Volume 5, Issue 1, January 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

3D-QSAR Study of Novel Tyrosine Inhibitory Aganist Matrix Metalloproteinase and Histone Deacetylase

Fatima Zohra Tassine*, Menana El Hallaoui
LIMME, Departement of Chimistry Faculty of Sciences Dhar El Mahraz, University Sidi Mohammed Ben Abd Ellah, B.P. 1796 Atlas. FES, Morocco

Abstract— The objective of the present work is to use artificial neural network (NN) to study the quantitative structure activity relationship (QSAR) of L-tyrosine derivatives, principal inhibitors of matrix metalloproteinases 2 (MMP-2) and histone deacetylase 8 (HDAC-8). A data set of 31 molecules belonging to the L-tyrosine family and represented by 9 descriptors (calculated using Molecular Modeling Pro plus MMP+) was analyzed. QSAR Models were built with Multiple Linear Regression (MLR) and NN methods. Good correlation is obtained both for Inhibitory activities against MMP-2 (rMLR=0.93, rNN=0.96) as for Inhibitory activities against HDAC-8 (rMLR=0.95, rNN=0.99). The predicted values of activities are in good agreement with the experimental results. The neural network models established with NN techniques considering the relevant descriptors selected by MLR, are statistically significant and show very good stability towards data variation with leave-one-out (LOO) cross validation (r CVMMP-2 =0.92 and r CV HDAC-8=0.95).

Keywords— L-tyrosine, MMP-2 inhibition, HDAC-8 inhibition, 3D-QSAR, MLR, NN, LOO.

I. INTRODUCTION

Cancer is one of the leading causes of disease and mortality worldwide [1]. Studies conducted over more than 40 years have revealed mounting evidence supporting that extracellular matrix remodeling proteinases, such as matrix metalloproteinases (MMPs), are the principal mediators of the alterations observed in the microenvironment during cancer progression [2,3].

The matrix metalloproteinases (MMPs) are a family of structurally related zinc-dependent endoproteinases that degrade and remodel structural proteins in the extracellular matrix [4]. They include more than 20 subtypes, among which MMP-2 is highly involved in the process of tumor invasion and metastasis and has been considered as a promising target for cancer therapy [5, 6].

A number of natural and synthetic HDAC inhibitors have been reported, and in recent years the importance of HDAC inhibitors has increased due to their efficacy against many malignant diseases [7]. Several of these HDAC inhibitors inhibit tumor growth and many of them are under clinical trials [8-9]. Quantitative structure–activity relationships (QSAR) studies have been successfully applied for modeling the biological activities of natural and synthetic chemicals [10, 11]. In this light, we are interested in studying a new set of novel L-tyrosine derivatives for their inhibitory activities based on simple but statistically sound QSAR models whose parameters can be easily obtained using commonly available and less costly computational programs. The generated models were the primary basis in designing new structures with potentially greater bioactivity.

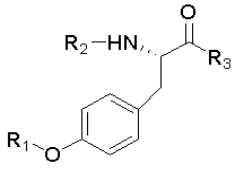
Herein, we report a quantitative structure-activity relationship (QSAR) study, which utilizes chemical properties obtained from quantum mechanical calculations (3-D parameters) and 2-D parameters known to derive predictive models from a series of 31 molecules belonging to the L-tyrosine family. The result of this work should facilitate further development of new selective L-tyrosine inhibitors.

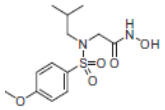
II. MATERIALS

A. Expérimental data

The L-tyrosine derivatives assayed for the inhibitory activities against MMP-2 and HDAC-8 [12] are included in Tables 1-2. The pharmacological activity IC50 has been expressed in µM]. In this work the activities MMP-2 and HDAC-8 are expressed in the logarithmic form (logIC50).

Table 1: studied compounds and their observed inhibitory activities against mmp-2 logic50 (obs), and calculated logic50 with mlr; nn and cv methods.



