3D Qsar Studies on a Series Of-[(1-Benzyl-1H-Imidazol-5-yl)-alkyl]-Amino Derivatives as Angiotensin II AT$_1$ Antagonists

Mukesh Chandra Sharma and Dharm Veer Kohli

Department of Pharmaceutical Sciences,
Dr. H.S. Gour University, Sagar (M.P.) 470 003, India

Abstract: Three dimensional quantitative structure activity relationship (3D QSAR) investigations were carried out on a series of - [(1-Benzyl-1H-Imidazol-5-yl)-alkyl]-amino derivatives for their Angiotensin II AT$_1$ Antagonists activity. The structures of all compounds were built on a workspace of VlifeMDS3.5 molecular modelling software and 3D QSAR models were generated by applying a partial least square (PLS) linear regression analysis coupled with a GA variable selection method. kNNMFA calculations for electrostatic, steric and hydrophobic field were carried out. The master grid maps derived from the best model has been used to display the contribution of electrostatic potential and steric field. The statistical results showed significant correlation coefficient q^2 (0.6275), r^2 for external test set (pred_r^2) 0.7531. These results should serve as a guideline in designing more potent and selective antihypertensive molecules.

Key words: Angiotensin II (Ang II) · kNNMFA · AT$_1$, VLife MDS
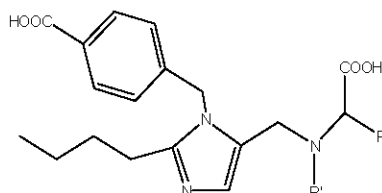
INTRODUCTION

The renin-angiotensin system (RAS) plays a key role in regulating cardiovascular homeostasis and electrolyte/ fluid balance in normotensive and hypertensive subjects. [1] Activation of the renin-angiotensin cascade begins with renin secretion from the juxtaglomerular apparatus of the kidney and culminates in the formation of the octapeptide angiotensin II (AII), which then interacts with specific receptors present in different tissues[2] .Two basic types of receptors, both having a broad distribution, have been characterized so far: the AT1 receptor, responsible for the majority of effects attributed to this peptide and the AT2 receptor, with a functional role yet uncertain. [3] Angiotensin II is implicated in the pathogenesis of essential hypertension, reno-vascular hypertension, congestive heart failure and other renal diseases associated with albuminuria in the renin-angiotensin system [4,5]. Blockade of the renin-angiotensin system with ACE inhibitors has provided effective treatment of these conditions; however, some of the adverse effects of ACE inhibitors appear to be unrelated to angiotensin II blockade. The effects like cough and angio-edema are due to other pathways of ACE inhibition, such as degradation of bradykinins and

prostaglandins[6]. Newly reported method k-Nearest Neighbor Molecular Field Analysis (k-NN MFA) adopts a k-Nearest Neighbor principle for generating relationship of molecular fields with the experimentally reported activity [7]. This method utilizes the active analogue principle that lies at the foundation of medicinal chemistry 10. A number of quantitative structure-activity relationship (QSAR) studies related to design of angiotensin II receptor antagonists drugs have also been reported [8-12]. Development of quantitative structure-activity relationship by aid of various physicochemical parameters has been an important task in lead optimization and has further facilitated in design novel therapeutic agents.

Experimental

Methodology: The Ang II receptor antagonistic activity data of D and L-N-[(1-Benzyl-1H-Imidazol-5-yl)-alkyl]-amino Acids as Angiotensin II AT$_1$ Antagonists were taken from the reported work [13] (Table 1). The biological activity data (IC$_{50}$ in nM) was converted to negative logarithmic mole dose (pIC$_{50}$) for quantitative structure activity relationship (QSAR) analysis. The observed and predicted biological activities of the training and test set molecules are presented in Table 1.

Corresponding Author: M.C. Sharma, Department of Pharmaceutical Sciences Dr. Hari Singh Gaur University,
Sagar. (M.P) 470003 India. E-mail: mukeshcsharma@yahoo.com.

