Acta Poloniae Pharmaceutica – Drug Research, Vol. 69 No. 3 pp. 535–543, 2012 ISSN 0001-6837
Polish Pharmaceutical Society

CHEMINFORMATICS BASED 3D-QSAR STUDY ON A SERIES OF 1,2-NAPHTHOQUINONE DERIVATIVES AS PTP 1B INHIBITORS

PAYAL PAHWA[1,2]* and MANJU PAPREJA[2,3]

[1]Bhagwan Parshuram Institute of Technology, Rohini (New Delhi)-110085, India
[2]National Institute of Medical Sciences, Jaipur (Rajasthan)-303121, India
[3]GVM Institute of Technology and Management, Sonipat (Haryana)-131001, India

Abstract: Self-organizing molecular field analysis (SOMFA), a simple three-dimensional quantitative structure activity relationship (3D-QSAR) based cheminformatics method was used in present case to study the correlation between the molecular properties and the PTP 1B inhibitory activities of a series of 1,2-naphthoquinone that acts as selective PTP 1B inhibitors. The statistical results, cross-validated r^2_{cv} and non cross-validated r^2, F-test value showed a satisfied predictive ability (r^2_{pred}). The spatial arrangement of the shape and electrostatic potential could be used as a guide for further development of selective and more potent PTP 1B inhibitors.

Keywords: cheminformatics, PTP 1B, 3D-QSAR, SOMFA

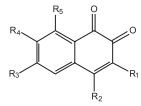
Cheminformatics is a scientific discipline that has evolved in the last 40 years at the interface between chemistry and computer science (1). Chemical informatics is the application of computer technology to chemistry in all of its manifestations (2). It is the cross between computer science and chemistry concerned with storing, retrieving and searching information and with storing relationship between bits of data (3). Three-dimensional quantitative structure-activity relations (3D-QSAR) have a profound impact on medicinal chemistry. QSAR provides the guidelines for making structural changes in the compound which generates statistically significant relationships between chemical structure and biological activity (4, 5). The basic principle is that the variations of biological activity within a series can be correlated with changes in measured or computed molecular features of the molecules. A validated 3D-QSAR model not only helps in better understanding of the structure activity relationships of any class of molecules, but also provides researcher an insight at molecular level about the lead molecules for further developments and information obtained from 3D-QSAR analysis provides important guidelines for drug design process (6, 7).

Protein tyrosine phosphatase 1B (PTP 1B), an intracellular enzyme which causes negative regulation of insulin receptor as well as leptin signaling system, has emerged as a highly validated, attractive target for the treatment of non-insulin dependent diabetes mellitus and obesity (8, 9). It has been involved in down-regulation of receptor tyrosine kinase activity following stimulation of the insulin or leptin receptors (10, 11). It is a cytosolic PTP, which appears to play a major role in insulin sensitivity and the dephosphorylation of the insulin receptor on the basis of many biochemical and cellular studies (12). A recent pivotal PTP 1B knockout mice study revealed that mice lacking functional PTP 1B exhibited increased sensitivity towards insulin and are resistant to obesity (13). These results taken together, establish a direct role for PTP 1B in down regulating the insulin and leptin functioning. As a result, there is a growing interest in the development of potent and specific inhibitors for this enzyme. PTP 1B inhibitors could potentially ameliorate insulin resistance and normalize plasma glucose and insulin without inducing hypoglycemia, and could therefore be a major advance in the treatment of type 2 diabetes (14).

Self-organizing molecular field analysis (SOMFA) is a novel simple three-dimensional quantitative structure activity relationship (3D-QSAR) methodology having similarities to both comparative molecular field analysis (CoMFA) and molecu-

* Corresponding author: e-mail: manju.papreja@gmail.com; phone: +91-130-2237245, fax: +91-130-2237245

Table 1. General structure of 1,2-naphthoquinone derivatives.



