Available online at www.scholarsresearchlibrary.com

Scholars Research Library

***J. Comput. Method. Mol. Design*, 2011, 1 (3): 49-56**
(http://scholarsresearchlibrary.com/archive.html)

ISSN : 2231- 3176
CODEN (USA): JCMMDA

Molecular docking and 3D QSAR studies of quinoxaline derivatives as potential influenza NS1A protein inhibitors

Vijay H. Masand[a*], Devidas T. Mahajan[a], Komalsing N. Patil[a], Nilesh Padole[b], Taibi Ben Hadda[c], Ahmed A. Alafeefy[d], I.G. Shibi[e]

[a] *Department of Chemistry, Vidya Bharati College, Camp, Amravati, Maharashtra, India.*
[b] *Department of Cosmetics, Vidya Bharati College, Camp, Amravati, Maharashtra, India.*
[c] *LaboratoireChimie des Matériaux, Université Mohammed Premier, Oujda-60000, Morocco.*
[d] *Department of Pharmaceutical Chemistry, College of Pharmacy, Alkharj University, Alkharj, 11451, P.O. Box 2457, Saudi Arabia*
[e] *Department of Chemistry, Sree Narayana College, Kollam- 691001, KERALA, India*

ABSTRACT

In present work, we have performed molecular docking and 3D QSAR analysis of quinoxaline derivatives, previously reported as potential influenza NS1A protein inhibitors..The docking analysis reveals that presence of water molecule inside the cavity of receptor play very crucial role. For better outcome, receptor based electrostatic potential map are also analyzed. The QSAR model is robust, statistically sound and validated thoroughly to avoid over fitting and chancy correlations. The three parametric model is with R^2 = 0.874, adj. R^2 = 0.859, pred. R^2 = 0.805, F value = 60.196. The analysis indicates that the biologic activity depends upon 3D descriptors. Combined use of different types of 3-D descriptors like WHIM, GATEAWAY and 3D MoRSE afforded valuable QSAR model. The analysis could be very useful in designing better influenza NS1A protein inhibitors.

Keywords: Molecular docking, 3D QSAR, Quinoxaline derivatives, anti-influenza, Drug Designing

---

INTRODUCTION

In 2009, world suffered from highly communicable respiratory disease “influenza” caused by influenza virus. All the three types of Influenza viruses viz. influenza A, influenza B, and

influenza C are a serious threat to humans[1]. Of the three types, the type A virus is more dangerous and its one of the subtype H1N1 caused the 2009 flu pandemic and H5N1 is a recent pandemic threat[2]. Due to this, development of new anti-influenza drug for effective treatment has always gained significant attraction. Literature survey reveals that in influenza virus many enzymes viz. NS1, NS3 etc. play vital role in the life cycle of virus and in addition some of them are highly conserved[3]. The NS1 protein which is absent in humans has been identified as a potential target for antiviral development. NS1 is highly essential for virus replication; therefore development of drug to suppress its normal functioning with better inhibitory profile is under progress. Modern drug designing methodologies viz. QSAR and Molecular Docking are very effective in developing new drugs with higher efficiency and lower toxicity[4].

In molecular docking is highly useful in understanding the way the drug interacts with the protein and the factors due to which drug binds with receptor. In QSAR, 2D and 3D descriptors are used to find mathematical correlations with biological activity.

The objectives of present work are (1) to perform molecular docking to understand the types of interactions involved between receptor and drug (2) to determine the structural features that governs the interactions of drug with receptor (3) to select appropriate number and type of descriptors to built QSAR model for anti-influenza activity of quinoxaline derivatives with no problem of "Over Fitting" (4) to develop robust and statistically sound QSAR model (5) to evaluate the QSAR model not only in terms of predictivity, but in terms of its ability to afford a chemical and structural explanation also. The results should serve as a guideline in designing more potent and selective anti-influenza.

