Sulphonamides as Inhibitors of Protein Tyrosine Phosphatase 1B: A Three-Dimensional Quantitative Structure–Activity Relationship Study Using Self-Organizing Molecular Field Analysis Approach

Suresh THAREJA, Ganesh Rajendra KOKIL, Saurabh AGGARWAL, Tilak Raj BHARDWAJ, and Manoj KUMAR*

University Institute of Pharmaceutical Sciences, Panjab University; Chandigarh–160014, India. Received November 26, 2009; accepted December 26, 2009; published online January 7, 2010

> **Protein tyrosine phosphatase 1B (PTP 1B), a cytosolic PTP involved in down-regulation of receptor tyrosine kinase activity following stimulation of the insulin or leptin receptors. Thus, PTP 1B inhibitors could potentially ameliorate insulin resistance and normalize plasma glucose and insulin levels without inducing hypoglycemia, and could therefore be a major advancement in the treatment of type 2 diabetes. A three-dimensional quantitative structure–activity relationship (3D-QSAR) study has been performed on a novel class of sulphonamides using self-organizing molecular field analysis (SOMFA) to correlate their chemical structures with their observed PTP 1B inhibitory activities. The master grid obtained for the various SOMFA models indicates electrostatic and shape potential contributions that can be mapped back onto structural features relating to the trends in inhibitory activities. On the basis of the spatial arrangement, steric and electrostatic factors should appropriately be taken into account for development of new potent inhibitors of PTP 1B for the management of type 2 diabetes.**

> **Key words** diabetes; insulin; protein tyrosine phosphatase 1B; three-dimensional quantitative structure–activity relationship; self-organizing molecular field analysis; sulphonamide

Diabetes mellitus is a chronic multifactorial metabolic disease resulting from insulin deficiency or insulin resistance. Diabetes is a life-long disease and there is no permanent cure. Diabetes mellitus, in the 21st century is considered to be the main threat to human health. The incidences of this disease are increasing day by day and are estimated to reach 210 million by the year 2010 and 300 million by the year 2025.^{1,2)} Protein tyrosine phosphatase 1B (PTP 1B) is a ubiquitously expressed intracellular enzyme which causes negative regulation of insulin receptor as well as leptin signaling system emerged as a potential target for treatment of type 2 diabetes.^{3,4)} It has been involved in down-regulation of receptor tyrosine kinase activity following stimulation of the insulin or leptin receptors.^{5,6)} Recent studies on PTP 1B knockout mice provided significant support for the view that PTP 1B is a key regulator of insulin signaling. PTP 1B deficient mice showed increased insulin sensitivity and obesity resistance, demonstrating that PTP 1B plays a major role in modulating both insulin sensitivity and fuel metabolism.^{7,8)} PTP 1B inhibitors could potentially ameliorate insulin resistance and normalize plasma glucose and insulin levels without inducing hypoglycemia, and could therefore be a major advancement in the treatment of type 2 diabetes. $9,10)$ Thus, PTP 1B is highly coveted by the pharmaceutical industry and makes it an ideal drug target for therapeutic intervention in common human diseases such as type 2 diabetes and obesity.¹¹⁾

Quantitative structure–activity relationship (QSAR) represents an approach to correlate structural descriptors of compounds with their biological activities. Three-dimensional quantitative structure activity–relationship (3D-QSAR) studies provide deeper insight into the mechanism of action of compounds that ultimately becomes of great importance in modification of the structure of compounds. In addition, 3D-QSAR also provides quantitative models, which permits pre-

diction of activity of compounds prior to the synthesis. Selforganizing molecular field analysis (SOMFA) is a novel 3D-QSAR methodology which has been developed by Robinson *et al.*12) It is a simple and intuitive in concept and avoids the complex statistical tools and variable selection procedures favored by other methods. The method has similarities to both comparative molecular field analysis (CoMFA) and molecular similarity studies.¹³⁾ Like CoMFA, a grid-based approach is used; however, no probe interaction energies need to be calculated. Like the similarity methods it is the intrinsic molecular properties, such as the molecular shape and electrostatic potential, which are used to develop QSAR models.¹⁴⁾

A SOMFA model could be based on any molecular property. In the present study we have used molecular shape and electrostatic potentials. A successful 3D-QSAR model not only helps in better understanding of the structure–activity relationship of any class of compounds, but also provides researcher an insight at molecular level about lead compounds for further developments. The inherent simplicity of this method allows the possibility of aligning the training compounds as an integral part of the model derivation process and of aligning prediction compounds to optimize their predicted activities.¹⁵⁾

Recent studies in our laboratory have focused on refining the molecular architecture using 3D-QSAR SOMFA approach for designing and optimization of new inhibitors for various targets.^{$16,17)$} The important aim of present studies is to correlate the 3D-structures of sulphonamide derivatives with their biological activities and to be able to predict the activity of new molecules prior to their synthesis with the hope that these molecules may be further be explored as potent anti-diabetic agents.

