3D QSAR Studies on Protein Tyrosine Phosphatase 1B Inhibitors: Comparison of the Quality and Predictivity among 3D QSAR Models Obtained from Different Conformer-Based Alignments

Gyanendra Pandey and Anil K. Saxena*

Medicinal and Process Chemistry Division, Central Drug Research Institute, Lucknow, India

Received June 1, 2006

A set of 65 flexible peptidomimetic competitive inhibitors (52 in the training set and 13 in the test set) of protein tyrosine phosphatase 1B (PTP1B) has been used to compare the quality and predictive power of 3D quantitative structure−activity relationship (QSAR) comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) models for the three most commonly used conformer-based alignments, namely, cocrystallized conformer-based alignment (CCBA), docked conformer-based alignment (DCBA), and global minima energy conformer-based alignment (GMCBA). These three conformers of 5-[(2*S*)-2-({(2*S*)-2-[(*tert*-butoxycarbonyl)amino]-3-phenylpropanoyl}amino)3-oxo-3-pentylamino)propyl]-2-(carboxymethoxy)benzoic acid (compound number 66) were obtained from the X-ray structure of its cocrystallized complex with PTP1B (PDB ID: 1JF7), its docking studies, and its global minima by simulated annealing. Among the 3D QSAR models developed using the above three alignments, the CCBA provided the optimal predictive CoMFA model for the training set with cross-validated r^2 (q^2) = 0.708, non-cross-validated r^2 = 0.902, standard error of estimate (s) = 0.165, and F = 202.553 and the optimal CoMSIA model with q^2 = 0.440, r^2 = 0.799, s = 0.192, and F = 117.782. These models also showed the best test set prediction for the 13 compounds with predictive r^2 values of 0.706 and 0.683, respectively. Though the QSAR models derived using the other two alignments also produced statistically acceptable models in the order DCBA > GMCBA in terms of the values of q^2, r^2, and predictive r^2, they were inferior to the corresponding models derived using CCBA. Thus, the order of preference for the alignment selection for 3D QSAR model development may be CCBA > DCBA > GMCBA, and the information obtained from the CoMFA and CoMSIA contour maps may be useful in designing specific PTP1B inhibitors.

INTRODUCTION

A quantitative structure−activity relationship (QSAR) gives a mathematical relationship between structural descriptors and biological activities.[1] The biological activities of unknown compounds are predicted from the model developed by the QSAR. The Hansch−Fujita method is a classical QSAR technique (2D QSAR), which constructs the multiple linear regression models between physicochemical descriptors of substituents and biological activities.[2,3] The predictive ability of the Hansch−Fujita model is restricted because the method uses only two-dimensional descriptors. The 2D descriptors are insufficient to explain complex structure−activity data and also to explain the drug−receptor interaction.

The 3D QSAR study includes the descriptors derived from individual atomic partial charges, highest occupied molecular orbital/lowest unoccupied molecular orbital energies, and molecular fields such as comparative molecular field analysis (CoMFA),[4] comparative molecular similarity indices analysis (CoMSIA),[5] and so forth. The improved techniques such as principal component analysis, principal component regression, and partial least-squares regression (PLS)[6] have been applied in 3D QSAR studies.

In the 3D QSAR studies, the molecular alignment and "active" conformation determination are so important that they affect the success of a model. One of the key steps in the 3D QSAR methodology is the selection of the active conformer for each compound in the series, followed by the molecular alignment rule. However, generating 3D conformations and the alignment of the compounds is a difficult and time-consuming process, especially when the compounds are large in size and contain several rotatable bonds.[7a] A choice of an active conformer and the corresponding alignment rule has to be undertaken, either in accordance with available experimental data or on the basis of hypothetical assumptions. Usually, a bioactive conformation of the ligand can be obtained from a structural determination of the ligand receptor complex by X-ray crystallography. One such example has been reported by Amin and Welsh,[7b] where the two different cocrystallized complexes were used for alignments to develop the CoMFA models. The field-fit procedure may be useful to define bioactive conformations of other derivatives. When no crystal structures are available, several theoretical methods are used to obtain the bioactive conformations.

The active analogue approach (AAA) is one of the techniques used to obtain the bioactive conformations of a set of molecules.[8] AAA generates the possible 3D conformations by conformational analysis and selects the best ones

* Corresponding author fax: 91-0522-26223405, e-mail: anilsak@gmail.com.

10.1021/ci600224n CCC: $33.50
Published on Web 10/21/2006

that satisfy the interatomic distances in a working hypothesis. However, this approach is more successful if there is one relatively rigid active compound or several somewhat rigid compounds that collectively restrict the conformational space. The conformation near the global energy minimum can also be considered as the active conformation for flexible molecule series. The three-way array data (multiway array, 2002) method is also used for solving the conformation problem in 3D QSAR studies.[9,10] In this method, three-way PLS was used to select the bioactive conformation from theoretically possible conformations. There are also published reports of successful 3D QSAR models where conformations and the alignment have been obtained by automated molecular docking.[11−15] There are some other examples where the moment of inertia is used for the molecular alignment in the CoMFA study. Generally, the global minima energy conformation of the most active molecule is used as a template for the molecular alignment in CoMFA and CoMSIA studies. Though several methods are available for the generation of the active conformation, comparisons of the quality and information value of 3D QSAR models based on different conformer-based alignments have been given little attention.

In the present study, the results and statistical quality of various models generated by the three most used conformer-based alignments, namely, cocrystallized conformer-based alignment (CCBA), docked conformer-based alignment (DCBA), and global minima energy conformer-based alignment (GMCBA), have been compared. The first conformer was generated by using the X-ray structure of a cocrystallized ligand as a template from the Protein Data Bank (PDB). Molecular docking using Autodock gave the second conformer, and the minimum energy conformation using simulated annealing of the ligand resulted in the third conformer. The alignments of the molecules were done on the basis of the above three different templates by using the most common structure (MCS) of the most active molecule. As the alignment of compounds plays a crucial role in the creation of CoMFA[4] and CoMSIA[5] models, the 3D QSAR models for each alignment were developed.

Protein tyrosine phosphatases constitute a growing family (>90 members) of receptor-like and cytoplasmic-signal-transducing enzymes.[16−19] These enzymes are classified on the basis of their cellular localization and are characterized by their common 230−280 amino acid catalytic domain termed as the PTP signature motif.[20,21] Several PTPases, including PTP1B, leukocyte-antigen-related tyrosine phosphatase, SH2-domain-containing phosphotyrosine phosphatase, PTPα, and PTPϵ, are capable of dephosphorylating the tyrosine residues[22,23] of the insulin receptor and thereby attenuating the tyrosine kinase activity. The phosphorylation (by protein tyrosine kinases) and dephosphorylation (by PTPases) of tyrosine residues in proteins is recognized as an important cellular regulatory mechanism, and many cellular functions could be artificially manipulated if one could exogenously control the activity of these kinases and phosphatases.[24] The disruption of the above process is at the root of a variety of disease states including cancer, inflammation, and diabetes.[25−28] Studies with PTP1B knockout mice, independently generated by two laboratories, have demonstrated that the targeted disruption of the PTP1B gene in mice resulted in an enhancement of insulin sensitivity and decreased susceptibility to diet-induced obesity.[29,30] Thus, PTP1B might be an attractive therapeutic target in type 2 diabetes as well as obesity and could provide a new therapeutic option to patients with at-risk obesity or type 2 diabetes. Several 3D QSAR studies of PTP1B inhibitors have been reported.[31−33] However, QSAR studies on the peptidomimetic competitive inhibitors, a major class of the PTP inhibitors, are not reported.

As a part of the ongoing work in our laboratory aimed at the search of antidiabetic compounds, it appeared of interest to develop 3D QSAR models on the peptidomimetic competitive inhibitors of protein tyrosine phosphatase 1B (PTP1B), reported by Larsen et al.[34] using the aforesaid strategy. These studies not only provided the insight into steric, electrostatic, hydrophobic, and hydrogen-bonding properties and essential structural features influencing the activity but also a comparative evaluation of three different conformer-based alignments in deriving different 3D QSAR models.

MATERIALS AND METHODS

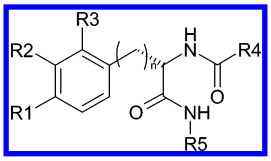
Data Set. The QSAR analysis, CoMFA, CoMSIA, and docking studies of PTB1B inhibitors were performed on a series of 66 peptidomimetic competitive inhibitors of protein tyrosine phosphatase 1B to quantitatively correlate the variation in structure and physicochemical properties with the in vitro PTP1 (rat homologue of PTP1B) inhibitory activity (Table 1). As the quality of the QSAR models not only depends on the high correlation coefficient (r^2) and r^2 (LOO) or q^2 values but also on the predictive r^2 value, the total number of 66 compounds with the measured activity data percentage inhibition, P ($0 < P > 100$) at a 100 μM dose, were divided into two sets, a training set with 53 molecules and a test set with 13 molecules covering the total activity range from least active to most active in both of the sets. The PTP1 activity data (P) were transformed as log(P/100 − P) for the QSAR analysis.[35]

Computational Approach. All of the molecular modeling studies were performed on a Silicon Graphics Octane R12000 workstation using Sybyl 6.9 (CoMFA, advance CoMFA, and CoMSIA), Insight II, and Autodock molecular modeling software.