Table 1: Structure and activities of N-[(1-Benzyl-1H-Imidazol-5-yl)-alkyl]-amino Acids

Compd	R_1	R_2	IC_{50}(nM)[a]	Log IC_{50}
1	H	H	298	2.47
2*	$C_6H_5CH_2$	H	517	2.71
3	$C_6H_5CH_2$	H	3.8	0.579
4	$C_6H_5CHCH_3$	H	176	2.24
5*	$C_6H_5CHCH_3$	H	28	1.44
6	2-thienyl-CH_2	H	673	2.82
7	2-thienyl-CH_2	H	36.7	1.56
8	3-indolyl-CH_2	H	48.3	1.68
9	3-indolyl-CH_2	H	2.9	0.46
10	H	$C_6H_5CH_2$	31	1.49
11*	H	2-thienyl-CH_2	169	2.22

*Test compound

The observed selection of test set molecules was made by considering the fact that test set molecules represents a range of biological activity similar to the training set. We hereby report the models, as generated by kNN-MFA in conjunction with stepwise (SW) forward-backward variable selection methods. In the kNN-MFA method, several models were generated for the selected members of training and test sets and the corresponding best models are reported herein. VLife Molecular Design Suite [14] allows user to choose probe, grid size and grid interval for the generation of descriptors. The variable selection methods along with the corresponding parameters are allowed to be chosen and optimum models are generated by maximizing q^2. k-nearest neighbor molecular field analysis (kNN-MFA) requires suitable alignment of given set of molecules. This is followed by generation of a common rectangular grid around the molecules. The steric and electrostatic interaction energies are computed at the lattice points of the grid using a methyl probe of charge +1. These interaction energy values are considered for relationship generation and utilized as descriptors to decide nearness between molecules. The optimal training and test sets were generated using the sphere exclusion algorithm [15].This algorithm allows the construction of training sets covering descriptor space occupied by representative points. Therefore it can be said that the predictive abilities of SW kNN MFA model is good. To derive the kNN-MFA descriptor fields, a 3D cubic lattice with grid spacing of 2 A° in x, y and z dimensions was created to encompass the aligned molecules. kNNMFA descriptors were calculated using an sp^3 carbon probe atom with a Vander Waals

radius of 1.52 A° and a charge of 1.0 with default cut-off energy 30 kcal/mol to generate steric field energies and electrostatic fields.All the structure tables 1 were constructed using the 2D draw application provided as a tool of main MDS window. The 2D structures were converted to 3D structure by exporting them to MDS 3D mode. Energy minimization and geometry optimization was conduct using MMFF method with RMS gradient set to 0.001 and iteration limit to 10000.Alignment of all the 11 compounds was done using template based alignment in MDS; the aligned structures were used for 3D QSAR studies. In template based alignment method, a template structure was defined and used as a basis for alignment of a set of molecules. Following molecule was the template used for template based alignment as it was common to all structures. All the 11 molecules were aligned with the template in this study and were used in 3D QSAR study.

Calculation of Field Descriptor Values: For calculation of field descriptor values, both electro static and steric and hydrophobic field type with cutoffs 10.0 and 30.0 kcal/mole respectively were selected and charge type was selected as Gasteiger-Marsili. Dielectric constant was set to 1.0 considering the distance dependent dielectric function. Probe setting was carbon atom with charge 1.0 and grid setting.

Nearest Neighbor (kNN) Method: The kNN methodology relies on a simple distance learning approach whereby an unknown member is classified according to the majority of its k-nearest neighbors in the training set.

The nearness is measured by an appropriate distance metric (e.g. a molecular similarity measure calculated using field interactions of molecular structures). The standard kNN method is implemented simply as follows: Calculate distances between an unknown object (u) and all the objects in the training set; select k objects from the training set most similar to object u, according to the calculated distances; and classify object u with the group to which the majority of the k objects belongs. An optimal k value is selected by optimization through the classification of a test set of samples or by leave-one out cross-validation.

kNN-MFA with Genetic Algorithm: The genetic algorithm (GA) [16] algorithm offers a new approach to the problem of building quantitative structure-activity relationship (QSAR) and quantitative structure-property relationship (QSPR) models. Replacing regression analysis with the GFA algorithm allows the construction of models competitive with or superior to those produced by standard techniques and makes available additional information not provided by other techniques. Unlike most other analysis algorithms, GFA as a result provides multiple models and the population of the model is created by evolving random initial model using a genetic algorithm. Genetic algorithms are derived from an analogy with the spread of mutations in a population. In this analogy, ''individuals'' are represented as a one-dimensional string of bits. An initial population of individuals is created, usually with random initial bits.