MATERIALS AND METHODS

2. Computational Method/Experimental protocol:

2.1 Data set:

The data set of 33 molecules was used to model anti-influenza activity. It comprises quinoxazoline derivatives with wide variety of substituents from electron donating to electron withdrawing located at several positions in the bicyclic core as shown in Fig. 1. The activities of these compounds have been reported elsewhere[3]. For the sake of convenience, the data reported in the form of %Binding at 50µm was converted to $-\log_{10}$(%Binding at 50µm) ie. p(Binding at 50µm). These are listed in table 2.

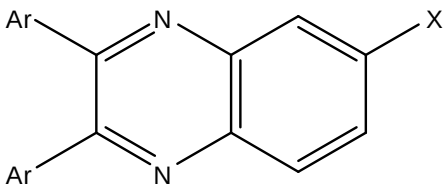

Fig. 1. 2,3,6-substituted quinoxaline derivatives

2.2. Preparation of the structures:
The 33 molecules were drawn in ChemSketch 12 freeware followed by optimization before saving in .mol file format. The descriptors were calculated using e-Dragon and PowerMV. The descriptors were 2D as well as 3D in nature. Since the calculations of these descriptors are well documented in the literature, it is not necessary to duplicate the same here.

2.3 Docking strategy:
Docking procedures were performed on NS1 effector domain as receptor, downloading its structure from the Protein Data Bank (PDB). Different files of NS1 effector domain are available from that web site. We selected PDB file (PDB code: 3EE9) on the basis of X-ray resolution, 2.14A^{o} in this case (from Research Collaboratory for Structural Bioinformatics (RCSB) http://www.rcsb.org/pdb). The structure of protein was validated by plotting Ramchandran plot using "Protein Geometry" module, which shows that no residue is an outlier (Fig.2).

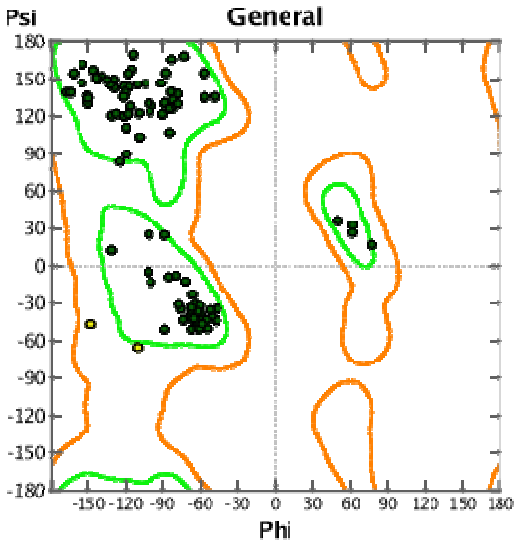

Fig. 2: Ramchandran plot of the PDB 3EE9 after optimization.

Before actual docking, the molecular structures were further prepared along with the proteins (charges and protonation states were assigned) by the docking engine. Docking of quinazoline derivatives to NS1 effector domain proteins was carried out using the standard procedure of Auto-Dock. The deprotonated form for the quinazoline derivatives was assigned and atomic charges were added using Gasteiger–Marsili formalism, which was the charge method, used in the calibration the AutoDock empirical free energy function. The ligand was set up for docking with the help of Autotors and the number of flexible torsions to be considered during the docking process was defined to 4, the hydroxyl and phenyl rotors.

Hydrogen bond analysis was performed on NS1 docked with different quinazoline derivatives to determine the possibility of hydrogen bonding or salt bridge formation between various quinazoline derivatives and the active site of NS1. The criteria for hydrogen bond interaction

used, when the distance between the hydrogen and the heteroatom was within the range of 2.5-3.5 Å and the bond angle was at 109º-110º. Each docking experiment consisted of 10 docking runs with 150 individuals and 500,000 energy evaluations. Other parameters were left to their default values.

With the assumption that the comparison of docking results obtained for most active and least active compounds from the series will give better structure based understanding, compound therefore compound 29,28 (higher binding) and 5, 9 (lower binding) were used for docking studies. Figure 3 contains best docking pose of each selected molecule.