The X-ray Conformer as Template. The X-ray conformer used as a template for the CoMFA and CoMSIA study was obtained from the X-ray structure of cocrystallized inhibitor PTP1B (PDB ID: 1JF7). The three-dimensional structure of PTP1B complexed with the inhibitor (compound number 66) was taken from the PDB (1JF7), and the PTP1B inhibitor was unmerged from the rest of the protein. This conformation was imported to Sybyl from Insight II, and the hydrogen atoms were added in Sybyl, and Gasteiger−Huckel charges with distance-dependent dielectrics were calculated.

Docking Protocol. Docking simulations on compound number 66 were performed with AutoDock 3.0.5 using a Lamarckian genetic algorithm.[36] The compound was sketched using standard facilities in Sybyl6.9; charges were calculated using the Gasteiger−Huckel method, and the geometry of the compound was subsequently optimized using the Tripos force field with default settings and a convergence criterion of 0.01 kcal/mol.

Table 1. Structures of the Inhibitors



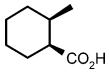
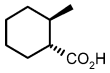
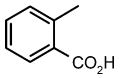
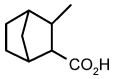
SN	R1	R2	R3	R4	R5	*n*
1	OSO2OH	H	H	$(CH_2)_2CO_2H$	*n*-C_5H_{11}	1
2	OCH(CO2H)2	H	H	$(CH_2)_2CO_2H$	*n*-C_5H_{11}	1
3	OCH(CO2H)2	H	H	$(CH_2)_2CO_2H$	*n*-C_5H_{11}	1
4	OCH(CO2H)2	F	H	$(CH_2)_2CO_2H$	*n*-C_5H_{11}	2
5	OCH(CO2H)2	F	F	$(CH_2)_2CO_2H$	*n*-C_5H_{11}	1
6	OCOCH2CO2H	H	H	$(CH_2)_2CO_2H$	*n*-C_5H_{11}	1
7	OCH2CO2H	H	H	$(CH_2)_2CO_2H$	*n*-C_5H_{11}	1
8[a]	OC(CO2H)═CHCO2H	H	H	$(CH_2)_2CO_2H$	*n*-C_5H_{11}	1
9[a]	OCH(CO2H)CH2CO2H	H	H	$(CH_2)_2CO_2H$	*n*-C_5H_{11}	1
10	CH═C(CO2H)2	H	H	$(CH_2)_2CO_2H$	*n*-C_5H_{11}	1
11	CH2CH(CO2H)2	H	H	$(CH_2)_2CO_2H$	*n*-C_5H_{11}	1
12	OCH2CO2H	H	H	$(CH_2)_2CO_2H$	*n*-C_5H_{11}	1
13[a]	OCH2CONH2	CO_2H	H	$(CH_2)_2CO_2H$	*n*-C_5H_{11}	1
14	OCH2CO2H	$CONH_2$	H	$(CH_2)_2CO_2H$	*n*-C_5H_{11}	1
15	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	*n*-C_5H_{11}	1
16	$OCH(CO_2Et)_2$	H	H	$(CH_2)_2CO_2H$	*n*-C_5H_{11}	1
17[a]	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2Bn$	*n*-C_5H_{11}	1
18	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CONH2$	*n*-C_5H_{11}	1
19	$OCH(CO_2H)_2$	H	H	$(CH_2)_3CO_2H$	*n*-C_5H_{11}	1
20	$OCH(CO_2H)_2$	H	H	$CH_2OCH_2CO_2H$	*n*-C_5H_{11}	1
21	$OCH(CO_2H)_2$	H	H	CH_2CO_2H	*n*-C_5H_{11}	1
22	$OCH(CO_2H)_2$	H	H	$CH_2C(Me)_2CH_2CO_2H$	*n*-C_5H_{11}	1
23[a]	$OCH(CO_2H)_2$	H	H	$CH_2C(c\text{-}CH2)_4CH_2CO_2H$	*n*-C_5H_{11}	1
24	$OCH(CO_2H)_2$	H	H		*n*-C_5H_{11}	1
25	$OCH(CO_2H)_2$	H	H		*n*-C_5H_{11}	1
26	$OCH(CO_2H)_2$	H	H		*n*-C_5H_{11}	1
27	$OCH(CO_2H)_2$	H	H		*n*-C_5H_{11}	1
28	$OCH(CO_2H)_2$	H	H		*n*-C_5H_{11}	1
29	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	*n*-C_6H_{13}	1
30	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	*n*-C_4H_9	1
31	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	*n*-$C_{10}H_{21}$	1
32	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	CH_2(c−C6H11)	1
33	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	$CH_2CH_2CH(Me)_2$	1
34	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	$(CH_2)_3Ph$	1
35[a]	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	CH2(1-naphthyl)	1
36	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	$(CH_2)_2CH(Ph)_2$	1
37	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	$CH_2CH(OEt)_2$	1
38	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	CH_2CH_2(4-SO_2NH_2−Ph)	1
39	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	CH_2CH_2(4-OH−Ph)	1
40	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	CH_2(4-CO_2H−Ph)	1
41[a]	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	CH_2(4-CF_3−Ph)	1
42	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	CH_2(3,4-OCH_2O−Ph)	1
43	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	CH_2CH_2OPh	1
44	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	2-(1-piperidinyl)ethyl	1
45[a]	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	3-(4-morpholinyl)propyl	1
46	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	3-(1-imidazolyl)propyl	1
47	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	(S)-Nle-NH_2	1
48	$OCH(CO_2H)_2$	H	H	$(CH_2)_2CO_2H$	(S)-leucinol	1
49	$OCH(CO_2H)_2$	H	H	$CH(CH_2CO_2H)NHCOO$(t-Bu)	*n*-C_5H_{11}	1
50	$OCH(CO_2H)_2$	H	H	$CH(CH_2CO_2H)NHCOCH_2CH_2CO_2H$	*n*-C_5H_{11}	
51	$OCH(CO_2H)_2$	H	H	$CH(CH_2CO_2H)NHCOC_5H_{11}$	*n*-C_5H_{11}	1
52[a]	$OCH(CO_2H)_2$	H	H	$CH(CH_2$(4-Bz-Ph))$NHCOO$(*t*-Bu)	*n*-C_5H_{11}	1
53	$OCH(CO_2H)_2$	H	H	$CH(CH_2$(4-Bz-Ph))$NHCOCH_2CH_2CO_2H$	*n*-C_5H_{11}	1
54[a]	$OCH(CO_2H)_2$	H	H	$CH(CH_2CH_2CO_2H)NHCOCH_2CH_2CO_2H$	*n*-C_5H_{11}	1
55	$OCH(CO_2H)_2$	H	H	$CH(CH_2CH_2CO_2H)NHCOO$(*t*-Bu)	*n*-C_5H_{11}	1
56	$OCH(CO_2H)_2$	H	H	$CH(CH_2CO_2H)NHCOOCH_2Ph$	*n*-C_5H_{11}	1
57	$OCH(CO_2H)_2$	H	H	$CH[R\text{-}CH_2$(4-Bz-Ph))$NHCOCH_2CH_2CO_2H$	*n*-C_5H_{11}	1

Table 1. (Continued)

SN	R1	R2	R3	R4	R5	*n*
58	$OCH(CO_2H)_2$	H	H	$CH[CH_2(4\text{-}BnOPh))NHCOCH_2CH_2CO_2H$	$n\text{-}C_5H_{11}$	1
59[a]	$OCH(CO_2H)_2$	H	H	$CH[CH_2\{4\text{-}(2,5\text{-}Cl_2Bn)OPh\}]NHCOCH_2CH_2CO_2H$	$n\text{-}C_5H_{11}$	1
60	$OCH(CO_2H)_2$	H	H	$CH[CH_2(4\text{-}MeOPh)]NHCOCH_2CH_2CO_2H$	$n\text{-}C_5H_{11}$	1
61	$OCH(CO_2H)_2$	H	H	$CH(CH_2Ph)NHCOCH_2CH_2CO_2H$	$n\text{-}C_5H_{11}$	1
62[a]	$OCH(CO_2H)_2$	H	H	$CH[CH_2\{4\text{-}(2,5\text{-}Cl_2Bn)OPh\}]NHCOO(t\text{-}Bu)$	$n\text{-}C_5H_{11}$	1
63	$OCH(CO_2H)_2$	H	H	$CH[CH_2(4\text{-}MeOPh)]NHCOO(t\text{-}Bu)$	$n\text{-}C_5H_{11}$	1
64	$OCH(CO_2H)_2$	H	H	$CH(CH_2Ph)NHCOO(t\text{-}Bu)$	$n\text{-}C_5H_{11}$	1
65[a]	$OCH(CO_2H)_2$	H	H	$CH(CH_3)]NHCOO(t\text{-}Bu)$	$n\text{-}C_5H_{11}$	1
66	OCH_2COOH	COOH	H	$CH(CH_2Ph)NHCOO(t\text{-}Bu)$	$n\text{-}C_5H_{11}$	1

[a] Test set molecules.

Hydrogens were added to PTP structure using Insight II, and potentials were fixed using the CVFF force field. The PTP structure was then imported to Autodock, and nonpolar hydrogens were merged followed by the calculation of Kollman-all-atom charges.