Cross-Validation Using Weighted K-Nearest Neighbor: This is done to test the internal stability and predictive ability of the QSAR models. Developed QSAR models were validated by the following procedure:

Internal and External Validation: A molecule in the training set was eliminated and its biological activity was predicted as the weighted average activity of the k most similar molecules (eq.1). The similarities were evaluated as the inverse of Euclidean distances between molecules (eq.2) using only the subset of descriptors corresponding to the current trial solution.

$$wi = \frac{Exp(-dj)}{\sum Exp(-d)} \qquad \begin{array}{c} \text{k-Nearest neighbor} \\ \hat{y}_i = \sum Wi\, y_i \end{array} \qquad (1)$$

$$d_{if} = \sum_{k=1}^{vn}\left(X_{ik} - X_{jk}\right)^2 \Big/^{1}\!\!\Big/_2 \qquad (2)$$

- Step 1 was repeated until every molecule in the training set has been eliminated and its activity predicted once.
- The cross-validated r^2 (q^2) value was calculated using eq. 3, where yi and yi are the actual and predicted activities of the ith molecule, respectively and $y$mean is the average k-Nearest neighbor activity of all molecules in the training set. Both summations are over all molecules in the training set. Since the calculation of the pair wise molecular similarities and hence the predictions, were based upon the current trial solution, the q^2 obtained is indicative of the predictive power of the current kNN-MFA model.

$$q^2 = 1 - \frac{\sum\left(y_i - \hat{y}_i\right)^2}{\sum\left(y_i - y_{mean}\right)^2}$$

External Validation: The predicted r^2 (pred_r^2) value was calculated using eq. 4, where yi and yi are the actual and predicted activities of the ith molecule in test set, respectively and y mean is the average activity of all molecules in the training set. Both summations are over all molecules in the test set. The pred_r^2 value is indicative of the predictive power of the current kNN-MFA model for external test set.

$$pred\_r^2 = 1 - \frac{\sum\left(y_i - \hat{y}_i\right)^2}{\sum\left(y_i - y_{mean}\right)^2}$$

RESULT AND DISCUSSION

This method utilizes the active analogue principle that lies at the foundation of medicinal chemistry [17-19]. The developed QSAR models are evaluated using the following statistical measures: n, (the number of compounds in regression); k, (number of variables); DF, (degree of freedom); optimum component, (number of optimum PLS components in the model); r^2 (the squared correlation coefficient), F test (Fischer's value) for statistical significance, q^2 (cross-validated correlation coefficient); pred_r^2, (r^2 for external test set); Z score, (Z score calculated by the randomization test); best_ran_q^2, (highest q^2 value in the randomization test); best_ran_r^2, (highest r^2 value in the randomization test). The regression coefficient r^2 is a relative measure of fit by the regression equation. It represents the part of the variation in the observed data that is explained by the regression.
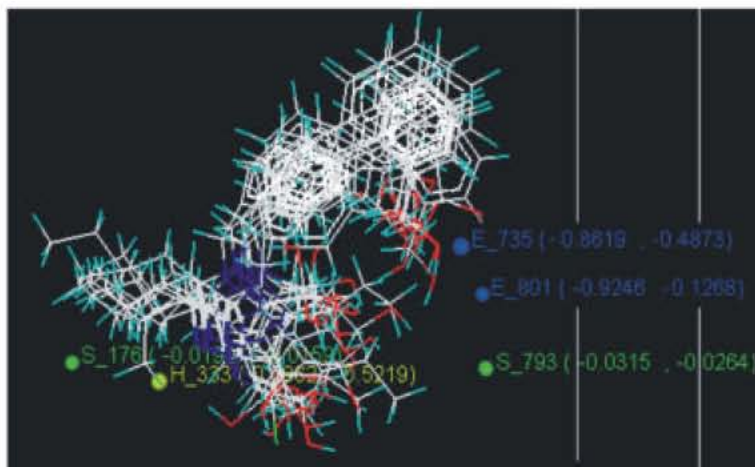
Fig. 1: Contribution plot of interactions GA kNNMFA model

Table 2: Calculated and Predicted pIC_{50} (by LOO method)

