by the training set. The predictive correlation (r^2_{pred}), based on the test set molecules, is defined as

$$r^2_{pred} = (SD - PRESS)/SD$$

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 test set.

3. Results and discussion

CoMSIA models were obtained with 25 molecules in the training set and 9 molecules in the test set. Combinations of different CoMSIA fields were used to get the best model (Table 2). The model CoMSIA-13, obtained with all the fields including steric, hydrophobic electrostatic, H-bond donor and acceptor was found to be the best one. The choice of model CoMSIA-13 over other models was based on all parameters including r^2, r^2_{cv}, number of components, standard error of estimate etc. The best model resulted in a cross-validated r^2 of 0.454 with minimum standard error and optimum numbers of components as 0.06 and 5, respectively (also the contour analysis was taken into account). On purely statistical count, it appears that CoMSIA-3, CoMSIA-7 are also equivalent to CoMSIA-13. However, our choice is CoMSIA-13 owing to the fact that it involves all types of fields and SEE is quite small. The summary of PLS statistics for the best model is shown in Table 3. The various orientations of the aligned molecules did not drastically alter the predictivity of the model (Table 4) though there is a slight variation in the value of r^2_{cv} for of *X*-180 and *Z*-180 rotations. A plot of predicted versus actual activity for training set molecules is shown in Figure 2. Figure 3 represents the histogram of the residuals for the test molecules. The actual, predicted and residual values of training and test set are given in Tables 5 and 6, respectively.

The training (Table 5) and test set (Table 6) were examined for outliers. The molecules having residuals above 1

Table 3. Statistical parameters for the optimised model CoMSIA-13

S. No.	Statistical parameters	Value
1	Number of molecules in the training set	25
2	Number of molecules in the test set	9
3	r^2_{cv}	0.454
4	Optimum number of components	5
5	r^2_{ncv}	0.988
6	SEE	0.066
7	*F*-value	245.9
8	Column filtering	2
Field contributions		
1	Steric	0.112
2	Electrostatic	0.279
3	Hydrophobic	0.297
4	Hydrogen bond donor	0.089
5	Hydrogen bond acceptor	0.223

Table 2. Statistical parameters obtained for different CoMSIA field contributions

S. No.	Title	Fields						PLS analysis		
		S	E	H	D	A	r^2_{cv}	*N*	r^2_{ncv}	SEE
1	CoMSIA-1	√					0.480	3	0.94	0.15
2	CoMSIA-2		√				0.119	5	0.97	0.09
3	CoMSIA-3			√			0.483	3	0.97	0.09
4	CoMSIA-4				√		−0.389	3	0.48	0.43
5	CoMSIA-5					√	−0.175	1	0.64	0.36
6	CoMSIA-6	√	√				0.364	3	0.98	0.09
7	CoMSIA-7	√		√			0.529	3	0.97	0.10
8	CoMSIA-8	√			√		0.126	6	0.87	0.21
9	CoMSIA-9	√				√	0.169	3	0.93	0.16
10	CoMSIA-10	√	√	√			0.469	3	0.98	0.88
11	CoMSIA-11	√		√	√		0.369	6	0.96	0.11
12	CoMSIA-12	√		√	√	√	0.508	6	0.96	0.11
13	CoMSIA-13	√	√	√	√	√	0.454	5	0.99	0.06

r^2_{cv} = cross-validated correlation coefficient; *N* = no. of components; r^2 = conventional correlation coefficient; SEE = standard error of estimate; S = steric field, E = electrostatic field, H = hydrophobic field, D = hydrogen bond donor field, A = hydrogen bond acceptor field.

Table 4. Statistical parameters obtained for different orientations

Different orientations	r^2_{cv}	r^2	Optimum no. of components	SEE	F
X-rotation	0.491	0.988	5	0.067	237.5
Y-rotation	0.453	0.987	5	0.068	231.0
Z-rotation	0.465	0.988	5	0.065	206.3

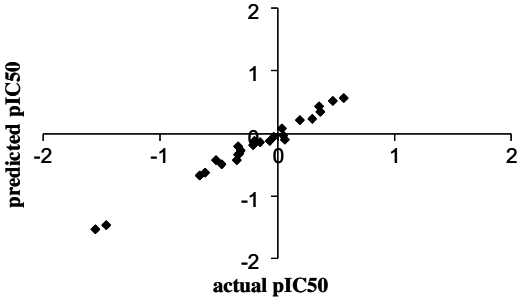


Figure 2. Plot of actual versus predicted pIC_{50} of the training set molecules.