Number	R_1	R_2	R_3	R_4	R_5
1	H	$-NH_2$	H	H	H
2	H	$-N(CH_3)C_6H_5$	H	H	H
3	H	$-OCH_3$	H	H	H
4	H	$-OCH_2\ CH_2CH_2OH$	H	H	H
5	H	$-\ C_6H_3$-2,5-Cl_2	H	H	H
6	H	$-\ C_6H_3$-2,5-F_2	H	H	H
7	H	$-\ C_6H_4$-4-OCH_2COOEt	H	H	H
8	H	$-\ C_6H_4$-4-OH	H	H	H
9	H	$-\ C_6H_4$-2-OH	H	H	H
10	H	$-\ C_6H_2$-3,5-di-*t*-butyl-4-OH	H	H	H
11	H	$-\ C_6H_4$-2-NO_2	H	H	H
12	H	-3-Indole	H	H	H
13	H	-3-Indole-5-carboxylic acid	H	H	H
14	H	-3-Indole-6-carboxylic acid	H	H	H
15	H	-Cyclohexyl	H	H	H
16	H	-Benzyl	H	H	H
17	H	- Cyclopentyl	H	H	H
18	H	-Biphenyl	H	H	H
19	H	-Isopropyl	H	H	H
20	$-O(CH_2)_4CH_3$	$-C_6H_5$	H	H	H
21	H	$-C_6H_5$	H	$-OCH_3$	H
22	H	$-C_6H_5$	H	$-OCH(Bn)COOCH_3$	H
23	H	$-C_6H_5$	H	H	-NHCOOBn
24	H	$-C_6H_5$	H	-NHCOOBn	H
25	H	$-C_6H_5$	H	H	-NHCOOEt
26	H	$-C_6H_5$	$-(CH_2)_2COOMe$	H	H
27	$-(CH_2)_2COOEt$	$-C_6H_5$	H	H	H
28	$-(CH_2)_2COOEt$	Indole	H	H	H
29	$-(CH_2)_2CONEt_2$	$-C_6H_5$	H	H	H
30	$-(CH_2)_2$Ph-4-O-CH_2COOBu^t	$-C_6H_5$	H	H	H
31	$-(CH_2)_2$Ph-4-O-CH_2COOBu^t	Indole	H	H	H
32	$-(CH_2)_2$Ph-4-O-CH_2COOH	$-C_6H_5$	H	H	H
33	$-(CH_2)_2$Ph-4-O-CH_2COOH	Indole	H	H	H
34	$-(CH_2)_2COOMe$	$-C_6H$	$-(CH_2)_2COOMe$	H	H
35	H	C_6H_5	H	H	H
36	H	$-\ C_6H_4$-2-OCH_2COOEt	H	H	H
37	H	- 1-Naphthyl	H	H	H
38	H	$-\ (CH_2)_5C_6H_5$	H	H	H
39	H	$-C_6H_5$	H	$-O(CH_2)_4CH_3$	H
40	H	Indole	H	-OCH(Bn)COOH	H
41	H	Indole	$-(CH_2)_2COOMe$	H	H
42	$-(CH_2)_2PhCOOEt$	$-C_6H_5$	H	H	H
43	$-(CH_2)_2COOMe$	Indole	H	$-(CH_2)_2COOMe$	H

1–34: training set molecules; 35–43: test set molecules.

Table 2. Actual and predicted activities for training set molecules from the best predictive SOMFA model I.

