Original Article

Computer assisted SOMFA Tool Validation using 3D-QSAR Study on Selective Inhibitors of Glucagon Receptor

Manju Papreja*[1,2] and Payal Pahwa[2,3]

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

Address for Correspondence

GVM Institute of Technology and Management, Sonipat (Haryana)-131001, India

E-mail: manju.papreja@gmail.com

ABSTRACT

Objective: A Cheminformatics based 3D-QSAR study was performed to test and validate the reliability of SOMFA tool for Drug Design.
Methods: For development of a statistically reliable model and validation of SOMFA tool, 27 molecules belonging to triarylimidaozle scaffold were taken from the reported studies and processed through SOMFA.
Results: During SOMFA investigation, best model obtained using atom based alignment showed good cross-validated correlation coefficient r^2_{cv} (q^2) (0.6911), non cross-validated correlation coefficient r^2 values (0.7197), low standard estimation of estimation S (0.5541) and high F-test value (51.3441), showing good statistical correlation.
Conclusion: The models thus obtained were accepted by various statistical parameters and thus validate the robustness and reliability of SOMFA tool for drug design.

Keywords: Cheminformatics, Computational Sciences, Drug Design, 3D-QSAR, SOMFA.

INTRODUCTION

Cheminformatics is the mixing of information resources to transform data into information and transformation of information for the decision making in the area of drug design and devlopement.[1] Cheminformatics involving application of computational and information tools for a range of problems in the field of drug design.[2] It is a interconnection of computer science and drug design for storing, retrieving, searching of information along with data.[3] A wide variety of computational tools are available for studying behavior of molecules which has been proved to be meaningful for designing novel molecules.[4] Three-dimensional quantitative structure-activity relations (3D-QSAR), is emerged as a powerful Cheminformatics technique which

changed the area of drug discovery. QSAR studies helpful in providing structural features in the form of molecular descriptor that can be useful for optimizing drug molecules.

A validated statistically significant correlation between structural feature and corresponding activity obtained using QSAR studies.[5,6] The approach is based on the assumption that changes in measured or computed molecular features can be correlated with variations in biological activity. A reliable, robust statistically validated QSAR model helped in studying the structure activity relationships of any class of molecules, but also provides insight at molecular level about the lead molecules for further developments.[7] Thus, 3D-QSAR analysis provides a useful framework which is used as a reliable approach in drug design.[8-11]

A novel three 3D-QSAR technique known as Self-organizing molecular field analysis (SOMFA) developed by Robinson et al. having similarity in concept with both molecular similarity studies and comparative molecular field analysis (CoMFA).[12, 13] It is a new technique based on molecular properties such as electrostatic and steric potential. It has been developed to tackle the alignment problem from which all 3D-QSAR methodologies suffers. This technique allows the possibility of aligning the training molecules as an integral part of the model derivation process and of aligning prediction molecules to optimize their predicted activities.[14]

Recently, there are various studies in the literature involving use of SOMFA software for 3D-QSAR studies for the purpose of refinement of molecular architecture for different classes of molecules.[15,16] In present study, it was considered of interest to carry out a Cheminformatics based 3D-QSAR study to test and validate the reliability of SOMFA tool for Drug Design.

Validation Procedure

Data Set

For development of a statistically reliable model and validation of SOMFA tool, 27 molecules belonging to triarylimidaozle scaffold were used from the reported studies and processed through SOMFA. [17] Due to presence of variety of chemical structure and potency profile, 3D-QSAR was performed on this series. Data set was rationally divided into training and test set. The training data set consists of 22 molecules which are used in model deriving process. The tests set consist of remaining 5 molecules having evenly distributed activities for testing the predictability of the developed models. **Table 1** depicted the representation of all the molecules (training and test set). The criteria for selection of test data set have been based on structural similarity of the molecules with the training set.