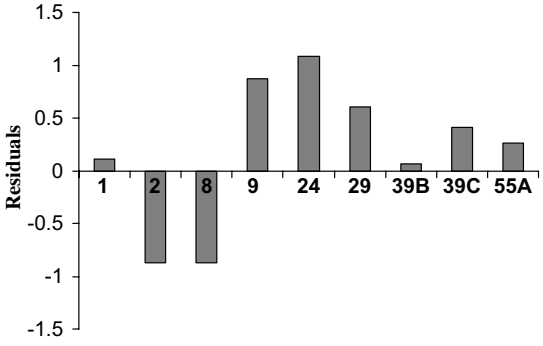


Figure 3. Histogram of residuals of the test set molecules.

Table 5. Actual and predicted activity of the training set molecules

S. No.	Compd.	Actual pIC_{50}	Predicted pIC_{50}	Residual
1	**10**	0.3	0.24	0.06
2	**11**	−0.19	−0.12	−0.07
3	**12**	−0.33	−0.34	0.01
4	**13**	−0.03	−0.06	0.03
5	**14**	0.36	0.43	−0.07
6	**15**	−0.2	−0.19	−0.01
7	**17**	−0.07	−0.12	0.05
8	**18**	−0.33	−0.20	−0.13
9	**19**	0.05	−0.03	0.08
10	**20**	−0.48	−0.49	0.01
11	**21**	−0.66	−0.68	0.02
12	**25**	−0.15	−0.14	−0.01
13	**26**	−0.52	−0.43	−0.09
14	**27**	−0.62	−0.62	0.00
15	**36d**	−0.31	−0.27	−0.04
16	**36f**	−0.32	−0.32	−0.00
17	**39a**	−0.35	−0.43	0.08
18	**4**	−1.46	−1.48	0.02
19	**42a**	0.48	0.51	−0.03
20	**42b**	0.04	0.07	−0.03
21	**48a**	0.37	0.35	0.02
22	**48b**	0.57	0.57	0.00
23	**5**	−1.56	−1.53	−0.03
24	**55d**	0.19	0.22	−0.03
25	**7**	0.07	−0.09	0.16

Table 6. Actual and predicted activity of the test set molecules

S. No.	Compd.	Actual pIC_{50}	Predicted pIC_{50}	Residual
1	**1**	−1.39	−1.50	0.11
2	**2**	−1.54	−0.67	−0.87
3	**8**	−0.70	0.17	−0.87
4	**9**	0.72	−0.15	0.87
5	**24**	0.49	−0.59	1.08
6	**29**	0.10	−0.51	0.61
7	**39b**	−0.13	−0.20	0.07
8	**39c**	0.19	−0.22	0.41
9	**55c**	−0.19	−0.45	0.26

logarithm unit were considered as outliers. According to this, the molecule **24** in the test set is an outlier having high residual value, under predicted by the CoMSIA model. There are several reasons that may account for the outliers, including an error in the experimental measurements, a unique structural difference, a different binding conformation and a significant difference in the physicochemical properties. Molecule **24** is the only molecule in the test set having cyclohexyl group at R_4 position. This molecule does not have many counter partners in the training set except molecule **27** and this might be a reason why CoMSIA is not able to predict the activity of this molecule accurately. Other reason for under prediction of this compound might be due to the different geometry adopted by the cyclohexyl ring compared to phenyl, naphthalene and indole rings present in many other molecules.

3.1. Contour interpretation

The contour maps obtained by CoMSIA show how 3D-QSAR methods can be used to identify features important for interaction between ligands and protein. The contour maps also give an indication of those portions of the molecule that require particular property to improve the binding affinity. A few of the interesting features are discussed below. The CoMSIA steric interactions (Fig. 4a) are represented by favoured green and disfavoured yellow contours while electrostatic interactions (Fig. 4b) are represented by negative charge favoured red and positive charge favoured blue contours. The hydrophobic interactions (Fig. 4c) are represented by favoured yellow and disfavoured white regions. Similarly, the fields representing hydrogen bond donors are shown by favoured cyan and disfavoured purple contours while the hydrogen bond acceptor fields are shown by favoured magenta and disfavoured red contours (Fig. 4d). The most active molecule **48b** is displayed in the background in Figure 4.