Computational Methods

Data Set and Biological Activities A dataset of 30 molecules belong-

Table 1. Structure of Sulphonamide Derivatives Used for the SOMFA Study

 \mathbf{R}

T: test set molecules.

ing to sulphonamide derivatives as PTP 1B inhibitors was taken from the literature and used for SOMFA analysis.¹⁸⁾ The above reported series of sulphonamide derivatives showed wide variations in their structures and potency profiles. The negative logarithm of the measured IC₅₀ (μ M) against PTP 1B enzyme as pIC_{50} (pIC₅₀ or log 1/IC₅₀) was used as dependent variable,19) thus correlating the data linear to the free energy change. Since some compounds exhibited insignificant/no inhibition, they were excluded from the present study. SOMFA (3D-QSAR) models were generated for this series using a training set of 22 molecules. The general structures of the training set and test molecules are presented in Table 1. Predictive power of the resulting models was evaluated by a test set of 8 molecules with uniformly distributed biological activities. The actual and predicted biological activities of the test set molecules are presented in Table 3. Selections of test set molecules was made by considering the fact that test set molecules represent structural features similar to compounds in the training set.²⁰⁾ Thus the test set compounds are true representative of the training set.

Molecular Modeling and Alignment The three-dimensional structures of the sulphonamide derivatives were constructed with the Chemdraw Ultra 8.0 running on an Intel Pentium IV 2.80 GHz Processor/Microsoft Win XP professional platform and were subjected to energy minimization using molecular mechanics (MM2). The minimization is continued until the root mean square (RMS) gradient value reaches a value smaller than 0.001 kcal/mol Å. The Hamiltonian approximations Austin model 1 (AM1) method²¹⁾ available in the MOPAC module²²⁾ of Chem3D is adopted for re-optimization until the RMS gradient attains a value smaller than 0.001 kcal/mol Å. Unless otherwise indicated, all parameters were kept default. The final active conformation search was performed by docking all the compounds into the active site using Molegro Virtual Docker (MVD) software.^{23,24)} The protein data bank (PDB) entry of PTP 1B enzyme used in docking experiments is 1XBO. Docked structure of most potent compound (**28**) in the active site of PTP 1B has been shown in Fig. 1.

Structural alignment is one of the most sensitive parameters in 3D-QSAR analyses. The accuracy of the prediction of a SOMFA model and the reliability of the SOMFA grids strongly depend on the structural alignment of the molecules. The selected template molecule is typically one of the following: (a) the most active compound; (b) the lead and/or commercial compound; (c) the compound containing the greatest number of functional groups.^{25,26)} Generally, the low energy conformation of the most active compound is set as a reference.²⁷⁾ In present study the compounds were aligned using docked conformation of the most active compound (**28**) used as the reference compound by atom based alignment technique where centroid of atoms were used for alignment shown in Table 2. The best model was obtained using alignment 1 where docked confirmation of the most active compound (**28**) was used as template structure. The superimposition of molecules was based on trying to minimize RMS differences in the fitting of selected atoms with those of a template molecule (Fig. 2).

SOMFA 3D-QSAR Models In the SOMFA study, a $40\times40\times40$ Å grid originating at $(-20, -20, -20)$ with a resolution of 0.5 and 1 Å respectively, was generated around the aligned compounds.^{28—31)} Table 4 reports four different models using different resolutions of grid under exploration using two atom based alignment techniques (Table 2).

For all of the studies, shape and electrostatic potential were generated. The partial least squares (PLS) algorithm^{32,33)} was used in conjugation with leave one out (LOO) cross-validation to develop final model. Partial least

Fig. 1. Docked Conformation of Most Active Compound (**28**) with PTP 1B Showing H-Bonding Interactions with the Active Site of the Enzymes Fig. 2. Superimposition of Compounds on Template Structure (**28**)

squares (PLS), the statistical method used in deriving the 3D-QSAR models, is an extension of multiple regression analysis in which the original variables are replaced by a small set of their linear combinations. These latent variables (components) so generated are used for multivariate regression, maximizing the commonality of explanatory and response variable blocks.