The standard docking procedure for a rigid protein and a flexible ligand was used. A grid of 60, 60, and 60 points in the *x*, *y*, and *z* directions was built, centered on the center of the mass of the catalytic site of the protein. A grid spacing of 0.375 Å and a distance-dependent dielectric constant were used for the calculation of the energy map. The default settings were used for all other parameters. At the end of docking, the ligand conformation with the most favorable free energy of binding in terms of hydrogen-bonding stabilization energy, van der Waals internal energy for the conformation of the compound, and complex interaction energy between the compound and the target protein in the resultant complex structures was selected as a template for DCBA.

Simulated Annealing. Simulated annealing was used for obtaining the global minima energy conformation of compound no 66. It was built using fragments in the Sybyl fragment library, and the geometry was fully optimized using the standard Sybyl settings. Partial charges were computed using the Gasteiger−Huckel method, and the structure was subjected to molecular dynamic simulation to heat the molecule at 700 K for 1000 fs followed by annealing the molecule to 300 K for 1500 fs and minimizing all conformers to a gradient of 0.01, and the minimum energy conformer thus obtained was used as a template for GMCBA.

Molecular Structures and Alignment. The above three conformers were used as three different templates for the building of molecular structures. The structures of all inhibitors were constructed by modifying each template by using Sybyl builder. The partial charges for all compounds were calculated using the Gasteiger−Huckel method, and the molecules were optimized for their geometry using the Tripos force field with a distance-dependent dielectric function and energy convergence criterion of 0.01 kcal/mol Å using 1000 iterations and standard Sybyl settings.

All molecules were aligned using the MCS-based alignment approaches using the most active molecule or the molecule having the highest number of functional groups as a criterion. The core used for common structure-based alignment is shown in Figure 1. Three different alignments obtained using three different conformers, namely, CCBA, DCBA, and gGMCBA are shown in Figures 2, 3, and 4, respectively.

CoMFA Studies. The superimposed molecules were kept in a 3D grid (spacing set at 2 Å), and steric and electrostatic

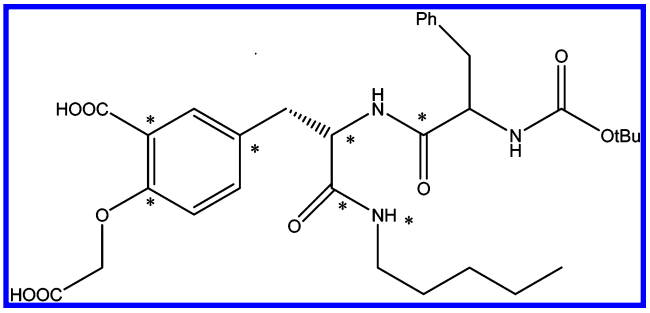


Figure 1. Structure of molecule 66 that was used as a template on which all of the other 65 molecules were superimposed using the MCS procedure. The stars indicate the atoms selected as the fitting centers.

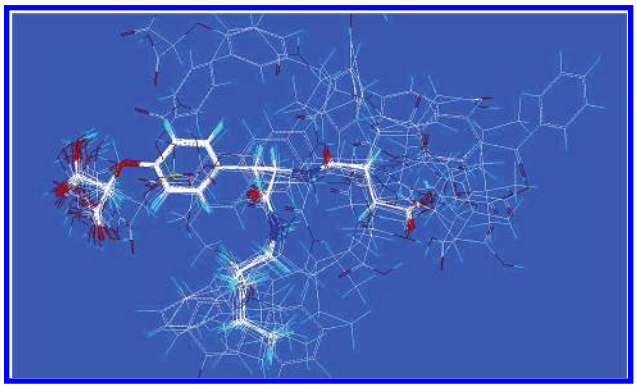


Figure 2. Cocrystallized conformer-based alignment (CCBA).

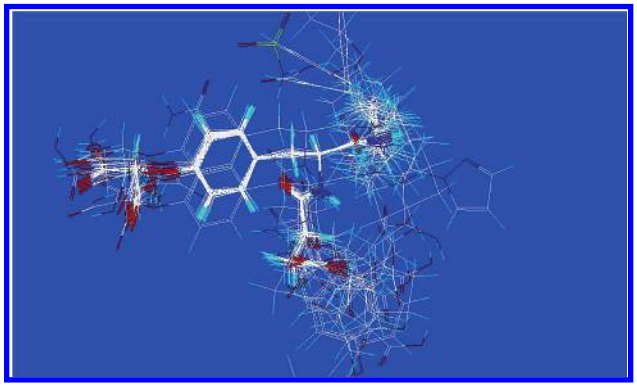


Figure 3. Docked conformer-based alignment (DCBA).

fields were calculated at various grid points using Lennard-Jonnes and Coulombic potentials, respectively, for CoMFA studies. An sp^3 carbon atom having a charge of +1 and a radius of 1.52 Å was used as a probe to calculate various steric and electrostatic fields for all three of the alignments. Various steric and electrostatic cutoffs and grid spacings were

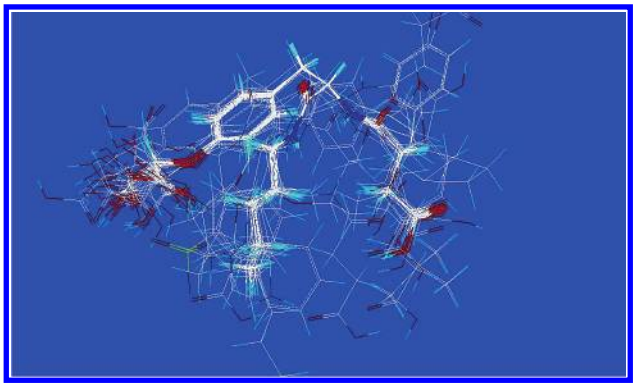


Figure 4. Global minima energy conformer-based alignments (GMCBA).

tried to investigate the influence of different parameter settings on CoMFA.

Advance CoMFA. H-bonding, indicator, and parabolic fields were used in the advanced CoMFA routine to extend the scope of CoMFA. The H-bonding fields are special indicator fields based on the description by Bohacek and McMartin.[37] Lattice points with steric energies below the steric cutoff energy are assigned nominal energies equal to the steric cutoff energy if they are close to H-bond-accepting or -donating atoms. Indicator fields[38] replace all lattice energies with magnitudes below a designated threshold with zero values. When both fields are included in a single CoMFA column, the greater of the steric and electrostatic cutoffs is used. The sign of the original lattice energy is retained. Parabolic fields are those in which the magnitude of a standard steric or electrostatic field at each lattice point has been squared; the original sign of the energy is retained.

CoMSIA Studies. Klebe introduced the CoMSIA technique in 1994, in which similarity indices are calculated at different points in a regularly spaced grid for prealigned molecules. It has several advantages over the CoMFA technique like greater robustness regarding both region shifts and small shifts within the alignments, no application of arbitrary cutoffs, and more intuitively interpretable contour maps. The standard settings (probe with charge +1, radius 1 Å and hydrophobicity +1, hydrogen-bond donating +1, hydrogen-bond accepting +1, attenuation factor α 0.3, and grid spacing 2 Å) were used in CoMSIA to calculate five different fields, namely, steric, electrostatic, hydrophobic, acceptor, and donor.

Partial Least-Squares Analysis (PLS). PLS is used to correlate the PTP inhibitory activity with the CoMFA and CoMSIA values containing the magnitude of steric, electrostatic, and other potentials. The models were assessed by leave-one-out r^2 (q^2) using the SAMPLS[39] method as implied in Sybyl. A strict criterion for the selection of the optimal number of components was applied by selecting the lowest S_{press} value, and wherever the last added component increased q^2 less than 5%, the less complex model was chosen. The PLS analysis was run (without cross-validation) with the optimal number of components to get predictive models. A column filtering of 2.0 kcal/mol was applied to reduce the noise and to speed up the calculation. CoMFA standard scaling was applied to all of the CoMFA analysis, and autoscaling was applied to all CoMSIA analyses.

Model Validation. The predictive power of all CoMFA and CoMSIA models was further validated by using an external test set (inhibitors marked with *a* in Table 1). The inhibitors in the test set were given exactly the same pretreatment as the inhibitors in the corresponding training sets. The correlation between the experimental and predicted activity for all models was calculated as a predictive r^2 value.

RESULTS

CoMFA, Advance CoMFA, and CoMFA Region Focusing. A total of 24 CoMFA models were developed from the training set of 52 PTP inhibitors using alignments from three different conformers, and the results of the CoMFA and CoMSIA analyses are presented in Tables 2−7. The models for each alignment describe the best results obtained at a cutoff of 30 kcal/mol both for steric and for electrostatic fields. The fields were focused using *StDeν*Coefficient* at a weight of 0.8 for the alignments.

Cocrystallized Conformer-Based Alignment. The CCBA produced the most consistent results among all of the three alignments, and the model with the indicator field showed the highest q^2 value (0.708) after region focusing. Among the rest of the seven models, the Tripos standard field showed the second highest q^2 (0.665) followed by the CoMFA parabolic field ($q^2 = 0.430$) and the rest with $0.430 > q^2 > 0.400$.

Docked Conformer-Based Alignment. Among the statistically significant models in terms of r^2, SE, Fischer's F-test for significance (F value), and r_{bs}^2 developed using the DCBA, the model with the Tripos standard field has the highest q^2 (0.451), while that with the hydrogen-bonding field had the lowest q^2 (0.250) with five and one components, respectively. After region-focusing of the models, the Tripos standard field produced the highest q^2 (0.579), while models with the indicator and parabolic field showed a q^2 of 0.531 and 0.501, respectively. The hydrogen-bonding field showed the lowest q^2 value (0.282) (Table 3).