Compound	Actual activity (p IC_{50})	Predicted activity	Residual activity
1	–1.390	–1.193	–0.197
2	–1.542	–0.961	–0.581
3	–1.464	–1.042	–0.422
4	–1.561	–0.825	–0.736
5	–0.703	–0.467	–0.236
6	0.301	–0.101	0.402
7	–0.029	0.225	–0.254
8	0.356	–0.182	0.538
9	–0.204	–0.422	0.218
10	–0.758	–0.397	–0.361
11	–0.068	–0.493	0.425
12	–0.053	–0.257	0.204
13	–0.477	–0.645	0.168
14	–0.658	–0.815	0.157
15	0.494	–0.116	0.610
16	–0.152	–0.385	0.233
17	–0.623	–0.724	0.101
18	–0.732	–0.488	–0.244
19	–1.006	–0.803	–0.203
20	–0.429	–0.424	–0.005
21	–0.310	–0.698	0.388
22	–0.405	–0.430	0.025
23	–0.352	–0.164	–0.188
24	–0.127	0.115	–0.242
25	–1.390	–1.193	–0.197
26	–1.542	–0.961	–0.581
27	–1.464	–1.042	–0.422
28	–1.561	–0.825	–0.736
29	–0.703	–0.467	–0.236
30	0.301	–0.101	0.402
31	–0.029	0.225	–0.254
32	0.356	–0.182	0.538
33	–0.204	–0.422	0.218
34	–0.758	–0.397	–0.361

lar similarity studies (15, 16). Like CoMFA, a grid-based approach is used; however, no probe interaction energies need to be evaluated. Like the similarity methods, it is the intrinsic molecular properties, such as the molecular shape and electrostatic potential, which are used to develop the QSAR models (17). A SOMFA model could suggest a method of tackling the all important alignment, which all 3D-QSAR methods have faced. 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 (18). There are numerous reported SOMFA based 3D-QSAR studies for optimizing the molecular architecture of inhibitors against various target (19–21).

Table 3. Actual and predicted activities for test set molecules from the best predictive SOMFA model I.

Compound	Actual activity (p IC_{50})	Predicted activity	Residual activity
1	0.066	–0.198	0.264
2	–0.332	–0.495	0.163
3	–0.332	–0.337	0.005
4	–0.518	–0.327	–0.191
5	0.036	–0.301	0.337
6	–0.316	–0.278	–0.038
7	0.036	0.165	–0.129
8	–0.004	–0.034	0.030
9	0.027	0.045	–0.018

Table 4. PLS statistical results of SOMFA.

Parameter	Resolution 1.0 Å (Model I)	Resolution 0.5 Å (Model II)
r^2	0.7155	0.7070
$r^2_{cv}(q^2)$	0.6826	0.6735
S	0.3209	0.3257
F	80.487	77.208
$r^2_{prediction}$	0.5534	0.5509
$S_{prediction}$	0.1822	0.1827
C_1	0.68	0.68

$r^2_{cv}(q^2)$: cross-validated correlation coefficient by leave one out method; r^2: conventional correlation coefficient; S: standard error of estimate; F: Fisher test value.

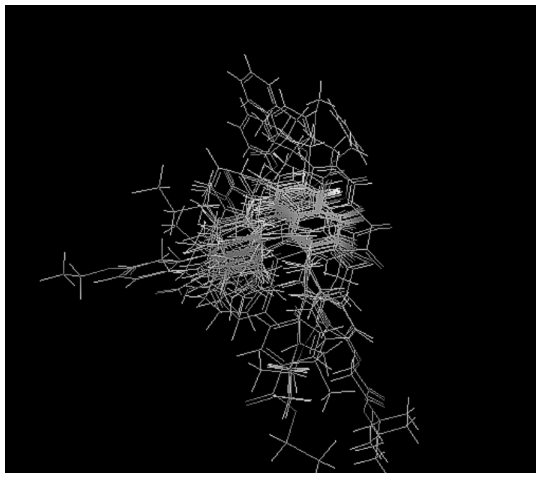


Figure 1. Superimposition of compounds on template

The main objective of the present 3D-QSAR study is to get a validated correlation between the structural features of 1,2-naphthoquinones and their PTP 1B inhibitory activities.