N°	R1	R2	R3	logIC50(Obs)	logIC50(MLR)	logIC50(NN)	logIC50(CV)
1	C6H5CH2	C6H5CO	OCH3	1,8	2,1	1,7	1,7
2	C6H5CH2	p-CH3C6H4CO	OCH3	1,9	1,9	1,9	1,9
3	C6H5CH2	CH3CO	OCH3	1,4	1,5	1,7	1,4
4	C6H5CH2	CH3SO2	OCH3	1,3	1,0	1,0	1,1
5	C6H5CH2	p-CH3C6H4SO2	OCH3	1,0	0,9	1,0	1,0
6	C6H5CH2CH2	C6H5CO	OCH3	1,9	1,1	1,5	1,4
7	C6H5CH2CH2	p-CH3C6H4CO	OCH3	1,9	1,4	1,7	1,6
8	C6H5CH2CH2	CH3CO	OCH3	1,1	1,1	1,7	1,0
9	C6H5CH2CH2	CH3SO2	OCH3	0,4	0,6	0,3	0,4
10	C6H5CH2CH2	p-CH3C6H4SO2	OCH3	0,8	0,8	0,9	1,0
11	C6H5CH2	C6H5CO	OH	1,0	0,8	1,1	0,7
12	C6H5CH2	p-CH3C6H4CO	OH	0,9	0,7	0,4	1,1
13	C6H5CH2	CH3CO	OH	0,5	0,7	0,4	0,4
14	C6H5CH2	CH3SO2	OH	0,1	0,1	0,1	0,2
15	C6H5CH2	p-CH3C6H4SO2	OH	0,4	0,0	0,0	0,3
16	C6H5CH2CH2	C6H4CO	OH	0,9	0,6	0,9	0,5
17	C6H5CH2CH2	p-CH3C6H4CO	OH	0,7	1,4	0,5	0,4
18	C6H5CH2CH2	CH3CO	OH	-0,1	0,2	-0,1	-0,6
19	C6H5CH2CH2	CH3SO2	OH	-0,9	-0,4	-0,6	-0,7
20	C6H5CH2CH2	p-CH3C6H4SO2	OH	-0,6	0,0	0,2	0,9
21	C6H5CH2	C6H5CO	NHOH	0,2	0,0	0,3	0,5
22	C6H5CH2	p-CH3C6H4CO	NHOH	-0,3	-0,5	0,0	-0,1
23	C6H5CH2	CH3CO	NHOH	-1,4	-1,3	-1,5	-1,3
24	C6H5CH2	CH3SO2	NHOH	-1,2	-1,8	-1,6	-1,5
25	C6H5CH2	p-CH3C6H4SO2	NHOH	-1,6	-0,8	-1,6	-1,6
26	C6H5CH2CH2	C6H5CO	NHOH	0,0	0,0	0,1	0,1
27	C6H5CH2CH2	p-CH3C6H4CO	NHOH	0,6	-0,2	-0,4	0,3
28	C6H5CH2CH2	CH3CO	NHOH	-1,9	-1,2	-1,7	-1,5
29	C6H5CH2CH2	CH3SO2	NHOH	-1,5	-2,0	-1,5	-1,5
30	C6H5CH2CH2	p-CH3C6H4SO2	NHOH	-1,8	-1,1	-1,3	-1,9
31 *				-1,9	-2,0	-1,7	-1,6

* NNGH (N-isobutyl-N-(4-methoxyphenylsulfonyl) glycyl hydroxamic acid), a potent inhibitor of MMP-2, was used as the positive control of MMP-2 assay.

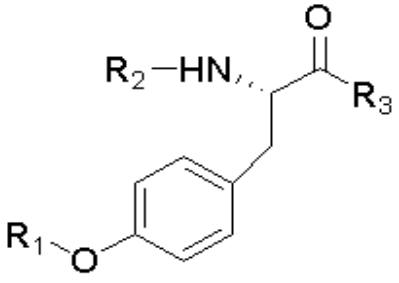
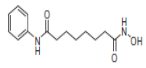


Table 2: studied compounds and their observed inhibitory activities against hdac-8 logic50 (obs), and calculated