Global Minima Energy Conformer-Based Alignment. Among the models developed using the GMCBA, the model with the Tripos standard field with three components showed a q^2 value of 0.404, while the models with indicator (two components), parabolic (two components), and hydrogen-bonding (one component) fields showed significant q^2 values of 0.419, 0.450, and 0.275, respectively. The application of the region-focusing improved the q^2 values, which was the highest (0.547) in the case of the Tripos standard field followed by the q^2 value (0.540) for the parabolic field (Table 4).

CoMSIA. A total of nine CoMSIA models were generated using the above three different alignments, namely, CCBA, DCBA, and GMCBA. The CoMSIA analysis was done at a grid spacing of 2 Å, and the column filtering was set at 2.0 kcal/mol while the other parameters were set as a default. The selection criterion for the number of components in the model was similar to the one used and discussed in the case of CoMFA studies.

Cocrystallized Conformer-Based Alignment. Among all of the fields, namely, steric, electrostatic, H-bond donor, H-bond acceptor, and hydrophobic, used in the analysis for CCBA, the highest q^2 value (0.44) was obtained for the CoMSIA steric field (five components) with the highest r^2 value (Table 5). The model generated using the donor and

Table 2. CoMFA, Advanced CoMFA, CoMFA Region-Focusing (CCBA)

parameters[a]	CoMFA	CoMFA (RF) Gs = 1 A^0	H-bond	H-bond (RF) Gs = 1 A^0	indicator	indicator (RF) Gs = 1 A^0	parabolic	parabolic (RF) Gs = 1 A^0
q^2	0.305	0.665	0.245	0.303	0.419	0.708	0.324	0.430
s_{press}	0.654	0.470	0.662	0.642	0.593	0.429	0.626	0.573
r^2	0.945	0.947	0.507	0.342	0.918	0.902	0.486	0.519
SE	0.185	0.187	0.535	0.624	0.224	0.165	0.546	0.528
F	200.821	111.610	51.424	12.729	132.076	202.553	47.286	54.013
component	4	7	1	2	4	5	1	1
fraction								
steric	0.436	0.494	0.484	0.856	0.939	0.732	0.478	0.546
electrostatic	0.564	0.506	0.516	0.144	0.061	0.268	0.522	0.454
r_{bs}^2 (100 runs)	0.954	0.977	0.596	0.375	0.949	0.972	0.557	0.557
r_{pred}^2	0.575	0.188	0.154	0.422	0.498	0.729	0.248	0.262
model no.	1	2	3	4	5	6	7	8

[a] q^2 = leave-one-out r^2, s_{press} = standard error of leave-one-out validation, r^2 = coefficient of determination, SE = standard error, F = Fischer's F value for test of significance, n = no. of components, fraction = fraction of the respective field, r_{bs}^2 = coefficient of determination after 100 bootstrapping runs, r_{pred}^2 = predictive r^2.

Table 3. CoMFA, Advanced CoMFA, CoMFA Region-Focusing (DCBA)

parameters[a]	CoMFA	CoMFA (RF) Gs = 1 A^0	H-bond	H-bond (RF) Gs = 1 A^0	indicator	indicator (RF) Gs = 1 A^0	parabolic	parabolic (RF) Gs = 1 A^0
q^2	0.451	0.579	0.250	0.282	0.409	0.531	0.433	0.501
s_{press}	0.584	0.494	0.660	0.645	0.610	0.538	0.598	0.561
r^2	0.939	0.747	0.486	0.350	0.923	0.866	0.926	0.912
SE	0.195	0.383	0.546	0.614	0.220	0.287	0.216	0.236
F	142.619	147.602	47.284	26.916	110.798	76.131	114.770	94.978
component	5	1	1	1	5	4	5	5
fraction								
steric	0.593	0.449	0.443	0.539	0.958	0.430	0.873	0.580
electrostatic	0.407	0.551	0.557	0.461	0.042	0.570	0.127	0.420
r_{bs}^2 (100 runs)	0.964	0.745	0.596	0.370	0.954	0.905	0.894	0.796
r_{pred}^2	0.167	0.569	0.204	0.276	0.018	0.531	0.114	0.124
model no.	9	10	11	12	13	14	15	16

[a] q^2 = leave-one-out r^2, s_{press} = standard error of leave-one-out validation, r^2 = coefficient of determination, SE = standard error, F = Fischer's F value for test of significance, n = no. of components, fraction = fraction of the respective field, r_{bs}^2 = coefficient of determination after 100 bootstrapping runs, r_{pred}^2 = predictive r^2.

Table 4. CoMFA, Advanced CoMFA, CoMFA Region-Focusing (GMCBA)

parameters[a]	CoMFA	CoMFA (RF) Gs = 1 A^0	H-bond	H-bond (RF) Gs = 1 A^0	indicator	indicator (RF) Gs = 1 A^0	parabolic	parabolic (RF) Gs = 1 A^0
q^2	0.404	0.547	0.275	0.348	0.419	0.510	0.450	0.540
s_{press}	0.594	0.512	0.648	0.634	0.589	0.533	0.571	0.516
r^2	0.765	0.612	0.517	0.539	0.742	0.598	0.774	0.631
SE	0.373	0.374	0.529	0.533	0.391	0.481	0.365	0.462
F	79.578	78.922	53.549	13.754	70.940	74.327	84.058	85.631
component	2	1	1	4	2	1	2	1
fraction								
steric	0.489	0.805	0.474	0.849	0.939	0.534	0.470	0.736
electrostatic	0.511	0.195	0.526	0.151	0.061	0.466	0.530	0.264
r_{bs}^2 (100 runs)	0.799	0.628	0.598	0.651	0.794	0.625	0.811	0.661
r_{pred}^2	0.418	0.599	0.601	0.366	0.624	0.659	0.624	0.697
model no.	17	18	19	20	21	22	23	24

[a] q^2 = leave-one-out r^2, s_{press} = standard error of leave-one-out validation, r^2 = coefficient of determination, SE = standard error, F = Fischer's F value for test of significance, n = no. of components, fraction = fraction of the respective field, r_{bs}^2 = coefficient of determination after 100 bootstrapping runs, r_{pred}^2 = predictive r^2.

acceptor field showed the q^2 value 0.330, while the model with steric and electrostatic field showed the q^2 value 0.255.

Docked Conformer-Based Alignment. Among the models generated using DCBA for all of the fields, the model with the CoMSIA steric field (two components) showed the highest q^2 (0.419) and r^2 (0.820) values. The q^2 values for the models with the donor and acceptor fields was 0.388, and that for the steric and electrostatic fields was 0.377 (Table 6).

Global Minima Energy Conformer-Based Alignment. The model generated for this alignment with the steric and electrostatic field (two components) showed the highest q^2 (0.396) value followed by the q^2 (0.339) for the model with the steric field (one component; Table 7).

DISCUSSION

The statistical information and quality of 3D QSAR models based on three different alignments have been

Table 5. CoMSIA Studies (CCBA)

parameters[a]	ST	EL	HP	DO	AC	ST&EL	DO&AC
q^2	0.440	0.185	0.254	0.254	0.202	0.255	0.330
s_{press}	0.605	0.687	0.657	0.608	0.680	0.651	0.629
r^2	0.799	0.344	0.402	0.660	0.387	0.412	0.684
SE	0.313	0.616	0.589	0.448	0.597	0.584	0.432
F	50.006	26.271	33.649	47.589	31.377	35.078	48.363
component	5	1	1	2	1	1	2
relative contribution							
S	1.0					0.391	
E		1.0				0.609	
H			1.0				
r_{bs}^2 (100 runs)	0.888	0.437	0.468	0.730	0.475	0.495	0.756
r_{pred}^2	0.653	0.227	0.282	0.210	0.145	0.320	0.426
model no.	25	26	27	28	29	30	31

[a] q^2 = leave-one-out r^2, s_{press} = standard error of leave-one-out validation, r^2 = coefficient of determination, SE = standard error, F = Fischer's F value for test of significance, n = no. of components, fraction = fraction of the respective field, r_{bs}^2 = coefficient of determination after 100 bootstrapping runs, r_{pred}^2 = predictive r^2 , ST = steric field, EL = electrostatic field, HP = hydrophobic field, DO = hydrogen-bond-donor field, AC = hydrogen-bond-acceptor field.

Table 6. CoMSIA Studies (DCBA)

parameters[a]	ST	EL	HP	DO	AC	ST&EL	DO&AC
q^2	0.419	0.257	0.336	0.376	0.350	0.377	0.388
s_{press}	0.567	0.670	0.627	0.609	0.620	0.607	0.602
r^2	0.626	0.799	0.605	0.791	0.606	0.701	0.671
SE	0.470	0.348	0.484	0.355	0.484	0.420	0.441
F	41.062	63.714	37.459	60.516	37.451	57.486	50.074
component	2	3	2	3	2	2	2
relative contribution							
S	1.0					0.435	0.429
E		1.0				0.565	0.571
H			1.0				
r_{bs}^2 (100 runs)	0.660	0.870	0.696	0.837	0.719	0.784	0.751
r_{pred}^2	0.650	0.441	0.438	0.561	0.588	0.640	0.013
model no.	32	33	34	35	36	37	38

[a] q^2 = leave-one-out r^2, s_{press} = standard error of leave-one-out validation, r^2 = coefficient of determination, SE = standard error, F = Fischer's F value for test of significance, n = no. of components, fraction = fraction of the respective field, r_{bs}^2 = coefficient of determination after 100 bootstrapping runs, r_{pred}^2 = predictive r^2 , ST = steric field, EL = electrostatic field, HP = hydrophobic field, DO = hydrogen-bond-donor field, AC = hydrogen-bond-acceptor field.