Computational methodology

A dataset of 43 molecules belonging to 1,2-naphthoquinone derivatives as PTP 1B inhibitors were taken from the literature and used for SOMFA analysis (22). The above reported series of 1,2-naphthoquinone derivatives showed wide variations in their structures and potency profiles. The negative logarithm of the measured IC_{50} (μM) against PTP 1B enzyme as pIC_{50} (log $1/IC_{50}$) was used as dependent variable, thus correlating the data linear to the free energy change (23). SOMFA (3D-QSAR) models were generated for this series using a training set of 34 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 9 molecules with uniformly distributed biological activities. The observed and predicted biological activities of the test set molecules are presented in Tables 2 and 3, respectively. 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 (24).

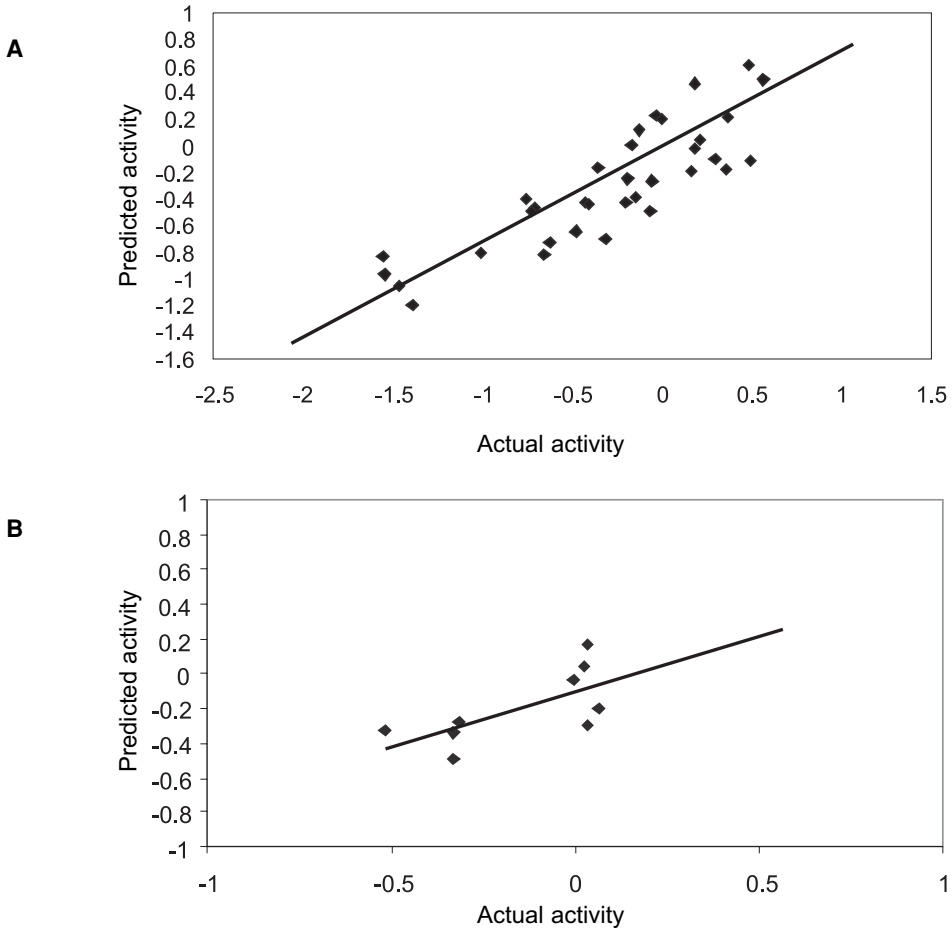


Figure 2. Graph of actual *vs.* predicted activities for training and test set molecules from the best predictive SOMFA model. A Training set, B Test set

The three-dimensional structures of the 1,2-naphthoquinone derivatives were constructed with the Chemdraw Ultra 8.0 running on an Intel Pentium IV 2.80GHz Processor 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) method available in the MOPAC module of Chem3D is adopted for re-optimization until the root mean square (RMS) gradient attains a value smaller than 0.001 kcal/mol Å (25, 26). Unless otherwise indicated, all parameters were kept default.