The cross-validated value $r_{\text{cv}}^2(q^2)$ can take up values in the range from 1, suggesting a perfect model, to less than 0 where errors of prediction are greater than the error from assigning each compound mean activity of the model.³⁴⁾ Since the final equations are not very useful to represent efficiently the SOMFA models, 3D master grid maps of the best are displayed by program Grid-Visualizer. They represent area in space where steric and electrostatic field interactions are responsible for the observed variations in the biological activities.

Results and Discussion

In the present 3D-QSAR study, SOMFA, a novel methodology, was employed for the analysis with the training set composed of 22 compounds whose inhibitory activities are known in order to find out the molecular features responsible for biological activities. Statistical results of SOMFA models obtained by PLS analysis *i.e.* cross-validated value $r_{\text{cv}}^2(q^2)$ value serves as a quantitative measure of the predictability of the SOMFA model.³⁵⁾

During the SOMFA investigation, grid spacings of 1 and 0.5 Å were investigated. The 1 Å grid spacing produces a good correlation equal to 0.5 Å grids. This has been decreased marginally with the 1 Å spacing used for the results presented (Table 4). Further increases in resolution have

Table 3. Actual and Predicted Activities for Training and Test Set Molecules from the Best Predictive SOMFA Model I

Compound	Actual activity	Predicted	Residual
	(pIC_{50})	activity	activity
$\mathbf{1}$	-1.079	-0.97	-0.109
2^T	-0.845	-0.456	-0.389
3	-0.623	-0.544	-0.079
4^T	-0.222	-0.259	0.037
5	-0.477	-0.582	0.105
6^T	-0.477	-0.48	0.003
7	-0.602	-0.584	-0.018
8	0.000	-0.005	0.005
9	-0.602	-0.676	0.074
10	-0.699	-0.357	-0.342
11^T	-0.602	-0.58	-0.022
12	-0.278	-0.576	0.298
13	-0.462	-0.266	-0.196
14	-0.204	0.216	-0.420
15	0.114	-0.026	0.140
16	0.200	0.368	-0.168
17	-0.041	0.752	-0.793
18^T	0.220	0.372	-0.152
19	0.000	-0.111	0.111
20	0.745	-0.107	0.852
21	1.130	0.364	0.766
22^T	0.657	0.225	0.432
23	-0.954	0.324	-1.278
24	-0.833	-0.039	-0.794
25	0.677	0.023	0.654
26^T	1.455	0.913	0.542
27	1.222	0.534	0.688
28	1.553	1.049	0.504
29^T	-0.041	0.614	-0.655
30	1.507	1.508	-0.001

T: test set molecules.

produced further small increases in model quality but not enough to warrant the extra computational time. From Table 4, we find that the results were less sensitive to resolution of grid. The best model using alignment 1 at 0.5 Å resolution shows good $r_{\text{cv}}^2(q^2)$ value than using other alignment. Good cross-validated correlation coefficient $r_{\text{cv}}^2(q^2)$ value (0.751), moderate non cross-validated correlation coefficient *r* ² values (0.797), high *F*-test value (78.387) and low standard error of estimation *S* (0.370) proves a good conventional statistical correlation which have been obtained, and we also found that the resultant SOMFA model have a satisfied predictive ability $r_{\text{pred}}^2 (0.616)$.

The actual and predicted activities of the training set are reported in Table 3 using model I. Figures 3A and 4 showed a good linear correlation and moderate difference between observed and predicted values of molecules in the training set. It is well known that the best way to validate a 3D-QSAR model is to predict biological activities for some compounds of test set. The SOMFA analysis of the test set composed of 8 compounds is reported in Table 4. Most of compounds in test set show good correlation between observed and predicted values (Figs. 3B, 5).

SOMFA calculation for both shape and electrostatic potentials were performed. The contribution of shape field and electrostatic field to QSAR equation is 66% and 34%, respectively. SOMFA analysis indicated that the electrostatic contribution is of a slightly low importance while shape contribution is of major importance (66%). The SOMFA electrostatic potential and shape for the analysis is presented as

Fig. 3. Graph of Actual *vs.* Predicted Activities for Training and Test Set Molecules from the Best Predictive SOMFA Model (A) Training set, (B) test set.