The steric field of CoMSIA shows green areas within which the occupation of bulky groups is likely to increase the activity and yellow areas within which occupation of bulky groups is likely to decrease the activity. There are two green contours present in close

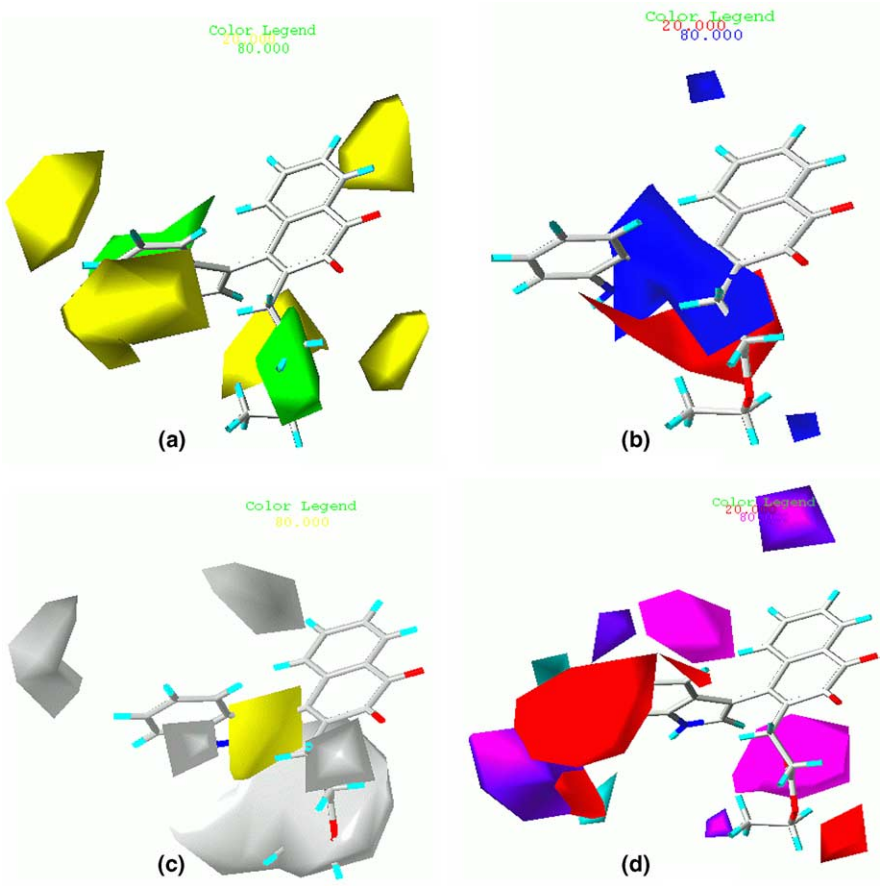


Figure 4. CoMSIA contour maps. (a) Sterically favoured areas (contribution level of 80%) in green, disfavoured areas (contribution level of 20%) in yellow. (b) Positive-charge favoured areas (contribution level of 80%) in blue, negative-charge-favoured areas (contribution level of 20%) in red. (c) Hydrophobic favoured areas (contribution level of 80%) in yellow, disfavoured (contribution level of 20%) in white. (d) Hydrogen-bond-donor favoured areas (contribution level of 80%) in cyan, disfavoured areas (contribution level of 20%) in purple. Hydrogen-bond-acceptor favoured areas (contribution level of 80%) in magenta, disfavoured areas (contribution level of 20%) in red.

proximity to R_3 and R_4 positions of naphthoquinone ring. It is noted, for many of the active compounds like **48b**, **42a**, **48a**, **14**, **10**, **19**, **55d** and **7**, there is no encroachment in the disfavoured yellow contour rather there is significant amount of bulk in the favoured green contours. On the other hand, the bulky groups of the less active molecules **5**, **4**, **21**, **27** overlap in the disfavoured yellow region. Likewise, many of the moderate active molecules like **19**, **42b**, **17**, **11**, **36f**, **26** present a portion of their bulky groups in the yellow regions.