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 (27, 28). Generally, the low energy conformation of the most active is set as a reference (29). In the present study, the compounds were aligned using low energy conformation of the most active compound (**28**) used as the reference compound by different alignment technique such as atom based and template based alignment. The best model was obtained using template based alignment where the most active compound (**28**) was used as template structure (Fig. 1).

In the SOMFA study, a 40×40×40 Å grid originating at (–20, –20, –20) with a resolution of 0.5 and 1 Å, respectively, was generated around the aligned compounds (30–32). Table 4 reports two different models using different resolution of grid under exploration using template based alignment. For all of the studies, shape and electrostatic potential were generated. To sum up the predictive power of these two properties into one final model, we combine their individual predictions using a

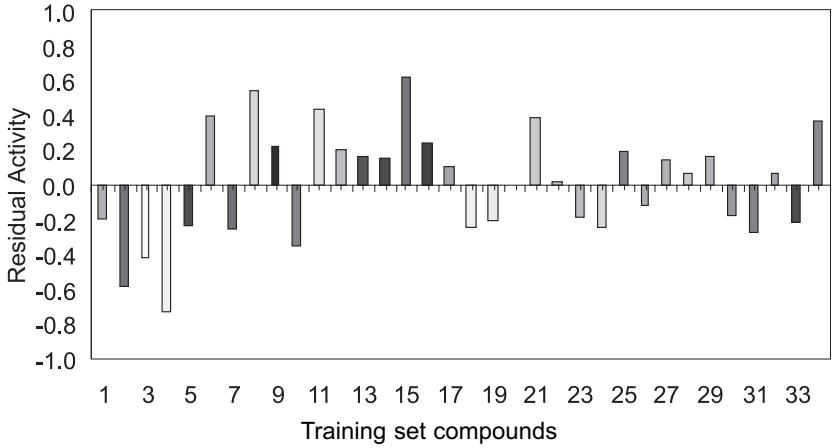


Figure 3. Histogram of SOMFA residual value for training set

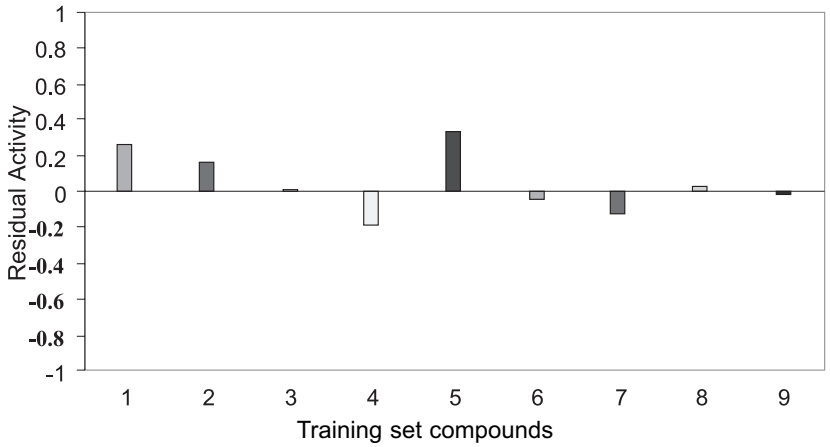


Figure 4. Histogram of SOMFA residual value for test set

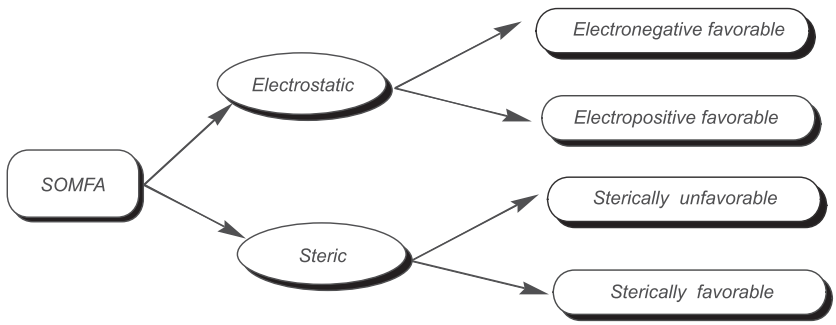


Figure 5. Representation of SOMFA grids

weighed average of the shape and electrostatic potential based QSAR, using a mixing coefficient (C_1) (33) as illustrated in eq. 1:

$$\text{Activity} = C_1\ \text{Activity}_{\text{shape}} + (1 - C_1)\ \text{Activity}_{\text{ESP}} \quad \text{eq.1}$$