N°	R1	R2	R3	logIC50(Obs)	logIC50(MLR)	logIC50(NN)	logIC50(CV)
1	C6H5CH2	C6H5CO	OCH3	3,0	2,7	2,8	2,8
2	C6H5CH2	p-CH3C6H4CO	OCH3	3,0	2,8	3,0	2,9
3	C6H5CH2	CH3CO	OCH3	2,9	2,6	2,7	2,7
4	C6H5CH2	CH3SO2	OCH3	2,9	2,9	2,9	2,9
5	C6H5CH2	p-	OCH3	3,0	3,5	3,0	3,0
6	C6H5CH2CH2	CH3C6H4SO2	OCH3	3,0	2,9	3,0	2,9
7	C6H5CH2CH2	C6H5CO	OCH3	3,0	2,9	3,0	2,9
8	C6H5CH2CH2	p-CH3C6H4CO	OCH3	2,9	2,8	2,9	2,9
9	C6H5CH2CH2	CH3CO	OCH3	3,0	3,3	3,1	3,1
10	C6H5CH2CH2	CH3SO2	OCH3	3,0	3,3	3,0	2,7
11	C6H5CH2	p-	OH	2,0	2,5	2,2	2,5
12	C6H5CH2	CH3C6H4SO2	OH	3,0	2,6	3,0	2,3
13	C6H5CH2	C6H4CO	OH	1,6	1,7	1,6	1,7
14	C6H5CH2	p-CH3C6H4CO	OH	1,7	1,5	1,7	1,7
15	C6H5CH2	CH3CO	OH	2,7	2,4	2,7	2,5
16	C6H5CH2CH2	CH3SO2	OH	2,4	2,4	2,6	2,6
17	C6H5CH2CH2	p-	OH	2,6	2,6	2,7	2,7
18	C6H5CH2CH2	CH3C6H4SO2	OH	2,1	1,9	2,0	2,3
19	C6H5CH2CH2	C6H5CO	OH	2,0	1,6	2,1	1,9
20	C6H5CH2CH2	p-CH3C6H4CO	OH	3,0	2,6	2,8	2,9
21	C6H5CH2	CH3CO	NHOH	1,0	1,4	1,2	1,2
22	C6H5CH2	CH3SO2	NHOH	1,5	1,6	1,3	1,3
23	C6H5CH2	p-	NHOH	0,8	0,9	1,0	0,7
24	C6H5CH2	CH3C6H4SO2	NHOH	0,6	0,4	0,6	0,7
25	C6H5CH2	C6H5CO	NHOH	1,6	1,8	1,7	2,0
26	C6H5CH2CH2	p-CH3C6H4CO	NHOH	1,4	1,4	1,4	1,3
27	C6H5CH2CH2	CH3CO	NHOH	1,7	1,8	1,5	1,4
28	C6H5CH2CH2	CH3SO2	NHOH	1,2	1,0	1,1	1,3
29	C6H5CH2CH2C	p-	NHOH	1,0	0,8	1,0	1,3
30	6H5CH2CH2	CH3C6H4SO2	NHOH	1,9	2,1	1,9	2,0
31 *		C6H5CO p-CH3C6H4CO CH3CO CH3SO2 p- CH3C6H4SO2		0,2	0,8	0,2	0,6

* SAHA (suberoylanilide hydroxamic acid), a potent HDACs inhibitor that has been registered as antitumor drug, was used as the positive control of HDAC-8 assay.

B. Descriptors

Molecular structures were drawn with ChemDraw Ultra, 8.0, Different types of descriptors, physicochemical, steric, geometrical, energetic and electronic, were calculated for the entire data set of 31 L-tyrosine derivatives (Table 3). Descriptors requiring 3D structures were calculated using MMP+ (Molecular Modeling Pro plus). 3D geometries were optimized with the molecular mechanics MM2 method incorporated in the referenced software [13].

Table 3: list of employed descriptors.

Descriptor symbol	Type of descriptor
BP	Boiling Point(K)
MW	Molecular Weight(Atomic Mass Units)
MR	Molar Refractivity
PC	Partition Coefficient (octanol/water)
BE	Bend Energy (Kcal/mol)
SE	Stretch Energy (Kcal /mol)
VDW	Van der Waals volume 1,4 energy (Kcal/mol)
DL	Dipole Length(ev)
RE	Repulsion Energy(ev)

III. METHODS

A. Multiple Linear Regressions

Multiple Linear Regression (MLR) is a supervised method that aims at establishing a mathematical relationship between a property of a given system and a set of molecular characteristics or descriptors that encode chemical information.