Biological Activities

The activity used for developing QSAR model has been used by taking -log of the measured IC_{50} (M) against glucagon receptor as pIC_{50}, which arranges the data in a linear manner.[18]

Computational Modeling and Alignment

Data set of the triarylimidaozle derivatives were framed using Chemdraw running and were initially subjected to molecular mechanics (MM2) algorithm for energy minimization and minimized until the root mean square (RMS) gradient value reaches a value.[19,20] The local minima structure of the most active molecule is used as the reference structure.[21] The alignment was done using local minima structure of the most potent compound[22] used as the reference compound by atom based alignment technique where centroid of atoms were used for alignment shown in **Fig. 1**. The structures were superimposed on a template structure of low energy depicted in **Fig. 2**).

SOMFA Models

A 40x40x40 Å grid model instigated at (-20,-20,-20) with a resolution of 0.5 À, was framed around the united molecules.[22-24]The statistical analysis was carried out using PLS algorithm in conjugation with leave one out (LOO) cross-validation to develop the final model. Results of PLS analysis generated using optimum components produces the final models without cross-validation for measuring the statistical importance of the model. The result from a cross-validation analysis was expressed as r^2_{cv} (q^2) value. The r^2_{cv} (q^2) can take up values in the range from 1, suggesting a perfect model, < 0 where errors of prediction are greater than the error from assigning each compound mean potency of the model.[25]

Fischer Statistics (F-Test), another useful statistical measure used to check reliability of the developed 3D-QSAR model. Higher value of F indicated higher statistical significant model.[26, 27]

RESULTS AND DISCUSSION

SOMFA tool was evaluated using a training data set comprises of 22 triarylimidazole derivatives to correlate with corresponding biological potential using PLS statistical technique in Microsoft excel. SOMFA calculation was done to develop the descriptors in the form of shape and electrostatic models. The involvement of shape and electrostatic model to QSAR equation is 52% and 48%, respectively which indicated that the electrostatic contribution is of a slightly lower importance than shape contribution (52%) (**Table 3)**. The SOMFA 3D-QSAR electrostatic potential and shape model has been presented in the form of 3D-Grids.It was earlier reported in the literature that 0.5 Å grid spacing produced a good correlation equal to 1.0 Å grids. Further increase in resolution has produced small increase in model quality but not enough to warrant the extra computational time. Therefore, in present SOMFA investigation, grid spacing of 0.5 Å was investigated to develop final 3D-QSAR model. The best 3D-QSAR model obtained using atom based alignment showed good correlation coefficient cross-validated r^2_{cv} (q^2) (0.6911), non cross-validated correlation coefficient r^2 values (0.7197), low standard estimation of estimation S (0.5541) and high F-test value (51.3441), showing good statistical correlation.

Table 2 showed the predicted and observed activities of the training data set molecules using SOMFA model. The predicted and observed activities of molecules in the dataset showed a good linear correlation and moderate difference Fig**. (3, 5)**.

The SOMFA model of the test dataset of 5 molecules also described in **Table 2**. The predictive efficiency was also confirmed using good non cross-validated prediction correlation coefficient (test set) r^2 values (0.6731). Almost all test dataset molecules set showed good correlation between predicted and observed activities **Fig. (4,5)**.

The SOMFA shape and electrostatic potential have been generated in the form of 3D- grids (**Fig. 6, 7**) using resolution of 0.5 Å. The master grid maps were used to display the contribution of shape and electrostatic potential. The master grid maps gave a direct visual indication regarding structural features responsible to differentiate the activities of molecules in the training set under study.

CONCLUSION

A good predictive statistical reliable SOMFA 3D-QSAR models for triarylimidazole having flexibility in structure and potency profile against glucagon receptor inhibitors have been developed successfully evidenced by statistical measures. The SOMFA steric potential map showed some important features including a high density of

red points around '$R_1$and R_3 ' of imidazole skeleton indicated presence of a encouraging steric feature while few blue points around 'R_2' indicated adverse steric feature for optimal activity. The SOMFA electrostatic potential map also showed some important features such as few red points around the substituent 'R_1 and R_3' indicated electropositive substituent's are favorable for optimal inhibitory activity while some blue points around 'R_2' indicated presence of electronegative substituent's for good inhibitory activity. SOMFA master grids indicated significant electrostatic and shape potential contributions. The models thus obtained were accepted by various statistical parameters and thus validate the robustness and reliability of SOMFA tool for drug design.