Clearly, multiproperty predictions could have been obtained through multiple linear regression. Using eq. 1 instead gives greater insight into the resultant model by allowing the study of the varia-

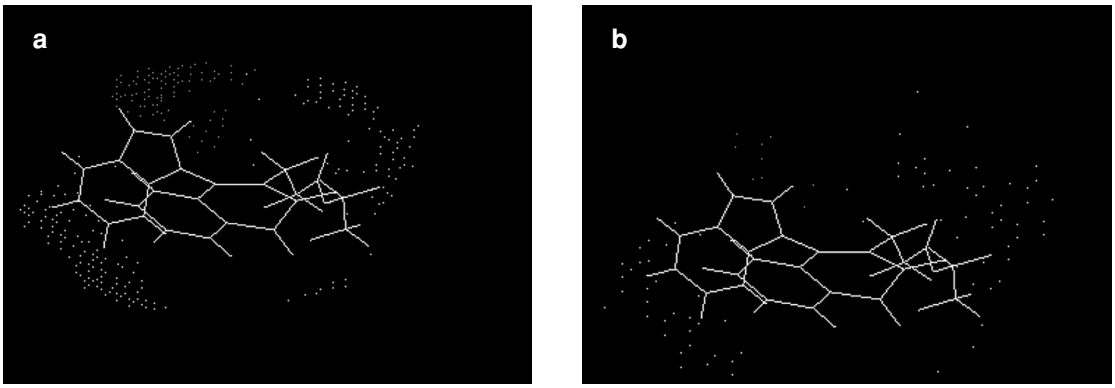


Figure 6. Electrostatic grids showing most active compound (**28**) in the background at different resolutions: a) 1 Å; b) 0.5 Å

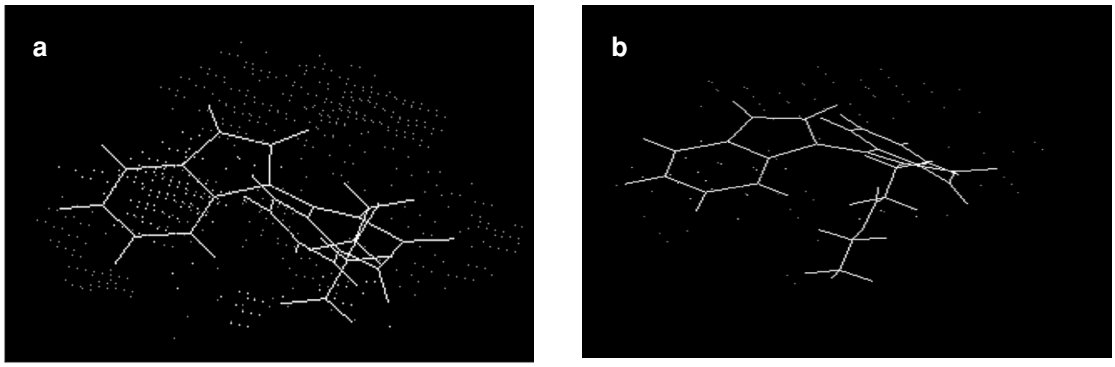


Figure 7. Shape grids showing most active compound (**28**) in the background. at different resolutions a) 1 Å; b) 0.5 Å.

tion in predictive power with different values of C_1. The partial least squares (PLS) algorithm was used in conjugation with leave one out (LOO) cross-validation to develop final model (34). This PLS analysis gave the optimum number of components that was used to generate the final models without cross-validation. The r^2_{cv} 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 (35). 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 of the biological activity.