2.4 QSAR strategy:

Correlation matrix was constructed to check correlation among the variables (descriptors) [6-9]. The QUIK rule was used to discard models with high predictor collinearity which might lead to chancy correlation. This rule is based on the K-multivariate correlation index which measures the total correlation among the variables, first in between the predictor variables (Kx), then the response variable is also added to this matrix and the correlation is recalculated (Kxy).

Table 1. Correlation matrix for biological activity and used descriptors

	p(%Binding at 50 µM)	Mor25u	G3u
	1.000		
Mor25u	0.625	1.000	
G3u	0.106	-0.255	1.000
HATS5p	0.561	-0.077	-0.070

The analysis of correlation matrix confirms that there is no correlation among the used descriptors. To establish mathematical correlation between the biological activity and descriptors, Multiple Linear Regression (MLR) was applied to build the models and the variables were selected using genetic algorithm as implemented in Weka 3.7. The optimum number of descriptors was found to be 3. The goodness of fit for each predictive model was evaluated by examining the square of correlation coefficient (R^2), adjusted R^2, the standard deviation (s), predictive R^2 and Y-randomization.

RESULTS AND DISCUSSIONS

Docking analysis: From figure 3 it is clear that the most probable reason behind the higher binding of molecule 29 are presence of H-bonding, hydrophobic and mild polar interactions with the receptor, whereas other molecules either lacks H-bonding or additional hydrophobic or a combination of both. Comparison of molecule 29 with molecule 9 indicates that even though both involves of H-bonding but molecule 29 is more tightly bound to receptor this means the water molecule (HOH 247) in the cavity of protein is very important.