Table 7. CoMSIA Studies (GMCBA)

parameters[a]	ST	EL	HP	DO	AC	ST&EL	DO&AC
q^2	0.339	0.184	0.245	0.259	0.192	0.369	0.277
s_{press}	0.619	0.603	0.644	0.655	0.684	0.589	0.648
r^2	0.492	0.529	0.443	0.456	0.434	0.543	0.516
SE	0.543	0.528	0.568	0.561	0.542	0.520	0.530
F	48.405	27.536	39.690	41.982	48.818	29.078	53.226
component	1	2	1	1	1	2	1
relative contribution							
S	1.0					0.462	
E		1.0				0.538	
H			1.0				
r_{bs}^2 (100 runs)	0.526	0.654	0.502	0.532	0.571	0.644	0.580
r_{pred}^2	0.366	0.072	0.114	0.014	0.104	0.274	0.056
model no.	39	40	41	42	43	44	45

[a] q^2 = leave-one-out r^2, s_{press} = standard error of leave-one-out validation, r^2 = coefficient of determination, SE = standard error, F = Fischer's F value for test of significance, n = no. of components, fraction = fraction of the respective field, r_{bs}^2 = coefficient of determination after 100 bootstrapping runs, r_{pred}^2 = predictive r^2 , ST = steric field, EL = electrostatic field, HP = hydrophobic field, DO = hydrogen-bond-donor field, AC = hydrogen-bond-acceptor field.

compared, as the alignment of the molecules is the most crucial step in the development of the 3D QSAR models using CoMFA and CoMSIA. The 3D QSAR models with a q^2 value > 0.3 are considered significant, although a q^2 value > 0.4 is preferred. The best models from CoMFA (6, 10, and 28) and CoMSIA (25, 32, and 44) were selected for each alignment. The predicted inhibitory activities of the 52 CoMFA and CoMSIA training set inhibitors with CoMFA model numbers 6, 10, and 28 and CoMSIA model numbers 25, 32, and 44 are listed in Table 8. There were no considerable outliers in the predictions by any model, indicating both their robustness and their predictive power. Among all 24 models generated using three different alignments, the best model was CoMFA model number 6,

Table 8. Observed and Predicted Activities of the Training and Test Set Compounds by CoMFA and CoMSIA Models

SN	observed activity (P)	log $P/(100-P)$	predicted activity[a]	predicted activity[b]
1	80	0.602	0.632	0.896
2	76	0.5	0.468	0.698
3	18	−0.658	−0.927	0.365
4	78	0.549	0.465	0.987
5	85	0.753	0.698	0.265
6	2	−1.69	−1.734	−1.365
7	14	−0.788	−0.697	−0.987
8[c]	50	0	−0.452	0.321
9[c]	34	−0.288	−0.266	0.021
10	33	−0.307	−0.138	−0.968
11	34	−0.288	−0.47	−0.654
12	93	1.123	1.325	1.365
13[c]	16	−0.72	−1.2	0.325
14	11	−0.907	−0.634	−0.865
15	39	−0.194	−0.111	−0.653
16	1	−1.995	−1.754	−1.865
17[c]	39	−0.194	0.123	−0.632
18	51	0.017	0.015	0.123
19	67	0.307	0.369	0.253
20	53	0.052	0.069	0.236
21	74	0.454	0.698	0.562
22	67	0.307	0.564	0.365
23[c]	72	0.41	0.869	0.685
24	62	0.212	0.236	0.658
25	72	0.41	0.456	0.652
26	57	0.122	0.256	0.865
27	63	0.231	0.365	0.985
28	37	−0.231	−0.08	−0.635
29	74	0.454	0.459	0.985
30	37	−0.231	−0.325	−1.23
31	51	0.017	0.105	0.362
32	73	0.431	0.546	0.652
33	48	−0.034	−0.02	−0.0652
34	72	0.41	0.658	0.635
35[c]	61	0.194	0.325	0.698
36	45	−0.087	0.027	−0.0154
37	46	−0.069	0.026	−0.052
38	63	0.231	0.325	0.962
39	43	−0.122	−0.153	−0.321
40	56	0.104	0.156	0.301
41[c]	66	0.288	1.236	0.326
42	63	0.231	0.365	0.352
43	57	0.122	0.236	0.325
44	23	−0.524	−0.304	−0.856
45[c]	35	−0.268	−1.98	−0.985
46	26	−0.454	−0.658	−0.852
47	93	1.123	1.265	1.985
48	57	0.122	0.369	0.652
49	90	0.954	0.985	0.125
50	96	1.38	1.36	1.698
51	91	1.004	0.132	1.056
52[c]	91	1.004	1.36	2.01
53	98	1.69	1.98	2.36
54[c]	96	1.38	1.56	1.365
55	83	0.688	0.693	1.23
56	90	0.954	0.698	1.652
57	54	0.069	0.987	0.123
58	98	1.69	1.125	1.982
59[c]	98	1.69	1.58	1.895
60	98	1.69	1.874	2.01
61	97	1.509	1.875	2.1
62[c]	85	0.753	0.987	0.852
63	90	0.954	1.23	0.968
64	88	0.865	1.02	0.652
65[c]	76	0.5	0.652	0.981
66	84	0.72	0.985	0.523

[a] Predicted activity by model no. 6. [b] Predicted activity by model no. 9. [c] Test set molecules.

having a q^2 of 0.708 for the CCBA. This model had better statistics than those based on the DCBA and GMCBA. The observed q^2 values > 0.400 for most of the fields before and after region-focusing for CCBA suggests that the CCBA may be considered most suitable for the CoMFA analysis and may be useful in explaining the variation of PTP activity of this class of compounds. Nevertheless, all of the alignment methods yielded statistically significant and predictive models.

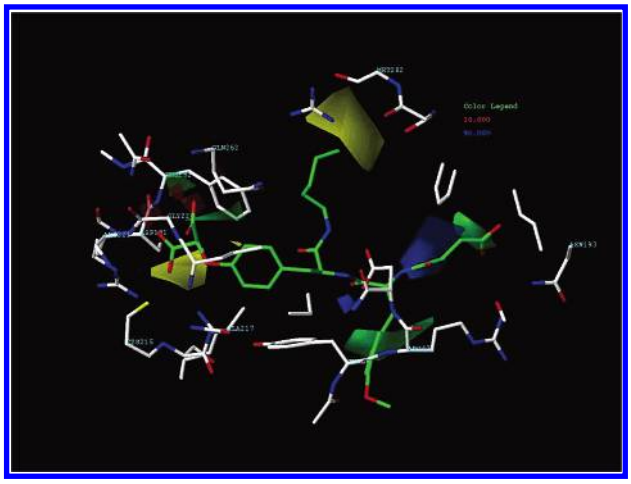


Figure 5. CoMFA contour map (focused tripos standard field) superimposed on the receptor active site (molecule number 60, in green).

The CoMSIA models also showed results similar to CoMFA models where the models generated from CCBA were slightly superior in terms of their statistics than those based on DCBA and GMCBA. The X-ray crystal-structure-bound conformer-based alignment produced the best 3D QSAR model in CoMSIA (model number 9) with the steric field ($q^2 = 0.440$). Nevertheless, the models (20 and 32) with other alignments were also statistically significant with the q^2 values 0.419 and 0.369, respectively. The relatively lesser difference in the q^2 values of these models as compared to the corresponding differences in the q^2 values of the three CoMFA models for the three alignments may be due to a lack of sensitivity of the CoMSIA method to small changes in the orientation of the compounds,[40] compared to the stronger dependence of conformation and orientation in CoMFA.

Comparison of CoMFA Contour Maps. Statistically relevant information from CoMFA analyses has been represented as 3D contour maps. Though there were some differences in the sizes of the favored and disfavored regions, the types of the regions were equivalent with no significant disagreements among the CoMFA and CoMSIA maps of the models based on different conformer-based alignment. Because the CCBA produced the most consistent and highly significant results, this alignment was used for generation of the contour maps for CoMFA and CoMSIA using the field type *StDev*Coefficient*. The contour maps using CoMFA region-focusing are of better quality because the region-focusing resulted in a sort of image enhancement.[41] The contour maps of CoMFA denote the region in the space where the aligned molecules would favorably or unfavorably interact with the receptor, while the CoMSIA contribution maps denote those areas within the specified region where the presence of a group with a particular physicochemical property will be favored or disfavored for good biological activity. The contour maps (Figures 5−8) show contributions for favorable and unfavorable interactions with the receptor

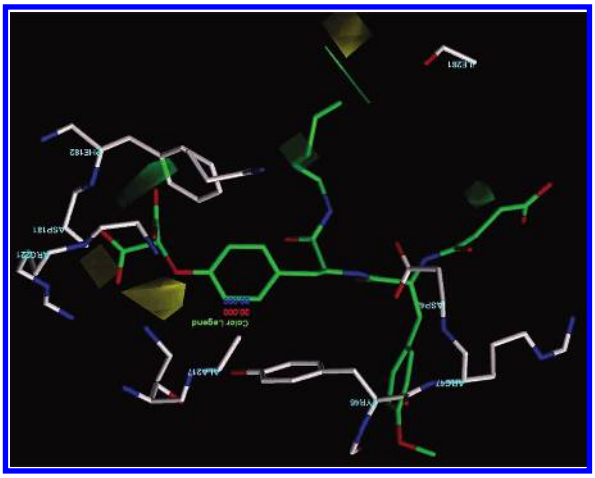


Figure 6. CoMFA contour map (focused tripos standard field) superimposed on the receptor active site (molecule number 60, in green).