RESULTS AND DISCUSSION

In the present study, SOMFA was employed for 3D-QSAR study with the training set composed of 34 compounds whose biological activities were known. SOMFA calculation for both shape and electrostatic potentials were performed, then combined to get an optimal coefficient $C_1 = 0.68$ according to eq 1. The SOMFA results indicated that the value of mixing coefficient C_1 is 0.68, which means that the contribution of shape field and electrostatic field to QSAR equation is 68% and 32%, respectively. SOMFA analysis indicated that the electrostatic contribution is of a slightly low importance while shape contribution is of major importance (C_1 = 0.68). During the SOMFA investigation, grid spacings of 1 and 0.5 Å were investigated. The 1Å

grid spacing produces a good correlation equal to 0.5 Å grids. This has been improved marginally with the 1 Å spacing used for the results presented here (Table 4). Further increases in resolution have produced further small increases in model quality but not enough to warrant the extra computational time. From Table 4, we found that the results were less sensitive to resolution of grid. The best model using template based alignment showed higher r^2_{cv} (q^2) value values than using other alignment techniques. Good cross-validated correlation coefficient r^2_{cv} (q^2) value (0.682), moderate non cross-validated correlation coefficient r^2 values (0.7155) proved a good conventional statistical correlation which have been obtained, and we also found that the resultant SOMFA model have a satisfied predictive ability.

The observed and predicted activities of the training set are reported in Table 2 using model I. Figures 2A and 3 show a good linear correlation and moderate difference between observed and predicted values of molecules in the training set. It's well known that the best way to validate a 3D-QSAR model is to predict biological activities for some compounds of test set (36). The SOMFA analysis of the test set composed of 9 compounds is reported in Table 3. Most of compounds in test set show good correlation between observed and predicted values (Figs. 2B, 4).

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 visual 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 and synthesize some novel compounds with much higher activities. Master grid map shows favorable and unfavorable effects (Fig. 5). The visualization of the electrostatic potential master grid and shape master grid of the best SOMFA model was showed in Figures 6 and 7, respectively, with most active compound (**28**) as the reference. The SOMFA electrostatic potential map shows some important features (Fig. 6); we observed a high density of points around the substituent R_1 and R_3 of the 1,2-naphthoquinone indicating some electronegative groups favorable for optimal inhibitory activities, while around R_2, R_4 and R_5 high density of points indicated that some electropositive groups are favorable for optimal inhibitory activities. Meanwhile, in the map of shape master grid (Fig. 7), a high density of points around various substituents were present indicating a favorable steric interaction; simultaneously, we observed also some points near regions where an unfavorable steric interaction may be expected to enhance activities.

CONCLUSION

In summary, we have developed a predictive SOMFA 3D-QSAR models for 1,2-naphthoquinones having wide variety of PTP 1B inhibitory activity evidenced by statistical measures. The master grid obtained for the various SOMFA models indicates that electrostatic potential and shape potential contributions, which can be mapped back onto structural features relating to the trends in activities of the molecules. The spatial arrangement of the shape and electrostatic potential could be used as a guide for further development of selective and more potent PTP 1B inhibitors.

Acknowledgement

The authors gratefully acknowledge GVM Institute of Technology for the support provided to carry out this project.