ACKNOWLEDGEMENTS

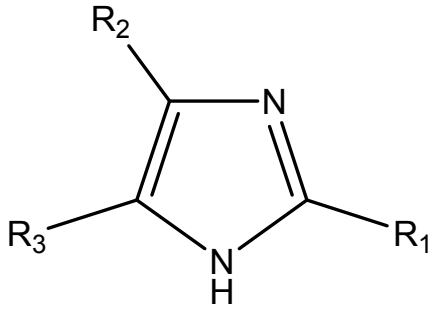
The authors gratefully acknowledge GVM Institute of technology for the support provided to carry out this project.

REFERENCES

1. Gaikwad, V. J. "Application of chemo informatics for innovative drug discovery."*International Journal of Chemical Sciences and Applications*, 1(1), pp.16-24, 2010.
2. Prakash, N.,Gareja, D.A. "Cheminformatics." *Journal of Proteomics and Bioinformatics,* 3(8), pp.249-252, 2010.
3. Bharati, D., *et al.* "Chemo informatics: Newer Approach for Drug Development."*Asian Journal of Research in Chemistry*, 2(1), pp.1-7, 2009.
4. Roberts, G., *et al.* "LeadScope: Software for Exploring Large Sets of Screening Data."*Journal of Chemical Information and Computer Sciences*, 40(6), pp.1302-1314, 2000.
5. Paliwal, S., Narayan, A.,Paliwal, S."Quantitative Structure Activity Relationship Analysis of DicationicDiphenylisoxazole as Potent Anti-Trypanosomal Agents."*QSAR & Combinatorial Science*, 28(11-12), pp.1367-1375, 2009.
6. Rogers, D.,Hopfinger, A.J. "Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships."*Journal of Chemical Information and Computer Sciences*, 34(4), pp.854-866, 1994.
7. Thareja, S., *et al.* "Sulphonamides as Inhibitors of Protein Tyrosine Phosphatase 1B: A Three-Dimensional Quantitative Structure–Activity Relationship Study Using Self-Organizing Molecular Field Analysis Approach." *Chemical and Pharmaceutical Bulletin*, 58(4), pp. 526-532, 2010.
8. Richon, A. B., An Introduction to QSAR Methodology. http://www.netsci.org /Science/ Compchem/feature19.html
9. Raju, Y. ; Suresh Babu, D. "An Effective Personalized Search Engine Architecture for Re-ranking Search Results Using User Behavior" *American Journal of Computer Science and Engineering Survey*, 3(6), pp.159-167, 2008.
10. Prithiviraj, P.; Porkodi, R. "A Comparative Analysis of Association Rule Mining Algorithms in Data Mining: A Study." *American Journal of Computer Science and Engineering Survey*, 3(1), pp.98-119, 2015.
11. Ali,S.,Drucker,D. J. "Benefits and limitations of reducing glucagon action for the treatment of type 2 diabetes." *American Journal of Physiology - Endocrinology and Metabolism*, 296(3), E415–421, 2009.
12. Robinson, D.D., *et al.* "Self-Organizing Molecular Field Analysis: A Tool for Structure-Activity Studies."*Journal of Medicinal Chemistry*, 42(4), pp.573-583, 1999.
13. Cramer, R.D., III, Patteerson, D.E.,Bunce, J.D. "Comparative Molecular Field Analysis (CoMFA).Effect of Shape on Binding of Steroids to Carrier Proteins."*Journal of the American Chemical Society*, 110 (18), pp.5959-5967, 1988.
14. Thareja, S., *et al.*"Self organizing molecular field analysis on a series of human 5α reductase inhibitors: Unsaturated 3-carboxysteroid."*European Journal of*