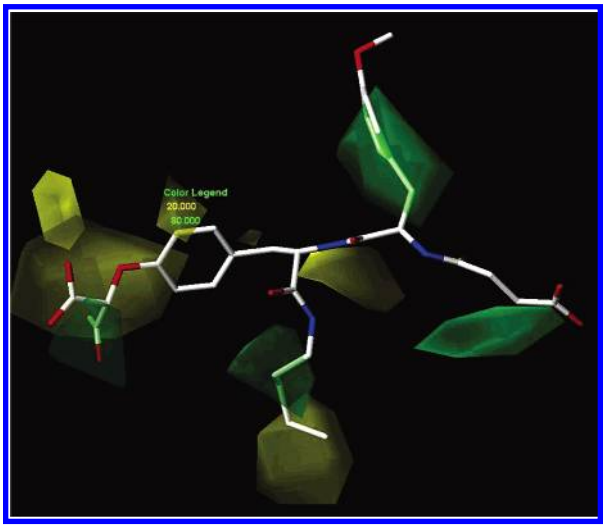


Figure 7. CoMSIA steric contour map (molecule number 60).

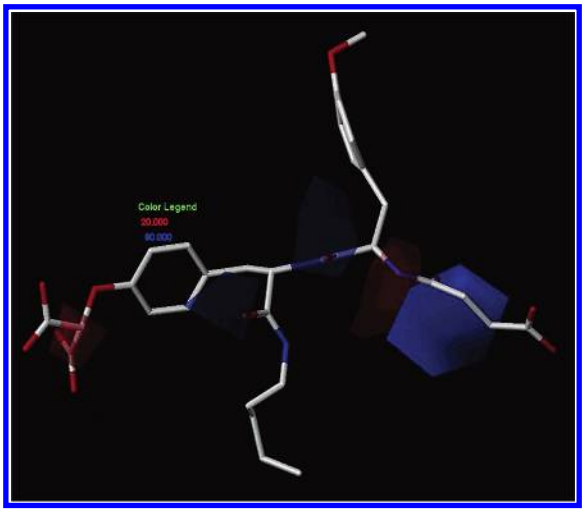


Figure 8. CoMSIA steric contour map (molecule number 60).

in terms of steric (80% green and 20% yellow) and electrostatic (80% blue and 20% red). Greater values of bioactivity are correlated with more bulk near green, less bulk near yellow, more positive charge near blue, and more negative charge near red. Both the CoMFA and CoMSIA electrostatic contours show the dominance of red polyhedra at the R1 position over the blue polyhedra, thus suggesting that the negative-charged substituents are favorable for activity at this position, thus suggesting that the binding pocket surrounding this group should have a positive field; such a binding pocket has also been suggested by Peters et al.[42] The absence of polyhedra at the R5 position indicates that the electrostatic nature of the substituents at this position has no influence on the inhibitory activity. As the position of R5 is far from the peptide backbone, the results corroborate with the studies carried out by Sarmiento et al.[43] and Jia et al.[44] where the major interactions of the inhibitors with the enzyme involve the peptide backbone. There is one red and two blue polyhedra present at the R4 position in the CoMFA and CoMSIA contour maps, respectively. Thus, positively charged groups are more important at R4 as compared to the negatively charged substituents.

Further, the steric contour maps indicate that less bulky groups are favorable for the activity at the R1 and R5 positions. It is also clear from the observed activity data that most of the active compounds have a *n*-C_5H_{11} group at the R5 position. Though a large steric variation at the R1 position is not reported in the paper, still it is clear from the contour maps that the activity may increase with less bulky groups with a more electronegative nature that are present at the R1 position. Thus, the results of these CoMFA and CoMSIA studies well corroborated with the knowledge available in the literature about the binding sites of the enzyme and their interactions with the ligands.

Superimposition of the CoMFA Contour Map on the Receptor Active Site. The CoMFA contour maps of the two best models were superimposed on the receptor active site (Figures 5 and 6) to check the complementarity of the CoMFA fields against that of the active site. It is evident from Figures 5 and 6 that the CoMFA contours are in good agreement with the receptor active site. The yellow contour showing the unfavorable region for steric bulk around the alkyl amide side chain (near Met282, Figures 5 and 6) is explainable as it has a restricted area and may not adjust more bulk. The green and the blue contours near the peptidic nitrogens of the ligand around Arg47 and Asp48 of the receptor are consistent with the observation that H-bonding between the molecule and carboxylate of Asp48 (Figure 5) is an important interaction, and increasing electropositive character on the peptide nitrogens here may enhance the interaction between the molecule and negatively charged aspartate. The second important interaction between Arg221 and the carboxylate of the molecule is also consistent with the red polyhedra in CoMFA contours (Figure 5) around this region, showing the need of increasing negative charge to enhance the interaction with positively charged arginine nitrogens. Further, the yellow polyhedra around Phe182 of the receptor indicate that steric bulk is not allowed, which is again consistent with the active site topology. Thus, it can be concluded from the above discussion that the CoMFA contour maps are consistent with the active site of the receptor and not only explain well the observed interactions at the receptor but may also suggest the structural modulation for increasing activity, and the CoMFA model may be used to quantitatively estimate the activity of new designed molecules.

Model Validation. The ability to predict the activity of inhibitors that were not included in the model building is

the most important test for the model's predictivity; hence, all of the above CoMFA- and CoMSIA-derived models were validated by using an external test set of 14 inhibitors. Predicted activities of the test set inhibitors for CoMFA model numbers 6, 13, and 24 and CoMSIA model numbers 9, 20, and 32 are presented in Table 5. Where the r_{pred}^2 values of CCBA-based models (2, 6, and 9) were higher than those of the other two DCBA- and GMCBA-based models. These results further substantiated that the CCBA produced statistically better CoMFA and CoMSIA models which also showed better external test set prediction. Thus, the use of the X-ray bioactive conformation in the alignment of the molecules for 3D QSAR model building using CoMFA and CoMSIA may be more helpful in getting robust models of high predictive value and providing pharmacophores interpretable in terms of interactions at the active site. Such models may be useful in optimizing the activity and in the design of new chemical entities (NCEs) as ligands.

CONCLUSION

The CoMFA and CoMSIA studies carried out using three different conformer-based alignments have led to the identification of the regions important for steric, hydrophobic, and electronic interactions. The derived models well explain the observed variance in the activity and also provide important insight into structural variations that can lead to the design of NCEs with high activity. The studies also demonstrate the comparative study among the three most commonly used conformers for the alignments in CoMFA and CoMSIA studies. The studies show that the X-ray bound conformation is best for the alignment of the molecules for the CoMFA and CoMSIA studies. The other two conformer-based alignments have also produced statistically significant and predictive models, suggesting that the docked and global minima energy conformers can be used for alignments in those cases where the X-ray structure of the protein is unknown. So, in cases where the X-ray structure of the ligand enzyme complex is known, the CCBA-based models may be given preference over the DCBA- and GMCBA-based models. These models also better explain the interaction at the binding site provably because of the consideration of the X-ray conformation of the ligand derived from its structure complexed with the enzyme. In our studies, these models maintained their superiority in explaining the variation in activity in both the training and test sets.

ACKNOWLEDGMENT

The author (G.P.) thanks the DST for the financial assistance in the form of a fellowship. Anshuman Dixit is acknowledged for valuable discussions. The technical assistance of A. S. Kushwaha is also acknowledged. This paper is Central Drug Research Institute communication no. 6805.

REFERENCES AND NOTES

(1) Sames, P. G.; Taylor, J. B. *Comprehensive Medicinal Chemistry*; Pergamon Press: Oxford, U. K., 1990; Vol. 4.

(2) Hansch, C.; Fujita, T. $p-\sigma-\pi$ Analysis: A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616−1626.

(3) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR: Hydrophobic, Electronic and Steric Constants*; American Chemical Society: Washington, DC, 1995.

(4) Cramer, R. D., III.; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.

(5) Klebe, G.; Abraham, U.; Mietzner, T. Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules to Correlate and Predict Their Biological Activity. *J. Med. Chem.* **1994**, *37*, 4130−4146.

(6) Geladi, P.; Kowalski, B. R. Partial Least-Squares Regression: A Tutorial. *Anal. Chim. Acta* **1986**, *185*, 1−17.

(7) (a) Green, S. M.; Marshall, G. R. 3D-QSAR: A Current Perspective. *Trends Pharmacol. Sci.* **1995**, *16*, 285−291. (b) Amin, E. A.; Welsh, W. J. Three-Dimensional Quantitative Structure−Activity Relationship (3D-QSAR) Models for a Novel Class of Piperazine-Based Stromelysin-1 (MMP-3) Inhibitors: Applying a "Divide and Conquer" Strategy. *J. Med. Chem.* **2001**, *44*, 3849−3855.