REFERENCES

1. Bharati D., Jagtap R.S., Kanase K.G. et al.: Asian J. Res. Chem. 2, 1 (2009).
2. Chemical Informatics at Indiana University, www.chembiogrid.org /related/resources/ciatiu.html.
3. Prakash N., Gareja D.A.: J. Proteom. Bioinform. 3, 249 (2010).
4. Roberts G., Myatt G.J., Johnson W.P. et al.: J. Chem. Inform. Comp. Sci. 40, 1302 (2000).
5. Paliwal S., Narayan A., Paliwal S.: QSAR Comb. Sci. 28, 1367 (2009).
6. Rogers D., Hopfinger A.J.: J. Chem. Inform. Comp. Sci. 34, 854 (1994).
7. Richon A. B., Young S.C.: An introduction to QSAR methodology. http://www.netsci.org /Science/Compchem/feature19.html
8. Thareja S., Aggarwal S., Bhardwaj T.R., Kumar M.: Med. Res. Rev. DOI 10.1002/med.20219 (article in press).
9. Kennedy B.P.: Biomed. Pharmacother. 53, 466 (1999).
10. Salmeen A., Andersen J.N., Myers M.P., Tonks N.K., Barford D.: Mol. Cell 6, 1401 (2000).
11. Zabeau L., Lavens D., Peelman F., Eyckerman S., Vandekerckhove J., Tavernier J.: FEBS Lett. 546, 45 (2003).
12. Byon J. C., Kusari J., Kowari A. B.: Mol. Cell. Biochem. 182, 101 (1998).

13. Elchebly M., Payette P., Michaliszyn E., Cromlish W., Collins S., Loy A.L. et al.: Science 283, 1544 (1999).
14. Liu G.: Curr. Med. Chem. 10, 1407 (2003).
15. Robinson D.D., Winn P.J., Lyne P.D., Richards W.G.: J. Med. Chem. 42, 573 (1999).
16. Cramer R.D., Patteerson D.E., Bunce J.D.: J. Am. Chem. Soc. 110, 5959 (1988).
17. Zheng L.Y.: Int. J. Mol. Sci. 7, 220 (2006).
18. Li M., Du L., Wu B., Xia L.: Bioorg. Med. Chem. 11, 3945 (2003).
19. Thareja S., Aggarwal S., Bhardwaj T. R. et al.: Eur. J. Med. Chem. 45, 2537 (2010).
20. Aggarwal S, Thareja S., Bhardwaj T.R. et al.: Eur. J. Med. Chem. 45, 476 (2010).
21. Thareja S., Aggarwal S., Bhardwaj T.R. et al.: Med. Chem. 6, 30 (2010).
22. Ahn J.H., Cho S.Y., Ha J.D., Chu S.Y., Jung S.H., Jung Y.S., Baek J.Y. et al.: Bioorg. Med. Chem. Lett. 12, 1941 (2002).
23. Puntambekar D.S., Giridhar R., Yadav M.R.: Eur. J. Med. Chem. 43, 142 (2008) .
24. Murthy V.S., Kulkarni V.M.: Bioorg. Med. Chem. 10, 2267 (2002).
25. Dewar M.J.S., Zoebisch E.G., Healy E.F. et al.: J. Am. Chem. Soc. 107, 3902 (1985).
26. Stewart J.J.: Comput. Aid. Mol. Des. 4, 1 (1990).
27. Agarwal A., Taylor E.W.: J. Comput. Chem. 14, 237 (1993).
28. Baurin N., Vangrevelinghe E., Allory L.M.: J. Med. Chem. 43, 1109 (2000).
29. Xu M., Zhang A., Han S., Wang L.: Chemosphere 48, 707 (2002).
30. Pedretti A., Villa L., Vistoli G.: J. Mol. Graph. Model. 21, 47 (2002).
31. VEGA ZZ Release 2.3.1.1 can be free downloaded from: http://www.ddl.unimi.it/vega/index2.htm
32. SOMFA2 v2.0.0 downloaded from: http://bellatrix.pcl.ox.ac.uk.(2006)
33. Li M., Fang H., Xia L.: Bioorg. Med. Chem. 15, 3216 (2005).
34. Zambre V. P., Murumkar P. R., Giridhar R., Yadav M.R.: J. Chem. Inf. Model. 49, 1298 (2009).
35. Golbraikh A., Tropsha A.: J. Mol. Graph. Model. 20, 269 (2002).
36. Kulkarni S.S., Gediya L.K., Kulkarni V.M.: Bioorg. Med. Chem. Lett. 7, 1475 (1999).

Received: 27. 05. 2011