Observed activity	Predict activity model
2.47	3.18
2.71	1.63
0.579	0.983
2.24	3.09
1.44	2.08
2.82	4.19
1.56	1.04
1.68	1.18
0.46	0.84
1.49	1.89
2.22	2.90

However, a QSAR model is considered to be predictive, if the following conditions are satisfied: $r^2 > 0.6$, $q^2 > 0.6$ and pred_$r^2 > 0.5$. In the present study, PLS coupled with stepwise variable selection method was used to develop 3D-QSAR models of -[(1-Benzyl-1H-Imidazol-5-yl)-alkyl]-amino Derivatives based on steric, electrostatic and hydrophobic fields. The total data set was divided into training and test sets using the sphere exclusion algorithm for diversity of the sampling procedure. The quality of the model was assessed by cross-validated q^2 in the training set and external validation was performed by calculating predictive r^2 (Pred_r^2) from the test-set compounds.

$pIC50 = 1.2654 - S\_145 (-0.0221 -0.0159) - E\_735 (-0.8619 - 0.4873) - E\_801 (-0.9246 -0.1268) - S\_793 (-0.0315, -0.0264) + H\_333 (0.2165, 0.5219) -0.8753$

$N_{training} = 11$, Optimum Components = 4, DF = 18, $r^2 = 0.7612$, $q^2 = 0.6275$, F test = 23.576, r^2_se = 0.6521, q^2_se = 0.2178, pred_$r^2 = 0.7531$, pred_r^2se = 0.3276, ZScore $Q^{\wedge 2} = 0.6598$, Best Rand $Q^{\wedge 2} = 0.2785$

Here, n represents number of observations, df is the degrees of freedom, r is the square root of the multiple R-squared for regression, $q2$ is the cross-validated r 2 and F is the F-statistic for the regression model. S_145, E_735, E_801, S_793 and H_333 are the steric and electrostatic field, hydrophobic energy of interactions between probe (CH_3) and compounds at their corresponding spatial grid points of 145, 735, 801, 793 and 333. GA kNN-MFA method have better q2 (0. 0.6275) and pred_r2 (0.7531) model validation correctly predicts activity 62.7% and 75.3% for the training and test set respectively. The randomization test suggests that the developed model have a probability of less than 1% that the model is generated by chance. It uses 2 steric descriptors with 4 k nearest neighbor to evaluate activity of new molecule, so model generated by GA kNN-MFA are best model. In the QSAR model, steric descriptors with negative coefficients represent regions of less steric tolerance; bulky substituent is favorable in this region. Steric descriptors with negative coefficients indicate regions where bulky substituent is favored. Electrostatic field descriptors whereas negative coefficient indicates that electronegative (electron-rich or electron-donating) groups are favorable in this region. From 3D-QSAR model and Fig. 1 it is observed that electrostatic field with negative coefficient (E_735, E_801) is -[(1-Benzyl-1H-Imidazol-5-yl)-alkyl]-amino, indicating that electronegative groups are unfavorable on this site and presence of electronegative groups decrease the activity of -[(1-Benzyl-1H-Imidazol-5-yl)-alkyl]-amino compounds. Positive contribution of H_333 to Hydrogen group nearer to R respectively indicates that positive hydrophobic field is favorable for increasing the activity. Hence less

hydrophobic or more hydrophilic substituent groups R is preferred. It is known that the CoMFA method provides significant value in terms of a new molecule design, when contours of the PLS coefficients are visualized for the set of molecules. Similarly, the kNN-MFA models provide direction for the design of new molecules in a rather convenient way. Negative value in electrostatic field descriptors indicates that negative electronic potential is required to increase activity and more electronegative substituents group is preferred in that position, positive range indicates that group that imparting positive electrostatic potential is favorable for activity so less electronegative group is preferred in that region. Negative range in steric descriptors indicates that negative steric potential is favorable for activity and less bulky substituents group is preferred in that region, Positive value of steric descriptors reveals that positive steric potential is favorable for increase in activity and more bulky group is preferred in that region.