The blue and red contours of electrostatic fields reveal that the increased positive charge will enhance the activity in the blue regions and increased negative charge will increase the activity in the red regions. There is a red contour, present near the position R_3 and the compounds of higher activity (**42a**, **48a**, **14**, **10**, **55d**) have their electronegative groups in this region and the least active molecules like **27**, **21**, **4 and 5** have their negative groups in the blue region. Figure 4c denotes the contours representing the hydrophobic fields. There is a yellow contour seen in the vicinity of the five-membered ring of indole of the most active molecule **48b**, which is connected to the naphthoquinone ring. This might be due to the presence of the hydrophobic rings present at position R_4 in many molecules. Fragments of white contours are seen spreading around the molecule (**48b**) in which a big white contour is seen near R_3. This indicates the presence of hydrophilic groups in these areas may increase the activity. This also gives an indication that probably the groups, represented by these regions are exposed to the solvent molecules. For example, one of the most active molecules **42a**, has its phenyl ring in the yellow region and $COOCH_3$ group in the hydrophilic region and one of the least active molecules **5**, has the methoxy group in the hydrophobic favoured region.

Literature studies on the PTB1B inhibitors indicate that the hydrogen bond interactions between the ligands and the amino acids, especially Asp48, Arg24, Arg254, etc. To understand importance of such interactions in the naphthoquinone series, we have performed the correlation of the hydrogen bond donor and acceptor fields (Fig. 4d). The two small cyan contours in Figure 4d show that the substitution of R_4 with hydrogen bond donating groups with a suitable linker may increase the activity. Similarly, there is a magenta contour seen in the vicinity of indole ring (R_4 of **48b**), which prefers hydrogen bonding acceptor atoms. The big magenta contour present near the flexible part at R_3 of naphthoquinone ring indicate that H-bond acceptor units R3 are favourable.

Among the molecules considered for CoMSIA there is a subset of molecules, which contains substituted phenyl groups at R_4 position (Compounds **7**–**17**, Table 1). It is interesting to analyse the contours and compare the actual and predicted values of this set of molecules and find the real correlations. For example, the experimental pIC_{50} value of the molecule **7** having phenyl at R_4 is 0.06. Replacement of phenyl with 2,5-dichloro phenyl (molecule **8**) decreases the activity to −0.70. At the same time, replacement of phenyl with a 2,5-difluoro phenyl increases the activity to 0.3. Also, between the two extremes of activity lie the other molecules with different substituents. The model explains well the real correlations among these molecules. Considering the subset, there are three molecules in the training set (**10**, **15** and **17**) having F, OH and NO_2 at the second position of phenyl and one molecule (**8**) in the test set having Cl. For example, a big red contour, favouring electro negative potential for the rise in activity is present near the *ortho*-substituted F atom in **10**, explains the observed changes. A careful examination of the predicted activity and residuals reveal that the activity of the molecules increases as the electronegative character of the groups increases. To further validate the correlation, we have predicted the activities of the compounds with NH_2, CCl_3, CF_3, etc. substituents at the 2-position of the phenyl ring at R_4. The predicted activities are in accordance with the expected trends in electro negativity of the substituents. In the case of Cl substituent also (molecule **8**) desired trend is observed, the residual value is slightly higher.

4. Conclusions

The drug discovery research in PTP1B is a challenging area to work with and many industrialists and academicians are focussing their research towards the development of potential PTP1B inhibitors. The most potential inhibitors of PTP1B known to date are peptidic in nature, which are not cell permeable. Hence the major efforts are aimed at developing small molecular non-peptidic inhibitors. In this context, the model developed on non-peptidic inhibitors serves as a guide to design more potent, small molecule non-peptidic inhibitors within this type of derivatives or a new series altogether. Since property filters have been used as a criterion for dataset selection to develop the model, it can be expected that the predicted new molecules also will have the desirable features. The analyses by positioning the molecules within the CoMSIA contours lead to interesting conclusions about the different functional groups on biological activity. With the crystal structure complex of PTP1B–napthoquinone is not known, the hydrogen bond and acceptor field of this CoMSIA study throws some light on the probable sites on the enzyme for explaining the interaction between this class of compounds and the protein.