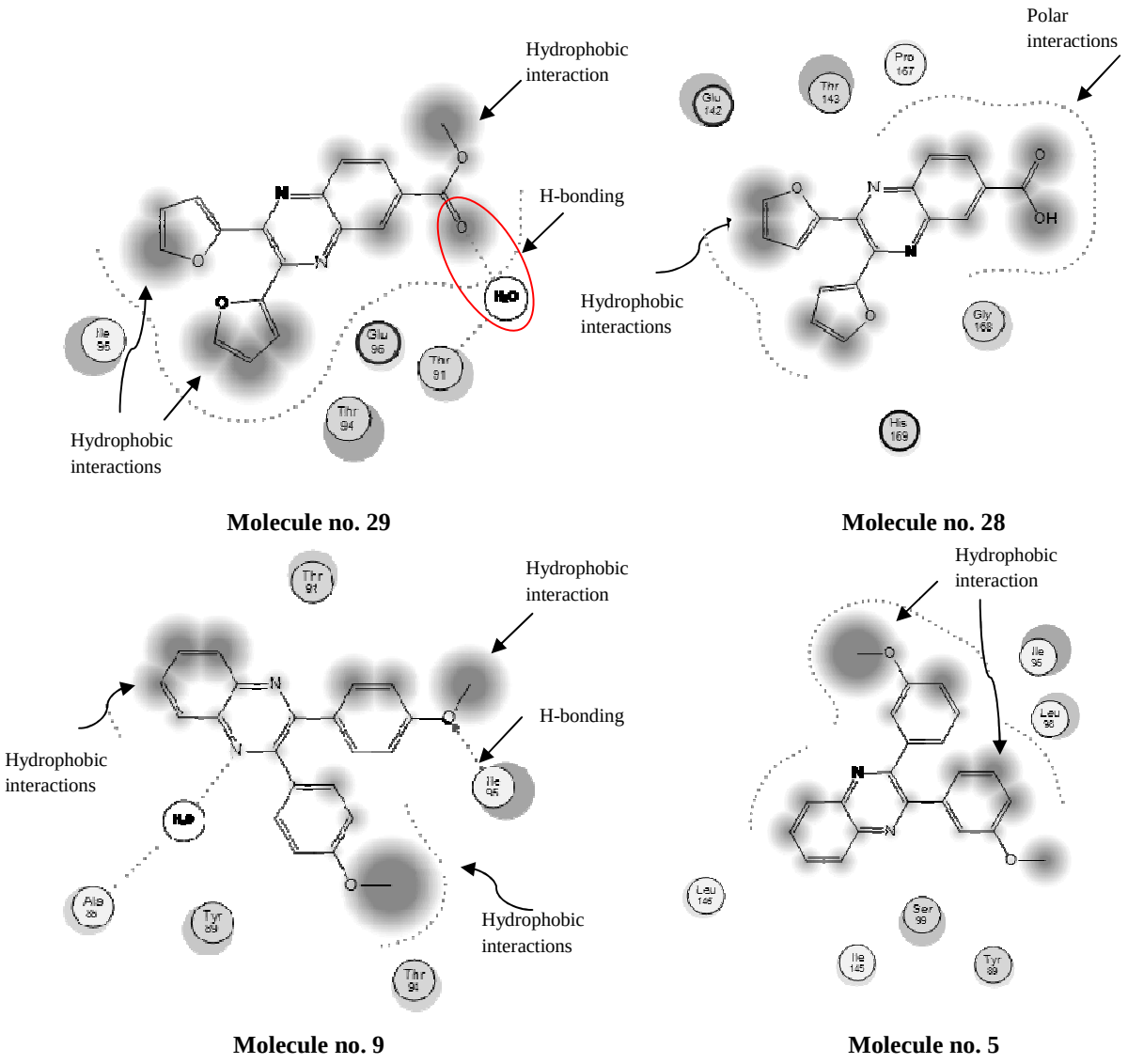

Fig. 3: The best docking pose of each selected molecule 29, 28, 9 and 5

QSAR analysis:

The best QSAR model based on three descriptors as follows along with the interpretation of QSAR model in terms of the specific contribution of substituent's and other molecular features to the modeled activity:

p(% Binding at 50 µM) = 0.701 x Mor25u + 1.091 x G3u + 37.727 x HATS5p - 5.961

N= 30, R^2 = 0.874, adj. R^2 = 0.859, pred. R^2 = 0.805, F value = 60.196

Where N is number of compounds in data set, R is the correlation coefficient, R^2 is the coefficient of determination, adj. $R^2$is adjusted coefficient of determination.

Deriving 3-parametric equations from 30 molecules may be done by chance. Therefore, in order to prove that the model is not chancy we have calculated R^2pred and adj. R^2 also. The rationale for using adjusted R^2 is that it varies with number of descriptors used and its value reduces with rise in the number of redundant descriptors. The high value of R, R^2 and pred. R^2 indicates that model has excellent statistical significance. Moreover the value of adj.R^2 which is considered as better parameter to judge the predictive power compared to R^2, is close to the value of R^2 thereby validating the high predictive power of model[6-9].

The model suggests that the binding of drug is directly related with Mor25u, G3u and HATS5p. Mor25u is 3D MoRSE descriptor and corresponds to 3D MoRSE signal 25/unweighted[10]. 3D-MoRSE descriptors are based on the idea of obtaining information from 3D atomic coordinates by the transform used in electron diffraction studies for preparing theoretical scattering curves. The positive coefficient of Mor25u indicates that increase in its value is positive factor for biologic activity.