Medicinal Chemistry, 44(12),pp.4920-4925, 2009.
15. Aggarwal, S., *et al.* "3D-QSAR studies on unsaturated 4-Azasteroids as human 5α-Reductase Inhibitors: A Self Organizing Molecular Field Analysis Approach."*European Journal of Medicinal Chemistry*, 45(2),pp.476-481, 2010.
16. Thareja, S., *et al.* "Self Organizing Molecular Field Analysis of 2, 4-Thiazolidinediones: A 3D-QSAR Model for the Development of Human PTP 1B Inhibitors."*European Journal of Medicinal Chemistry*, 45(6), pp.2537-2546, 2010.
17. Chang, L.L., *et al.*"Substituted imidazoles as glucagon receptor antagonists."*Bioorganic & Medicinal Chemistry Letters*, 11(18), pp.2549-2553, 2001.
18. Aggarwal, S.,*et al.* "Self Organizing Molecular Field Analysis on Pregnane derivatives as Human Steroidal 5α-Reductase Inhibitors."*Steroids*, 75(6), pp.411-418, 2010.
19. Dewar, M.J.S, *et al.* "Development and use of Quantum Mechanical Molecular Models AM1: A new general purpose Quantum Mechanical Molecular Model."*Journal of the American Chemical Society*,107 (13), 3902-3909, 1985.
20. Stewart, J." MOPAC: A Semiempirical Molecular Orbital Program."*Journal of Computer-Aided Molecular Design*, 4(1),pp.1-103, 1990.
21. Xu, M., *et al.* "Studies of 3D-Quantitative Structure-Activity Relationships on a set of nitro aromatic compounds: CoMFA advanced CoMFA and CoMSIA."*Chemosphere*, 48(7), pp.707-715, 2002.
22. VEGA ZZ Release 2.3.1.1 can be free downloaded from: http://www.ddl.unimi.it/vega/index2.htm.
23. SOMFA2 v2.0.0 downloaded from: http://bellatrix.pcl.ox.ac.uk.(2006).
24. Li, M.Y., Fang, H., Xia, L. "Pharmacophore based design, synthesis, biological evaluation, and 3D-QSAR studies of aryl-piperazines as $[\alpha]_1$-adrenoceptor antagonists." *Bioorganic & Medicinal Chemistry Letters*, 15(13), pp.3216-3219, 2005.
25. Golbraikh, A.,Tropsha, A. "Beware of q^2!"*Journal of Molecular Graphics and Modelling*, 20(4), pp.269-276, 2002.
26. Puntambekar, D.S., Giridhar, R., Yadav, M.R. "Insights into the structural requirements of farnesyltransferase inhibitors as potential anti-tumor agents based on 3D-QSAR CoMFA and COMSIA Models."*European Journal of Medicinal Chemistry*, 43(1), pp.142-154, 2008.
27. Puntambekar, D., Giridhar, R.,Yadav, M.R. "3D-QSAR Studies of Farnesyltransferase inhibitors: A comparative Molecular Field Analysis Approach."*Bioorganic & Medicinal Chemistry Letters*, 16(7), pp.1821-1827, 2006.