The success of a MLR approach in the development of interpretative and predictive model equations requires that data should be representative and homogeneous, the absence of redundancy among descriptors, a ratio between number of compounds in data set and number of descriptors of at least three and the fulfillment of strict validation procedures [14].

So For the purpose of this work, the Data set is submitted to MLR Analysis, using the software SYSTAT, version 12, so that, the proposed MLR model served primarily to select the descriptors used as the input parameters for a back propagation network (NN), and in the other hand it could be used to predict inhibitory activities logIC 50.

B. Neural Network (NN)

NN is artificial systems simulating the function of the human brain. Three components constitute a neural network: the processing elements or nodes, the topology of the connections between the nodes, and the learning rule by which new information is encoded in the network. While there are a number of different ANN models, the most frequently used type of ANN in QSAR is the three-layered feed-forward network [15].

The Multi-Layer Perceptron is a specific type of Feedforward Neural Network. The nodes are organized in layers (input, hidden and output layers) and each neuron is connected with one more nodes of the following layers only. There is a special type of node called ‘‘bias’’ which has no connection with neurons in the previous layers. It is used to shift the y-intercept value of the activation function for the next layers and therefore enhance the flexibility of the network. Some authors [16] have proposed a parameter ρ, leading to determine the number of hidden neurons, which plays a major role in determining the best ANN architecture defined as follows:

ρ = (Number of data points in the training set / Sum of the number of connections in the NN).

The output layer represents the calculated activity values log (IC50). In this work the ANN architectures for the inhibitory activities against MMP-2 and HDAC-8, respectively (5-4-1) and (4-4-1) are depicted in fig. 1 and fig. 2.

All calculations of NN are done on Matlab 7 using our program written in C language.

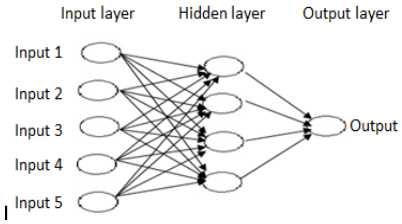


Fig. 1 Feedforward Multi-Layer Perceptron neural network for MMP-2 activities.

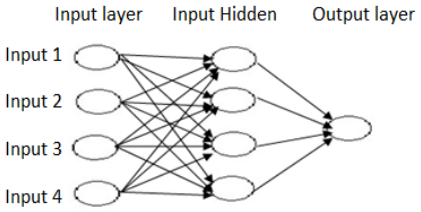


Fig. 2 Feedforward Multi-Layer Perceptron neural network for HDAC-8 activities.

C. Cross Validation

Cross-validation is a popular technique used to explore the reliability of statistical models. Based on this technique, a number of modified data sets are created by deleting in each case one or a small group, these proceedings are named respectively “leave-one-out” and “leave-some-out” [17-19]. For each data set, an input-output model is developed. The model is evaluated by measuring its accuracy in predicting the responses of the remaining data (the ones that have not been used in the development of the model). In this study we used, the leave-one-out (LOO) procedure.

IV. RESULTS AND DISCUSSION

A. Multiple Linear Regression (MLR)

The built QSAR model for Inhibitory activities against MMP-2 using multiple linear regression (MLR) method is represented by the following Equation:

$$\mathbf{LogIC_{50}=8,123-0,037MW+0,760PC+1,011BE-10,533SE+0,714VDW} \quad (1)$$

n = 31 **r** = 0.93 **s** = 0.47

Where n is the number of compounds, r is the correlation coefficient, s is the standard deviation

The t-values are shown in Table 4. The values of predicted logIC50 (log50-MLR) calculated from equation (1), and observed logIC50values (logIC50-obs) are given in Table 1. The correlation of predicted logIC50 and observed logIC50 is illustrated in fig. 2.

Table 4. T-values of regression equation (1) (ratio of the parameter's regression coefficient and the standard error) for inhibitory activities against mmp-2.