Fig. 4. Histogram of SOMFA Residual Value for Training Set

Table 4. PLS Statistical Results of SOMFA

Parameter	Alignment 1 Resolutions		Alignment 2 Resolutions	
	0.5 Å (Model I)	1 Å (Model II)	0.5 Å (Model III)	1 Å (Model IV)
r^2	0.797	0.7922	0.725	0.721
$r_{\rm cv}^2(q^2)$	0.751	0.7448	0.678	0.674
S	0.370	0.374	0.430	0.433
F	78.387	76.253	52.656	51.702
prediction	0.616	0.617	0.547	0.542
<i>prediction</i>	0.466	0.465	0.505	0.508

 r_{cv}^2 : cross-validated correlation coefficient by leave one out method; r^2 : conventional correlation coefficient; *S*: standard error of estimate; *F*: Fisher test value; $r_{\text{prediction}}^2$: predictive correlation coefficient; $S_{\text{prediction}}$: standard error of prediction.

master grid.

The master grid maps derived from the best model were used to display the contribution of electrostatic potential and shape molecular field. The master grid maps gave a direct vi-

Fig. 5. Histogram of SOMFA Residual Value for Test Set

sual indication of which parts of the compounds differentiate the activities of compounds in the training set under study. The master grid also offered an interpretation as to how to design some novel compounds with much higher activities. The visualization of the electrostatic potential master grid and shape master grid of the best SOMFA model were shown in Figs. 6 and 7, respectively, with most active compound (**28**) as the reference. Each master grid map was colored in two different colors for favorable and unfavorable effects. In other words, the electrostatic features were red (more positive charge increases activity, or more negative charge decreases activity) and blue (more negative charge increases activity, or more positive charge decreases activity), and the shape feature are red (more steric bulk increases activity) and blue (more steric bulk decreases activity), respectively.

The SOMFA electrostatic potential map shows some important features, we find a high density of blue points around substituent R_2 , R_6 of the sulphonamide derivatives, which indicates some electronegative groups are favorable while around R_1 , R_3 , R_4 and R_5 red points indicating some electropositive groups are favorable. Meanwhile, in the map of shape master grid, we can find a high density of red points around R_3 , R_6 which means a favorable steric interaction; simultaneously, we also find blue points around R_1, R_2, R_5 and $R₇$ where an unfavorable steric interaction may be expected to enhance activities. The SOMFA results obtained above can be correlated with interactions through most active compound (28) in which R_1 and R_3 having electropositive –OH, $-N(CH_3)_2SO_2$ group interacts through H-bond with Gln 262, Lys 120 while R_6 having electronegative –COOH group interacting with Ala 217 of the active site of PTP 1B.

Conclusion

We have developed predictive SOMFA 3D-QSAR model for sulphonamide derivatives as inhibitors of PTP 1B based on docked conformer based alignment and it exhibited statistically significant predictability. The master grid obtained from the SOMFA models indicates electrostatic and shape potential contributions that can be mapped back onto structural features relating to the trends in activities of the molecules. The present SOMFA study investigates the indispensable molecular features of sulphonamide derivatives which can be exploited for further structural modifications of these lead molecules in order to achieve improved PTP 1B inhibitory activity for the management of type 2 diabetes.

Acknowledgements We gratefully acknowledge Indian Council of Medical Research (ICMR), New Delhi for Senior Research Fellowship

Fig. 6. SOMFA Derived Electrostatic Grids Showing Most Active Compound (**28**) in the Background

Blue and red indicates region where more electronegative groups or electropositive groups, respectively, will enhance the activity at different resolutions (A) 0.5 Å , (B) 1 Å.

 (A)

Fig. 7. SOMFA Derived Shape Grids Showing Most Active Compound (**28**) Is Displayed in the Background

Red and blue indicates region where more steric bulk or less steric bulk, respectively, will enhance the activity at different resolutions (A) 0.5 Å, (B) 1 Å.

(SRF) to Suresh Thareja. We are also thankful to Dr. Daniel Robinson (Computational Chemistry Research Group, Oxford University, U.K.) for the use of the SOMFA software and Dr. Rene Thomson for use of Molegro Virtual Docker.