Table 1. Structure of triarylimidazole derivatives



Compound Number	R_1	R_2	R_3
1.	(4-Br)Ph	(4-F)Ph	4-pyridyl
2.	(3-Br)Ph	(4-F)Ph	4-pyridyl
3.	(4-Cl)Ph	(4-F)Ph	4-pyridyl
4.	(4-F)Ph	(4-F)Ph	4-pyridyl
5.	(4-I)Ph	(4-F)Ph	4-pyridyl
6.	(4-Me)Ph	(4-F)Ph	4-pyridyl
7.	(4-iPrPh	(4-F)Ph	4-pyridyl
8.	(4-Ph)Ph	(4-F)Ph	4-pyridyl
9.	(4-NH_2)Ph	(4-F)Ph	4-pyridyl
10.	(4-OMe)Ph	(4-F)Ph	4-pyridyl
11.	(4-CNPh	(4-F)Ph	4-pyridyl
12.	(4-COOMe)Ph	(4-F)Ph	4-pyridyl
13.	(4-SMe)Ph	(4-F)Ph	4-pyridyl
14.	(4-Br)Ph	Ph	4-pyridyl
15.	(4-Cl)Ph	(4-F)Ph	3-Me(4-pyridyl)
16.	(4-Cl)Ph	(4-Cl)Ph	4-pyridyl
17.	(4-Cl)Ph	(4-I)Ph	4-pyridyl
18.	(4-Cl)Ph	(4-Ph)Ph	4-pyridyl
19	(4-Cl)Ph	(4-t-Bu)Ph	4-pyridyl
20.	(4-Cl)Ph	(4-n-Bu)Ph	4-pyridyl
21	(4-Cl)Ph	(3-Ph)Ph	4-pyridyl
22	(4-Cl)Ph	(2-OPh)Ph	4-pyridyl
23	(4-Cl)Ph	(3-OPh)Ph	4-pyridyl
24	(4-Cl)Ph	(4-OPh)Ph	4-pyridyl
25	(4-Cl)Ph	(2O-n-Bu)Ph	4-pyridyl
26	(4-Cl)Ph	(2,4-(O-n-Pr)$_2$)Ph	4-pyridyl
27	(4-Cl)Ph	(2,4-(O-n-Bu)$_2$)Ph	4-pyridyl

Table 2. Actual and Predicted activities for Training and Test set molecules from SOMFA model

Compound	Actual Activity (pIC_{50})	Predicted Activity	Residual Activity
1	6.568	6.050	0.491
2^{T}	5.853	6.045	-0.156
3	6.398	6.067	0.282
4	5.699	6.042	-0.371
5^{T}	6.292	6.050	0.197
6	5.886	6.016	-0.187
7	6.155	5.944	0.216
8	5.000	5.901	-0.946
9	5.699	6.077	-0.361
10^{T}	4.886	5.953	-1.094
11	5.097	5.949	-0.914
12	5.06	5.724	-0.758
13	6.31	5.907	0.352
14	6.107	6.420	-0.027
15^{T}	5.959	6.181	-0.489
16	6.721	6.126	0.464
17	6.886	6.277	0.644
18	6.854	6.588	0.236
19	6.886	6.716	0.292
20^{T}	7.131	6.665	0.605
21	7.215	6.735	0.551
22	8.187	8.242	-0.065
23	7.886	7.391	0.22
24	7.569	7.428	0.072
25	8.071	7.183	1.107
26	7.886	8.684	-0.700
27	8.187	8.861	-0.718

T-Test Set Molecules

Table 3: Statistical results of 3D-QSAR studies

Parameter	Resolution (0.5 Å)
q^2	0.6911
r^2	0.7197
S	0.5541
F	51.3441
r^2_{pred}	0.6731
Contributions	*Shape 52 %* *Electrostatic 48%*

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;
r^2_{pred}: Correlation coefficient f or prediction (test) set