(8) Sufrin, J. R.; Dunn, D. A.; Marshall, G. R. Steric Mapping of the L-Methionine Binding Site of ATP: L-Methionine *S*-Adenosyltransferase. *Mol. Pharmacol.* **1981**, *19*, 307−313.

(9) Hasegawa, K.; Arakawa, M.; Funatsu, K. 3D-QSAR Study of Insecticidal Neonicotinoid Compounds Based on 3-Way Partial Least Squares Model. *Chemom. Intell. Lab. Syst.* **1999**, *47*, 33−40.

(10) Hasegawa, K.; Arakawa, M.; Funatsu, K. Rational Choice of Bioactive Conformations through Use of Conformation Analysis and 3-Way Partial Least Squares Modeling. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 253−261.

(11) Sippl, W. Receptor-Based 3D QSAR Analysis of Estrogen Receptor Ligands − Merging the Accuracy of Receptor-Based Alignments with the Computational Efficiency of Ligand-Based Methods. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 559−572.

(12) Liu, H.; Huang, X.; Shen, J.; Luo, X.; Li, M.; Xiong, B.; Chen, G.; Yang, Y.; Jiang, H.; Chen, K. Inhibitory Mode of 1,5-Diarylpyrazole Derivatives against Cyclooxygenase-2 and Cyclooxygenase-1: Molecular Docking and 3D QSAR Analyses. *J. Med. Chem.* **2002**, *45*, 4816−4827.

(13) Buolamwini, J. K.; Assefa, H. CoMFA and CoMSIA 3D QSAR and Docking Studies on Conformationally Restrained Cinnamoyl HIV-1 Integrase Inhibitors: Exploration of a Binding Mode at the Active Site. *J. Med. Chem.* **2002**, *45*, 841−852.

(14) Cui, M.; Huang, X.; Luo, X.; Briggs, J. M.; Ji, R.; Chen, K.; Shen, J.; Jiang, H. Molecular Docking and 3D-QSAR Studies on Gag Peptide Analogue Inhibitors Interacting with Human Cyclophilin A. *J. Med. Chem.* **2002**, *45*, 5249−5259.

(15) Huang, X.; Xu, L.; Luo, X.; Fan, K.; Ji, R.; Pei, G.; Chen, K.; Jiang, H. Elucidating the Inhibiting Mode of AHPBA Derivatives against HIV-1 Protease and Building Predictive 3D-QSAR Models. *J. Med. Chem.* **2002**, *45*, 333−343.

(16) Neel, B. G.; Tonks, N. K. Protein Tyrosine Phosphatases in Signal Transduction. *Curr. Opin. Cell Biol.* **1997**, *9*, 193−204.

(17) Denu, J. M.; Stuckey, J. A.; Saper, M. A.; Dixon, J. E. Form and Function in Protein Dephosphorylation. *Cell* **1996**, *87*, 361−364.

(18) Jia, Z. Protein Phosphatases: Structures and Implications. *Biochem. Cell Biol.* **1997**, *75*, 17−26.

(19) Zhang, Z. Y. Protein-Tyrosine Phosphatases: Biological Function, Structural Characteristics, and Mechanism of Catalysis. *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33*, 1−5.

(20) Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; Fitzhugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczky, J.; LeVine, R.; McEwan, P.; McKernan, K.; Meldrim, J.; Mesirov, J. P.; Miranda, C.; Morris, W.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, N.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman, R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer, S.; Milne, S.; Mullikin, J. C.; Mungall, A.; Plumb, R.; Ross, M.; Shownkeen, R.; Sims, S.; Waterson, R. H.; Wilson, R. K.; Hillier, L. W.; McPherson, J. D.; Marra, M. A.; Mardis, E. R.; Fulton, L. A.; Chinwalla, A. T.; Pepin, K. H.; Gish, W. R.; Chissoe, S. L.; Wendl, M. C.; Delehaunty, K. D.; Miner, T. L.; Delehaunty, A.; Kramer, J. B.; Cook, L. L.; Fulton, R. S.; Johnson, D. L.; Minx, P. J.; Clifton, S. W.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.; Doggett, N.; Cheng, J. F.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E.; Frazier, M.; Gibbs, R. A.; Muzny, D. M.; Scherer, S. E.; Bouck, J. B.; Sodergren, E. J.; Worley, K. C.; Rives, C. M.; Gorrell, J. H.; Metzker, M. L.; Naylor, S. L.; Kucherlapati, R. S.; Nelson, D. L.; Weinstock, G. M.; Sakaki, Y.; Fujiyama, A.; Hattori, M.; Yada, T.; Toyoda, A.; Itoh, T.; Kawagoe, C.; Watanabe, H.; Totoki, Y.; Taylor, T.; Weissenbach, J.; Heilig, R.; Saurin, W.; Artiguenave, F.; Brottier, P.; Bruls, T.; Pelletier, E.; Robert, C.; Wincker, P.; Smith, D. R.; Doucette-Stamm, L.;

Rubenfield, M.; Weinstock, K.; Lee, H. M.; Dubois, J.; Rosenthal, A.; Platzer, M.; Nyakatura, G.; Taudien, S.; Rump, A.; Yang, H.; Yu, J.; Wang, J.; Huang, G.; Gu, J.; Hood, L.; Rowen, L.; Madan, A.; Qin, S.; Davis, R. W.; Federspiel, N. A.; Abola, A. P.; Proctor, M. J.; Myers, R. M.; Schmutz, J.; Dickson, M.; Grimwood, J.; Cox, D. R.; Olson, M. V.; Kaul, R.; Raymond, C.; Shimizu, N.; Kawasaki, K.; Minoshima, S.; Evans, G. A.; Athanasiou, M.; Schultz, R.; Roe, B. A.; Chen, F.; Pan, H.; Ramser, J.; Lehrach, H.; Reinhardt, R.; McCombie, W. R.; De la Bastide, M.; Dedhia, N.; Blocker, H.; Hornischer, K.; Nordsiek, G.; Agarwala, R.; Aravind, L.; Bailey, J. A.; Bateman, A.; Batzoglou, S.; Birney, E.; Bork, P.; Brown, D. G.; Burge, C. B.; Cerutti, L.; Chen, H. C.; Church, D.; Clamp, M.; Copley, R. R.; Doerks, T.; Eddy, S. R.; Eichler, E. E.; Furey, T. S.; Galagan, J.; Gilbert, J. G.; Harmon, C.; Hayashizaki, Y.; Haussler, D.; Hermjakob, H.; Hokamp, K.; Jang, W.; Johnson, L. S.; Jones, T. A.; Kasif, S.; Kaspryzk, A.; Kennedy, S.; Kent, W. J.; Kitts, P.; Koonin, E. V.; Korf, I.; Kulp, D.; Lancet, D.; Lowe, T. M.; McLysaght, A.; Mikkelsen, T.; Moran, J. V.; Mulder, N.; Pollara, V. J.; Ponting, C. P.; Schuler, G.; Schultz, J.; Slater, G.; Smit, A. F.; Stupka, E.; Szustakowski, J.; Thierry-Mieg, D.; Thierry-Mieg, J.; Wagner, L.; Wallis, J.; Wheeler, R.; Williams, A.; Wolf, Y. I.; Wolfe, K. H.; Yang, S. P.; Yeh, R. F.; Collins, F.; Guyer, M. S.; Peterson, J.; Felsenfeld, A.; Wetterstrand, K. A.; Patrinos, A.; Morgan, M. J.; De Jong, P.; Catanese, J. J.; Osoegawa, K.; Shizuya, H.; Choi, S.; Chen, Y. J. Initial Sequencing and Analysis of the Human Genome. *Nature* **2001**, *409*, 860−921.