References and Notes

- 1) Vats R. K., Kumar V., Kothari A., Mital A., Ramachandaran U., *Cur. Sci.*, **88**, 241—249 (2005).
- 2) Defranzo R. A., *Diabetes*, **37**, 667—687 (1988).
- 3) Forsell P. A. L., Boie Y., Montalibet J., Collins S., Kennedy B. P., *Gene*, **260**, 145—153 (2000).
- 4) Lund I. K., Bilestup N., *J. Mol. End.*, **15**, 339—351 (2005).
- 5) Byon J. C., Kusari J., Kusari A. B., *Mol. Cell. Biochem.*, **182**, 101— 108 (1998).
- 6) Kennedy B. P., Ramachandran C., *Biochem. Pharmacol.*, **60**, 877— 883 (2000).
- 7) 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., *Science*, **283**, 1544—1548 (1999).
- 8) 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., *Mol. Cell. Biol.*, **20**, 5479—5489 (2000).
- 9) Zhang Z. Y., Lee S. Y., *Expert Opin. Investing. Drugs*, **12**, 223—233 (2003).
- 10) Sachan N., Thareja S., Aggarwal R., Kadam S. S., Kulkarni V. M., *Int. J. PharmTech Res.*, **1**, 1625—1631 (2009).
- 11) Tonks N. K., *FEBS Lett.*, **546**, 140—148 (2003).
- 12) Robinson D. D., Winn P. J., Lyne P. D., Richards W. G., *J. Med. Chem.*, **42**, 573—583 (1999).
- 13) Cramer R. D., Patteerson D. E., Bunce J. D., *J. Am. Chem. Soc.*, **110**, 5959—5967 (1988).
- 14) Li S., Zheng Y., *Int. J. Mol. Sci.*, **7**, 220—229 (2006).
- 15) Li M., Du L., Wu B., Xia L., *Bioorg. Med. Chem.*, **11**, 3945—3951 (2003).
- 16) Thareja S., Aggarwal S., Bhardwaj T. R., Kumar M., *Eur. J. Med.*

Chem., **44**, 4920—4925 (2009).

- 17) Aggarwal S., Thareja S., Bhardwaj T. R., Kumar M., *Eur. J. Med. Chem.*, **45**, 476—481 (2010).
- 18) Holmes C. P., Li X., Pan Y., Xu C., Bhandari A., Moody C. M., Steven J. A., Ferla W., Defrancisco M. N., Frederick B. T., Zhou S., Macher N., Jang L., Irvine J. D., Grove J. R., *Bioorg. Med. Chem. Lett.*, **15**, 4336—4341 (2005).
- 19) Puntambekar D., Giridhar R., Yadav M. R., *Bioorg. Med. Chem. Lett.*, **16**, 1821—1827 (2006).
- 20) Murthy V. S., Kulkarni V. M., *Bioorg. Med. Chem.*, **10**, 2267—2282 (2002).
- 21) Dewar M. J. S., Zoebisch E. G., Healy E. F., Stewart J. J. P., *J. Am. Chem. Soc.*, **107**, 3902—3909 (1985).
- 22) Stewart J., *J. Comput. Aid. Mol. Des.*, **4**, 1—45 (1990).
- 23) Thomsen R., Christensen M. H., *J. Med. Chem.*, **49**, 3315—3321 (2006).
- 24) Molegro Virtual Docker 3.2.0 http://www.molegro.com/mvdproduct.php
- 25) Baurin N., Vangrevelinghe E., Allory L. M., *J. Med. Chem.*, **43**, 1109—1122 (2000).
- 26) Agarwal A., Taylor E. W., *J. Computat. Chem.*, **14**, 237—245 (1993).
- 27) Xu M., Zhang A., Han S., Wang L., *Chemosphere*, **48**, 707—715 (2002)
- 28) Pedretti A., Villa L., Vistoli G., *J. Mol. Graph. Model*, **21**, 47—49 (2002).
- 29) Thareja S., Aggarwal S., Bhardwaj T. R., Kumar M., *Med. Chem.*, **6**, (2010).
- 30) VEGA ZZ Release 2.0.5.52: http://www.ddl.unimi.it/vega/index2.htm
- 31) SOMFA2 v2.0.0: http://bellatrix.pcl.ox.ac.uk. (2006)
- 32) Li M. Y., Fang H., Xia L. *Bioorg. Med. Chem. Lett.*, **15**, 3216—3219 (2005)
- 33) Kulkarni S. S., Gediya L. K., Kulkarni V. M., *Bioorg. Med. Chem. Lett.*, **7**, 1475—1485 (1999).
- 34) Golbraikh A., Tropsha A., *J. Mol. Graph. Model.*, **20**, 269—276 (2002).
- 35) Puntambekar D. S., Giridhar R., Yadav M. R., *Eur. J. Med. Chem.*, **43**, 142—154 (2008).