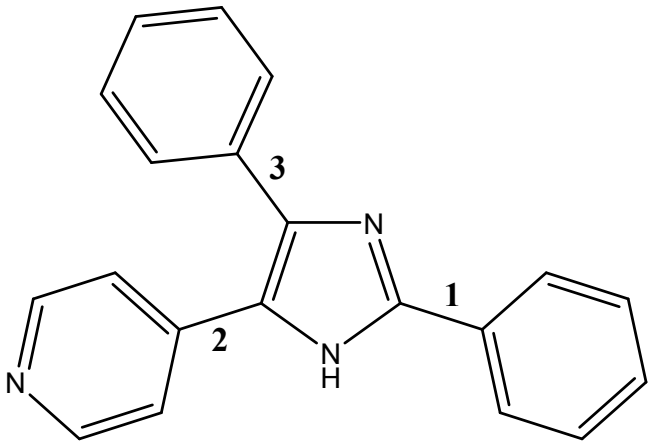


Figure 1. Alignment of atoms used for developing 3D-QSAR model

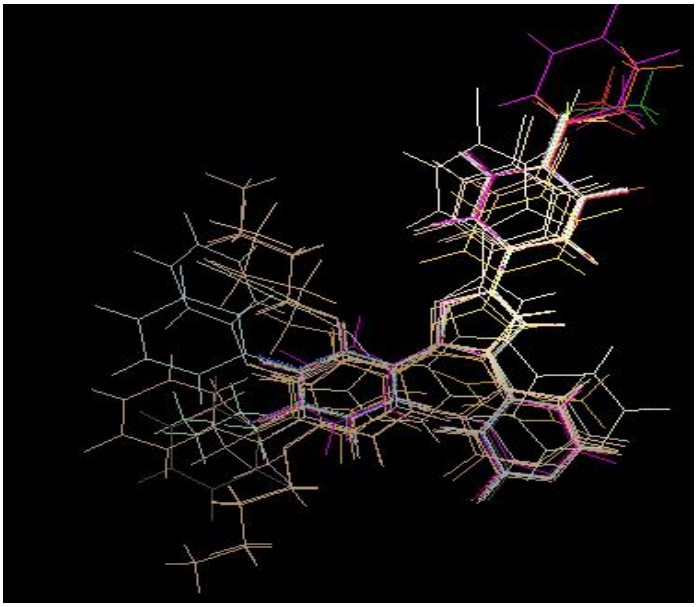


Figure 2. Superimposition of all structures on the most potent compound[22]