(21) Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J.; Gabor Miklos, G. L.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, J.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Di Francesco, V.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R. R.; Ke, Z.; Ketchum, K. A.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z.; Wang, A.; Wang, X.; Wang, J.; Wei, M.; Wides, R.; Xiao, C.; Yan, C.; Yao, A.; Ye, J.; Zhan, M.; Zhang, W.; Zhang, H.; Zhao, Q.; Zheng, L.; Zhong, F.; Zhong, W.; Zhu, S.; Zhao, S.; Gilbert, D.; Baumhueter, S.; Spier, G.; Carter, C.; Cravchik, A.; Woodage, T.; Ali, F.; An, H.; Awe, A.; Baldwin, D.; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, M. L.; Curry, L.; Danaher, S.; Davenport, L.; Desilets, R.; Dietz, S.; Dodson, K.; Doup, L.; Ferriera, S.; Garg, N.; Gluecksmann, A.; Hart, B.; Haynes, J.; Haynes, C.; Heiner, C.; Hladun, S.; Hostin, D.; Houck, J.; Howland, T.; Ibegwam, C.; Johnson, J.; Kalush, F.; Kline, L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McIntosh, T.; McMullen, I.; Moy, M.; Moy, L.; Murphy, B.; Nelson, K.; Pfannkoch, C.; Pratts, E.; Puri, V.; Qureshi, H.; Reardon, M.; Rodriguez, R.; Rogers, Y. H.; Romblad, D.; Ruhfel, B.; Scott, R.; Sitter, C.; Smallwood, M.; Stewart, E.; Strong, R.; Suh, E.; Thomas, R.; Tint, N. N.; Tse, S.; Vech, C.; Wang, G.; Wetter, J.; Williams, S.; Williams, M.; Windsor, S.; Winn-Deen, E.; Wolfe, K.; Zaveri, J.; Zaveri, K.; Abril, J. F.; Guigo, R.; Campbell, M. J.; Sjolander, K. V.; Karlak, B.; Kejariwal, A.; Mi, H.; Lazareva, B.; Hatton, T.; Narechania, A.; Diemer, K.; Muruganujan, A.; Guo, N.; Sato, S.; Bafna, V.; Istrail, S.; Lippert, R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J.; Blick, L.; Caminha, M.; Carnes-Stine, J.; Caulk, P.; Chiang, Y. H.; Coyne, M.; Dahlke, C.; Mays, A.; Dombroski, M.; Donnelly, M.; Ely, D.; Esparham, S.; Fosler, C.; Gire, H.; Glanowski, S.; Glasser, K.; Glodek, A.; Gorokhov, M.; Graham, K.; Gropman, B.; Harris, M.; Heil, J.; Henderson, S.; Hoover, J.; Jennings, D.; Jordan, C.; Jordan, J.; Kasha, J.; Kagan, L.; Kraft, C.; Levitsky, A.; Lewis, M.; Liu, X.; Lopez, J.; Ma, D.; Majoros, W.; McDaniel, J.; Murphy, S.; Newman, M.; Nguyen, T.; Nguyen, N.; Nodell, M.; Pan, S.; Peck, J.; Peterson, M.; Rowe, W.; Sanders, R.; Scott, J.; Simpson, M.; Smith, T.; Sprague, A.; Stockwell, T.; Turner, R.; Venter, E.; Wang, M.; Wen, M.; Wu, D.; Wu, M.; Xia, A.; Zandieh, A.; Zhu, X. The Sequence of the Human Genome. *Science* **2001**, *291*, 1304−1351.

(22) Evans, J. L.; Jallal, B. Protein Tyrosine Phosphatases: Their Role in Insulin Action and Potential as Drug Targets. *Expert Opin. Invest. Drugs* **1999**, *8*, 139−160.

(23) Burke, T. R.; Zhang, Z. Y., Jr. Protein-Tyrosine Phosphatases: Structure, Mechanism, and Inhibitor Discovery. *Biopolymers* **1998**, *47*, 225−241.

(24) Ahn, N. Introduction: Protein Phosphorylation and Signaling. *Chem. Rev.* **2001**, *101*, 2207−2208.

(25) Bandyopadhyay, D.; Kusari, A.; Kenner, K. A.; Liu, F.; Chernoff, J.; Gustafson, T. A.; Kusari, J. Protein-Tyrosine Phosphatase 1B Complexes with the Insulin Receptor in Vivo and is Tyrosine-Phosphorylated in the Presence of Insulin. *J. Biol. Chem.* **1997**, *272*, 1639−1645.

(26) Wang, X. Y.; Bergdahl, K.; Heijbel, A.; Liljebris, C.; Bleasdale, J. E. Analysis of in Vitro Interactions of Protein Tyrosine Phosphatase 1B with Insulin Receptors. *Mol. Cell. Endocrinol.* **2001**, *173*, 109−119.

(27) Jacob, K. K.; Sap, J.; Stanley, F. M. Receptor-Like Protein-Tyrosine Phosphatase α Specifically Inhibits Insulin-Increased Prolactin Gene Expression. *J. Biol. Chem.* **1998**, *273*, 4800−4809.

(28) Goldstein, B. J. Protein-Tyrosine Phosphatases and the Regulation of Insulin Action. *J. Cell. Biochem.* **1992**, *48*, 33−42.

(29) Elchebly, M.; Payette, P.; Michaliszyn, E.; Cromlish, W.; Collins, S.; Loy, A. L.; Normandin, D.; Cheng, A.; Himms-Hagen, J.; Chan, C. C.; Ramachandran, C.; Gresser, M. J.; Tremblay, M. L.; Kennedy, B. P. Increased Insulin Sensitivity and Obesity Resistance in Mice Lacking the Protein Tyrosine Phosphatase-1B Gene. *Science* **1999**, *283*, 1544−1548.

(30) Klaman, L. D.; Boss, O.; Peroni, O. D.; Kim, J. K.; Martino, J. L.; Zabolotny, J. M.; Moghal, N.; Lubkin, M.; Kim, Y. B.; Sharpe, A. H.; Stricker-Krongrad, A.; Shulman, G. I.; Neel, B. G.; Kahn, B. B. Increased Energy Expenditure, Decreased Adiposity, and Tissue-Specific Insulin Sensitivity in Protein-Tyrosine Phosphatase 1B-Deficient Mice. *Mol. Cell Biol.* **2000**, *20*, 5479−5489.

(31) Murthy, V. S.; Kulkarni, V. M. 3D-QSAR CoMFA and CoMSIA on Protein Tyrosine Phosphatase 1B Inhibitors. *Bio. Med. Chem.* **2002**, *10* (7), 2267−2282.

(32) Hu, X.; Stebbins, C. E. Molecular Docking and 3D-QSAR Studies of *Yersinia* Protein Tyrosine Phosphatase YopH Inhibitors. *Bio. Med. Chem.* **2005**, *13* (4), 1101−1109.

(33) Zhou, M.; Ji, M. Molecular Docking and 3D-QSAR on 2-(Oxalylamino) Benzoic Acid and Its Analogues as Protein Tyrosine Phosphatase 1B Inhibitors. *Bio. Med. Chem. Lett.* **2005**, *15* (24), 5521−5525.

(34) Larsen, S. D.; Barf, T.; Liljebris, C.; May, P. D.; Ogg, D.; O'Sullivan, T. J.; Palazuk, B. J.; Schostarez, H. J.; Stevens, F. C. Synthesis and Biological Activity of a Novel Class of Small Molecular Weight Peptidomimetic Competitive Inhibitors of PTP1B. *J. Med. Chem.* **2002**, *45*, 598−622.

(35) Desai, B.; Sureja, D.; Naliapara, Y.; Shah, A.; Saxena, A. K. Synthesis and QSAR Studies of 4-Substituted Phenyl-2,6-dimethyl-3,5-bis-*N*-(substituted phenyl) Carbamoyl-1,4-dihydropyridines as Potential Antitubercular Agents. *Bio. Med. Chem.* **2001**, *9* (8), 1993−1998.

(36) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(37) Bohacek, R. S.; McMartin, C. Definition and Display of Steric, Hydrophobic, and Hydrogen Bonding Properties of Ligand Binding Sites in Proteins using Lee and Richards Accessible Surface: Validation of a High-Resolution Graphical Tool for Drug Design. *J. Med. Chem.* **1992**, *35*, 1671−1684.

(38) Kroemer, R. T.; Hecht, P. Replacement of Steric 6−12 Potential-Derived Interaction Energies by Atom-Based Indicator Variables in CoMFA Leads to Models of Higher Consistency. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 205−212.

(39) Bush, B. L.; Nachbar, R. B., Jr. Sample-Distance Partial Least Squares: PLS Optimized for Many Variables, with Application to CoMFA. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587−619.

(40) Böhm, M.; Stürzebecher, J.; Klebe, G. Three-Dimensional Quantitative Structure−Activity Relationship Analyses Using Comparative Molecular Field Analysis and Comparative Molecular Similarity Indices Analysis to Elucidate Selectivity Differences of Inhibitors Binding to Trypsin, Thrombin, and Factor Xa. *J. Med. Chem.* **1999**, *42*, 458−477.

(41) Dixit, A.; Kashaw, S. K.; Gaur, S.; Saxena, A. K. Development of CoMFA, Advance CoMFA and CoMSIA Models in Pyrroloquinazolines as Thrombin Receptor Antagonist. *Bioorg. Med. Chem*. **2004**, *12*, 63−69.

(42) Peters, G. H.; Frimurer, T. N.; Andersen, J. N.; Olsen, O. H. Molecular Dynamic Simulations of Protein Tyrosine Phosphatase 1B.II Substrate Enzyme Interactions and Dynamics. *Biophys. J.* **2000**, *78*, 2191−2200.

(43) Sarmiento, M.; Puius, Y. A.; Vetter, S. W.; Keng, Y. F.; Wu, L.; Zhao, Y.; Lawrence, D. S.; Almo, S. C.; Zhang, Z. Y. Structural Basis of Plasticity in Protein Tyrosine Phosphatase 1B Substrate Recognition Biochemistry. *Biochemistry* **2000**, *39*, 8171−8179.

(44) Jia, Z.; Barford, D.; Flint, A. J.; Tonks, N. K. Structural Basis of Phosphotyrosine Peptide Recognition by PTP1B. *Science* **1995**, *268*, 1754−1758.

(45) Sekar, N.; Li, J.; Shechter, Y. Vanadium Salts as Insulin Substitutes: Mechanisms of Action, a Scientific and Therapeutic Tool in Diabetes Mellitus Research. *Crit. Rev. Biochem. Mol. Biol.* **1996**, *31*, 339−359.
(46) Goldstein, B. J.; Bittner-Kowalczyk, A.; White, M. F.; Harbeck, M. Tyrosine Dephosphorylation and Deactivation of Insulin Receptor Substrate-1 by Protein-Tyrosine Phosphatase 1B. Possible Facilitation by the Formation of a Ternary Complex with the GRB2 Adaptor Protein. *J. Biol. Chem.* **2000**, *275*, 4283−4289.

CI600224N