ACKNOWLEDGEMENT

REFERENCES

1. Ferrario, C.M., 1990. The Renin-Angiotensin System: Importance in Physiology and Pathology. J. Cardiovasc. Pharmacol., 15(3): 51-55.
2. Vallotton, M.B., 1987. The Renin-Angiotensin System. Trends. Pharmacol. Sci., 8: 69.
3. Nahmias, C. and A.D. Strosberg, 1995. The angiotensin AT_2 Receptor Searching for Signal-Transduction Pathways and Physiological Function. Trends. Pharmacol. Sci., 16: 223-225.
4. Burnier, M., 2001. Angiotensin II type 1 receptor blockers. Circulation., 103: 904-912.
5. Burnier, M. and H.R. Brunner, 2003. Angiotensin II receptor antagonists. Lancet., 355: 637-645.
6. Rodgers, J.E. and J.H. Patterson, 2001. Angiotensin II-receptor blockers: clinical relevance and therapeutic role. Am. J. Health. Syst. Pharm., 58: 671-683
7. Ajmani, S., K. Jhadav and S.A. Kulkarni, 2006. Three-dimensional QSAR using k-nearest neighbor method and its interpretation. J. Chem. Inf. Model., 46: 24-31.
8. Belvisi, L., G. Bravi. G. Catalano, M. Mabiliab, A. Salimbeni and C. Scolastico, 1996. A 3D QSAR CoMFA study of nonpeptide angiotensin II receptor antagonists. J. Comput. Aided. Mol. Des., 10: 567-582.
9. Sharma, M.C., S. Sharma, N.K. Sahu and D.V. Kohli, 2011. QSAR Studies of some Substituted imidazolinones Derivatives angiotensin II receptor antagonists using Partial Least Squares Regression (PLSR) Based Feature Selection. Jour. Saud. Chem. Soc., (In press)
10. Sharma, M.C., S. Sharma, N.K. Sahu and D.V. Kohli, 2011. QSAR Studies of some Substituted imidazolinones Derivatives angiotensin II receptor antagonists using Partial Least Squares Regression (PLSR) Based Feature Selection. Jour. Saud. Chem. Soc., (In press)
11. Sharma, M.C., D.V. Kohli, S.C. Chaturvedi and S. Sharma, 2009. Molecular Modeling Studies of Some substituted 2-butylbenzimidazoles angiotensin II receptor antagonists as antihypertensive agents. Digest. J. Nanomat. Biostruct., 4(4): 843-856.
12. Yoo, S.E., S.K. Kim, S.H. Lee, K.Y. Yi and D.W. Lee, 1999. A comparative molecular field analysis and molecular modeling studies on pyridylimidazole type of angiotensin II antagonists. Bioorg. Med. Chem., 7: 2971-2976.
13. David, T., G.R. Hill, J.W. Girard, M. Richard, E.F. Edwards, E.O. Weidley, E. Catherine, E.B. Peishoff and A. Nambi, 1995. Bioorg. Med. Chem. Lett., 5(1): 19-24.
14. Vlife MDS software package, version 3.5, supplied by Vlife science technologies Pvt. Ltd, 1, Akshay 50, Anand park, Aundh, Pune, India, 411007.
15. Golbraikh, A. and A. Tropsha, 2002. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. J. Comp. Aided. Mol. Design., 16: 357-369.
16. Holland, J.H., 1992. Genetic Algorithms, Sci. Am., 267: 66-72.
17. Sharma, M.C and S. Sharma, 2010. 3D- Quantitative Structure-Activity Relationship Analysis of Some 2-Substituted Halogen benzimidazoles Analogues with Antimycobacterial activity. Int. J. Chem. Tech. Res., 2(1): 606-614.

18. Sharma, M.C. S. Sharma. D.V. Kohli and S.C. Chaturvedi, 2010. Three Dimensional Quantitative Structural-Activity Relationship (3D-QSAR) Studies some 3-{4-[3-(2-aryl-phenoxy) butoxy]-phenyl} Propionic acids as novel PPAR γ/δ Agonists. Der. Pharma. Chemica., 2(1): 82-90.

19. Sharma, M.C., S. Sharma. D.V. Kohli and S.C. Chaturvedi, 2010. QSAR and k-Nearest Neighbour Molecular Field Analysis (k-NN MFA) Classification Analysis of Studies of Some Bemzimidazoles Derivatives Antibacterial activity Against Escherichia coli. Der. Pharmacia. Lett., 2 (1): 150-161