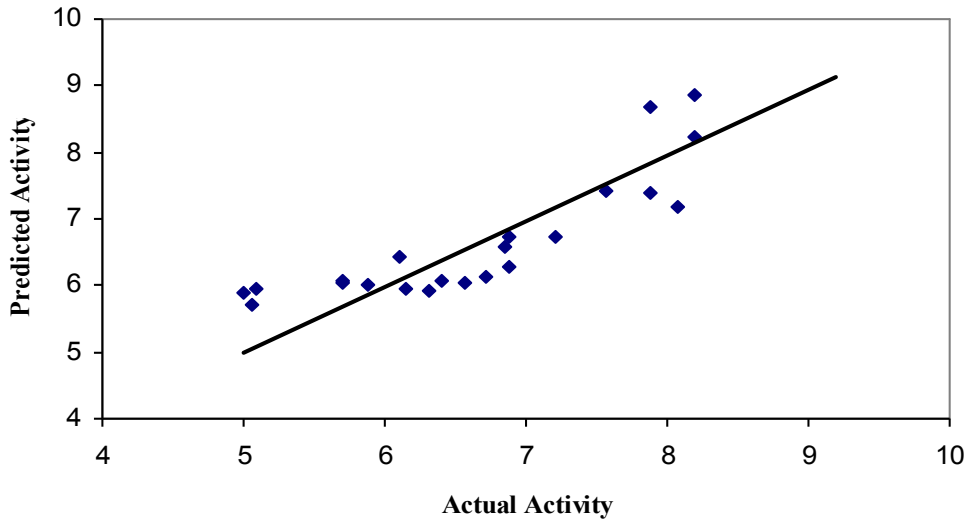


Figure 3. Plot of actual vs. predicted activities of training set molecules

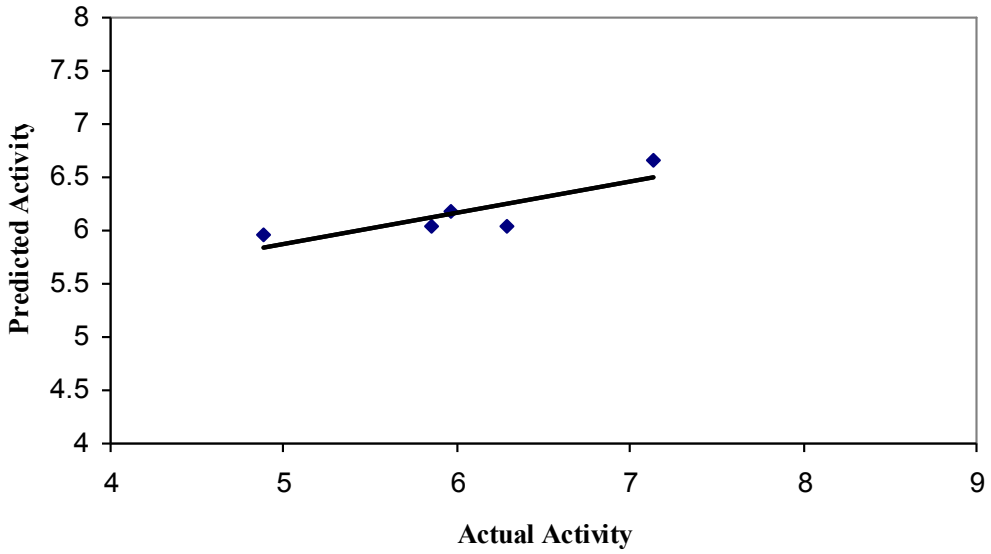


Figure 4. Plot of actual vs. predicted activities of test set molecules

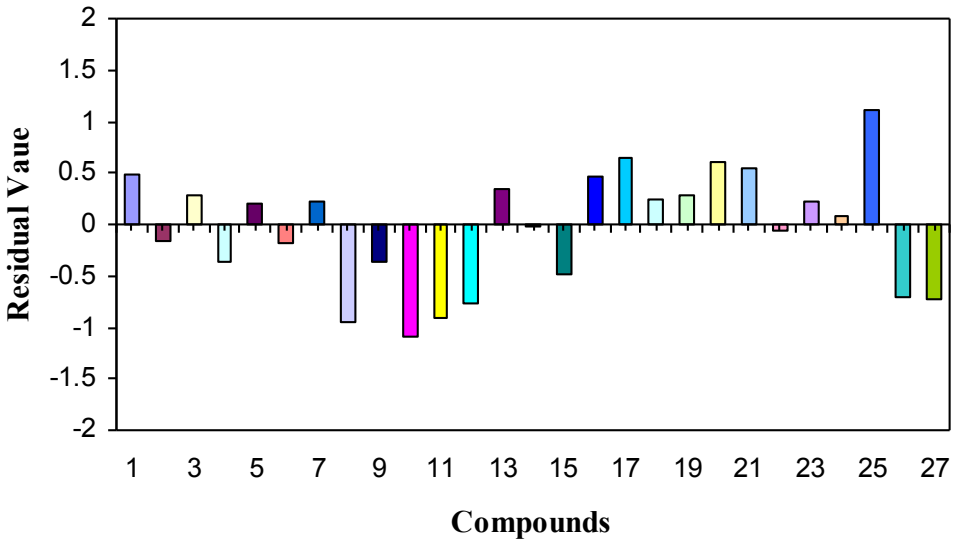


Figure 5. Histogram of SOMFA residual value for all the compounds