References and notes

1. Kenner, K. A.; Anyanwu, E.; Olefsky, J. M.; Kusari, J. *J. Biol. Chem.* **1996**, *271*, 19810.
2. Goldstein, B. J.; Ahmad, F.; Ding, W.; Li, P.-M.; Zhang, W.-R. *Mol. Cell. Biochem.* **1998**, *182*, 91.
3. Kennedy, B. P.; Ramachandran, C. *Biochem. Pharmacol.* **2000**, *60*, 877.
4. Elchebly, M.; Payette, P.; Michaliszyn, E.; Cromlish, W.; Collins, S.; Loy, A. L.; Normandin, D.; Cheng, A.; Hagen, J. H.; Chan, C.-C.; Ramachandran, C.; Gresser, M. J.; Tremblay, M. L.; Kennedy, B. P. *Science* **1999**, *283*, 1544.
5. Barford, D.; Flint, A. J.; Tonks, N. K. *Science* **1994**, *263*, 1397.
6. Groves, M. R.; Yao, Z.-J.; Roller, P. P.; Burke, T. R., Jr.; Barford, D. *Biochemistry* **1998**, *37*, 17773.
7. Huang, P.; Ramphal, J.; Wei, J.; Liang, C.; Jallal, B.; McMahon, G.; Tang, C. *Bioorg. Med. Chem.* **2003**, *11*, 1835.
8. Puius, Y. A.; Zhao, Y.; Sullivan, M.; Lwarence, D. S.; Almo, S. C.; Zhang, Z. Y. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 13420.
9. Asante-Appiah, E.; Patel, S.; Dufresne, C.; Roy, P.; Wang, Q.; Patel, V.; Friesen, R.; Ramachandran, C.; Becker, J. W.; Leblanc, Y.; Kennedy, B. P.; Scapin, G. *Biochemistry* **2002**, *41*, 9043.
10. Iversen, L. F.; Moller, K. B.; Pedersen, A. K.; Peters, G. H.; Petersen, A. S.; Andersen, H. S.; Branner, S.; Mortensen, S. B.; Moller, N. P. H. *J. Biol. Chem.* **2002**, *277*, 19982.
11. Salmeen, A.; Andersen, J. N.; Myers, M. P.; Meng, T. C.; Hinks, J. A.; Tonks, N. K.; Barford, D. *Nature* **2003**, *423*, 769.
12. Elizabeth, A. H.; Nigel, L. *Cur. Opin. Invest. Drugs* **2003**, *4*, 1179.
13. Blaskovich, M. A.; Kim, H.-O. *Expert Opin. Ther. Patent* **2002**, *12*, 871.
14. Taylor, S. D. *Curr. Top. Med. Chem.* **2003**, *3*, 759.
15. Liu, G. *Curr. Med. Chem.* **2003**, *10*, 1241.
16. Desiraju, G. R.; Gopalakrishnan, B.; Jetti, R. K.; Nagaraju, A.; Raveendra, D.; Sarma, J. A.; Sobhia, M. E.; Thilagavathi, R. *J. Med. Chem.* **2002**, *45*, 4847.
17. Chakraborti, A. K.; Gopalakrishnan, B.; Sobhia, M. E.; Malde, A. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1403.
18. Aboye, T. L.; Sobhia, M. E.; Bharatam, P. V. *Bioorg. Med. Chem.* **2004**, *12*, 2709.
19. Chakraborti, A. K.; Gopalakrishnan, B.; Sobhia, M. E.; Malde, A. *Eur. J. Med. Chem.* **2003**, *38*, 975.
20. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.
21. Murthy, V. S.; Kulkarni, V. M. *Bioorg. Med. Chem.* **2002**, *10*, 897.
22. Ahn, J. H.; Cho, S. Y.; Ha, J. D.; Chu, S. Y.; Jung, S. H.; Jung, Y. S.; Baek, J. Y.; Choi, I. K.; Shin, E. Y.; Kang, S. K.; Kim, S. S.; Cheon, H. G.; Yang, S.-D.; Choi, J.-K. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1941.
23. SYBYL6.8, Tripos Associates Inc.: 1699, S Hanley Rd., St. Louis, MO631444, USA.
24. Hyperchem6.0, Hypercube Inc., 1115, HW 4th Street, Gainsville, Il 32601, USA.
25. Powell, M. J. D. *Math. Program.* **1977**, *12*, 241.
26. Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
27. Viswanadhan, V. N.; Ghose, A. K.; Revenkar, G. R.; Robins, R. *J. Chem. Inf. Comput. Sci.* **1989**, *28*, 163.
28. Klebe, G. *J. Mol. Biol.* **1994**, *237*, 212.
29. Singh, J.; Vlijmen, H.-V.; Lee, W.-C.; Liao, Y.; Lin, K.-C.; Ateep, H.; Cuerrvo, J.; Zimmerman, C.; Hammond, C.; Karpusar, M.; Palmer, R.; Chattopadhyay, T.; Adams, S. P. *J. Comput. Aided. Mol. Design* **2002**, *36*, 201.
30. Cramer, R. D.; Bunce, J. D.; Patterson, D. E. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18.