Table 2. Actual and Predicted values of % Binding at 50 µM

Compound No.	% Binding at 50 µM Experimental	p(% Binding at 50 µM) Experimental	p(% Binding at 50 µM) Predicted	Residual values
1	1.8	-0.2552	-0.26555	0.01027
2	4.5	-0.6532	-0.40801	-0.24519
3	10.9	-1.0374	-1.01363	-0.02379
4	10.9	-1.0374	-0.99625	-0.04116
5	25	-1.3979	-1.44544	0.04750
6	28.7	-1.4578	-1.46565	0.00777
7	42.8	-1.6314	-1.46860	-0.16283
8	10.2	-1.0086	-1.07749	0.06889
9	5	-0.6989	-0.32848	-0.37048
10	38.8	-1.5888	-1.33635	-0.25247
11	13.3	-1.1238	-1.57190	0.44805
14	19.5	-1.2900	-1.38141	0.09138
15	1.5	-0.1760	-0.28758	0.11149
16	23.2	-1.3654	-1.26116	-0.10432
17	56.6	-1.7528	-1.94465	0.19183
18	7.3	-0.8633	-0.96705	0.10373
19	54.3	-1.7348	-1.39036	-0.34443
20	60.9	-1.7846	-1.57175	-0.21286
21	76	-1.8808	-1.93577	0.05496
22	4.4	-0.6434	-0.70746	0.06400
23	25.5	-1.4065	-1.53784	0.13130
24	15.7	-1.1959	-1.10446	-0.09143
25	7.8	-0.8920	-0.85520	-0.03688
26	2.8	-0.4471	-0.65361	0.20645
27	6.7	-0.8260	-0.67573	-0.15034
28	1.4	-0.1461	-0.29616	0.15003
29	7.2	-0.8573	-0.68512	-0.17220
30	9.1	-0.9590	-0.82493	-0.13410
32	1.5	-0.1760	-0.60803	0.43193
33	0.5	0.30103	0.07812	0.22290

Symmetry of the molecule is included in variable G3u. It is among the WHIM descriptors and corresponds to 3st component symmetry directional WHIM index/ unweighted[10]. This means G3u is a directional WHIM symmetry descriptor which encodes the symmetry along the third component. The positive coefficient of G3u indicates that increase in its value increases biologic activity.

The third descriptor HATS5p is GETAWAY descriptor and corresponds to leverage-weighted autocorrelations of lag5/ weighted by atomic polarizabilities[10]. For good biological activity, the compounds should have atoms at a topological distance of 5 with different polarizability as a tendency. This means that one atom *i* should have polarizability greater than the polarizability of the molecule and the other atom *j* should exhibit the reverse.

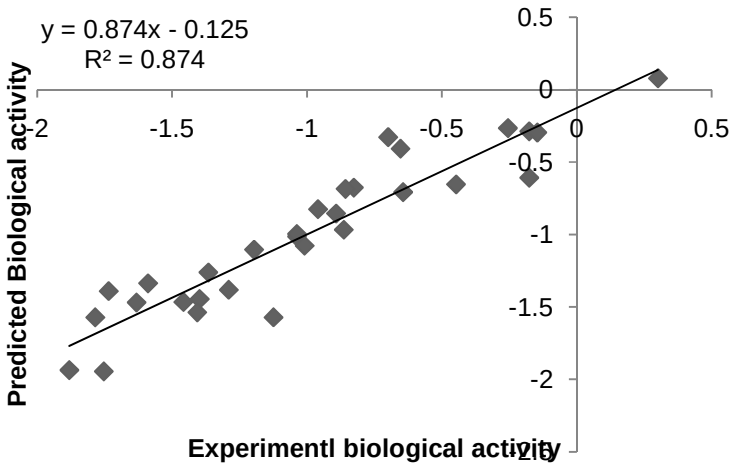

Fig.5. Experimental vs. Predicted % Binding at 50 µM

From figure 5, it is clear that there is good relation between the experimental and predicted % Binding at 50 µM, in addition the graph between experimental biological activity and residual is with good scattering of points thereby indicating statistical stability of model[6-9]. For evaluation, we performed Y-randomization also.