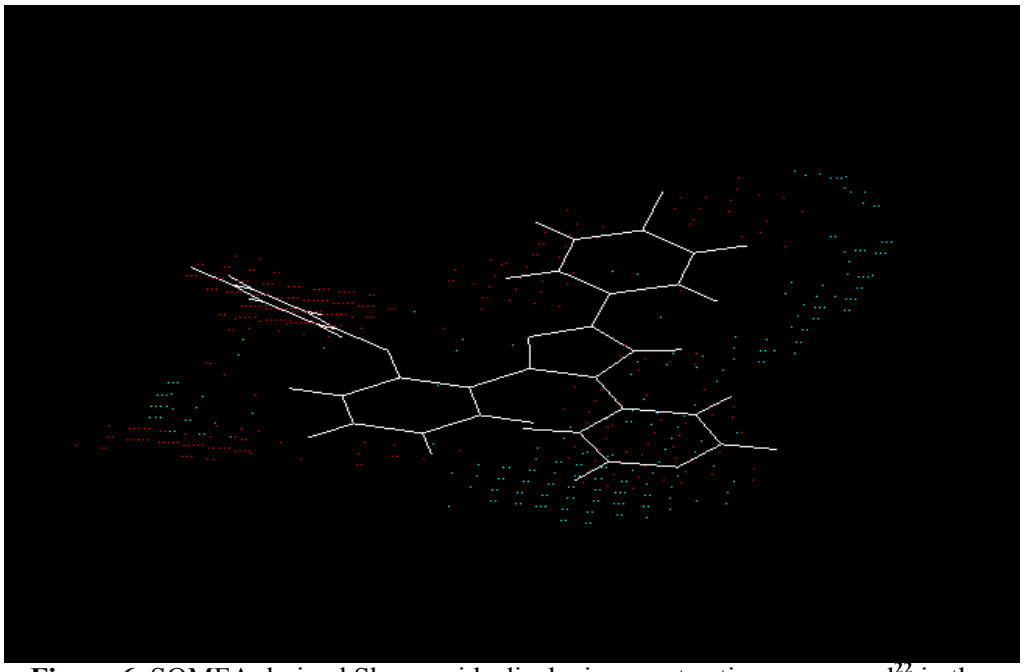


Figure 6. SOMFA derived Shape grids displaying most active compound[22] in the background at 0.5 Å resolution

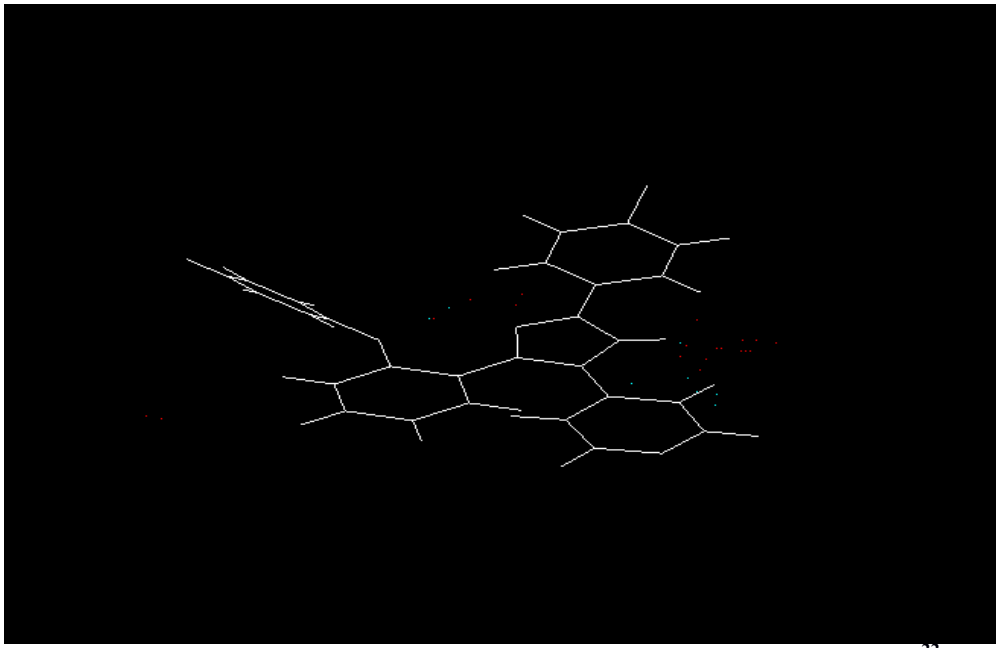


Figure 7. SOMFA derived Electrostatic grids displaying most active compound[22] in the background at 0.5 Å resolution