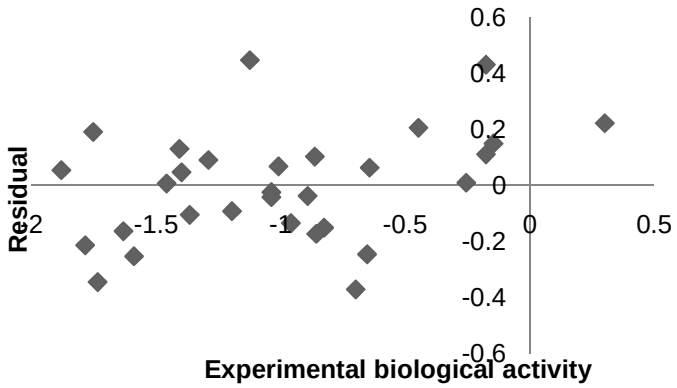

Fig.6. Experimental % Binding at 50 µM vs. residual

Y-Randomization test:
The robustness of a given QSAR model can be established using Y-Randomization[6-9]. In Y-randomization, dependent variable (%Binding at 50 µM in present study) is shuffled randomly and a new QSAR model is constructed using the original independent variables. If the new QSAR models have lower R^2 values for several trials, then the given QSAR model is thought to be robust. Thus Y-randomization is useful to avoid any chancy correlation between dependent variable vector and independent variables. The model has lower R^2 even after many Y-randomizations.

CONCLUSIONS

From the result and discussion it is clear that (1) molecule 29 is tightly bound to receptor because of H-bonding, polar and hydrophobic interactions whereas other molecules either lacks H-bonding or hydrophobic interactions. (2) The molecule no. HOH 247 present inside the cavity of receptor play crucial role. (3) Only three 3D descriptors are sufficient for predicting the activity.(45) the derived model is statistically reliable and non-chancy.(5) the biologic activity depends upon 3D variables Mor25u, G3u and HATS5p.

Acknowledgement
We are thankful to Dr. F. C. Raghuwanshi for his support and Sumit O. Bajaj, C. Eugene Bennett Department of Chemistry, West Virginia University, Morgantown, West Virginia 26506-6045, USA for providing useful data and helpful discussions.

REFERENCES

[1] I. M. Lagoja, E. De Clercq, *Med. Res. Rev*. **2008**, 28, 1-38
[2] J. K. Taubenberger, D. M. Morens, *Annu. Rev. Pathol. Mech. Dis*., **2008**, 3, 499-522.
[3] Lei You, Eun Jeong Cho, John Leavitt, Li-Chung Ma, Gaetano T. Montelione, Eric V. Anslyn, Robert M. Krug, Andrew Ellington, Jon D. Robertus, *Bioorg. Med. Chem. Lett*., **2011**, 21: 3007–3011 and references cited therein.
[4] Rahul D. Jawarkar, Vijay H. Masand, Komalsingh N. Patil, Devidas T. Mahajan, Moulay Hfid Youssoufi, Taibi Ben Hadda, Santosh L. Kumbhare, *Der Pharma Chemica*, **2010**, 2(6): 302-310
[5] Rahul D. Jawarkar, Devidas T. Mahajan, Vijay H. Masand, Taibi Ben Hadda, G. H. Kurhade, *Der Pharmacia Lettre*, **2010**, 2(6):350-357
[6] Anil M. Manikrao, Pravin N. Khatale, Rahul D. Jawarkar, J. V. Vyas, D. T. Mahajan, Vijay H. Masand, Taibi ben hadda, *J. Comput. Method. Mol. Design*, **2011**, 1 (2):9-13
[7] G. M. Nazeruddin, S. B. Suryawanshi, Hassan A. Osman, Vijay H. Masand, Komalsingh N. Patil, *J. Chem. Pharm. Res*., **2010**, 2(3):544-550
[8] Vijay H. Masand, Komalsingh N. Patil, Devidas T. Mahajan, Rahul D. Jawarkar, Ghulam M. Nazerruddin, *Der Pharma Chemica*, **2010**, 2(4): 298-308
[9] V. H. Masand, K. N. Patil, G. M. Nazerruddin, R. D. Jawarkar, S. O Bajaj., *Org. Chem.: An Ind. J.*, **2010**, 6(1), 30-38
[10] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, first ed. Wiley-VCH, Mannheim, **2000**