Descriptor	Coefficient	Standard error	t-value
MW	-0,037	0,004	8,969
PC	0,760	0,125	6,063
BE	1,011	0,206	4,900
SE	-10,533	1,639	-6,428
VDW	0,714	0,110	6,463

The correlation coefficient is r = 0.93, and the standard error s = 0.471.

The descriptors related to the Constitutional descriptors, Molecular Weight (MW), Partition Coefficient (PC), Bend Energy (BE), Stretch Energy (SE) and Van der Waals (VDW) are the most important parameters in the establishment of MMP-2 QSAR model for L-tyrosine derivatives.

For Inhibitory activities against HDAC-8, the built QSAR model using multiple linear regression (MLR) method is represented by the following Equation:

$$\mathbf{LogIC_{50}= 6,640-0,009BP-0,312BE-0,227DL \quad (2)}$$

n=31 r=0, 95 S=0, 29

Table 5. T-values of regression equation (2) (ratio of the parameter's regression coefficient and the standard error) for inhibitory activities against hdac-8.

Descriptor	Coefficient	Standard errer	t-value
BP	-0,009	0,001	-12,497
BE	-0,312	0,100	-3,118
DL	-0,227	-0,538	-7,479
RE	0,000	0,000	11,218

The correlation coefficient is r = 0,950, and the standard error s = 0,293.

We notice that the descriptors related to the Constitutional descriptors, Boiling Point (BP), Bend Energy (BE), Dipole Length (DL), and Repulsion Energy (RE) are the most important in the establishment of HDAC-8 QSAR model for L-tyrosine derivatives. The correlation of predicted logIC50 and observed logIC50 is illustrated in fig. 3.

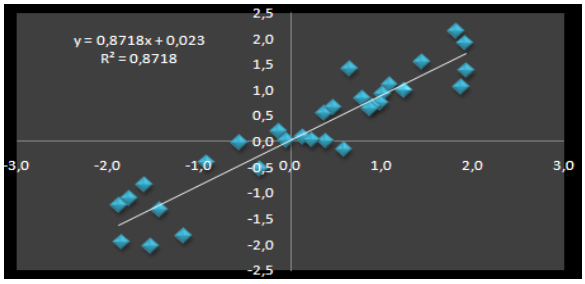


Fig. 2 The correlation of predicted logIC50 (logIC50-MLR) for Inhibitory activities against MMP-2 calculated Using Multiple Linear Regression and observed logIC50 (LogIC50-obs).

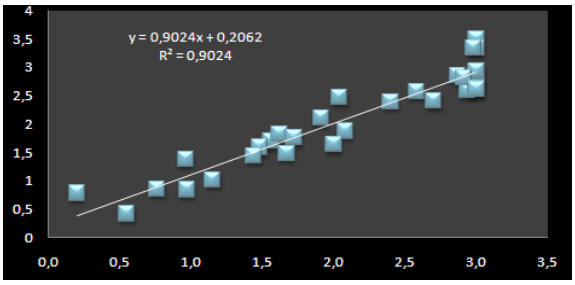


Fig. 3 The correlation of predicted logIC50 (logIC50-MLR) for Inhibitory activities against HDAC-8 calculated using Multiple Linear Regression and observed logIC50 (logIC50-obs).

A good correlation was obtained with MLR method (rMLR= 0, 930 for Inhibitory activities against MMP-2 and rMLR= 0.950 for Inhibitory activities against HDAC-8) so the predictive power of this two models is very significant.

B. Neural networks

Neural networks (NN) are used to generate predictive models of quantitative structure–activity relationships (QSAR) between a set of molecular descriptors obtained from the MLR analysis and observed activity. The correlation of the observed activities with the NN calculated ones, MMP-2 and HDAC-8, are illustrated in Fig. 4 and Fig. 5 respectively.

The correlation coefficient of Inhibitory activities against MMP-2, r = 0.961 and that of Inhibitory activities against HDAC-8, r = 0,990, show that the selected descriptors by LMR are pertinent and that models proposed to predict activity are relevant.

Fig. 4 and Fig. 5 displays a plot between observed activity (logIC50 (Obs)) and predicted activity (logIC50 (NN))

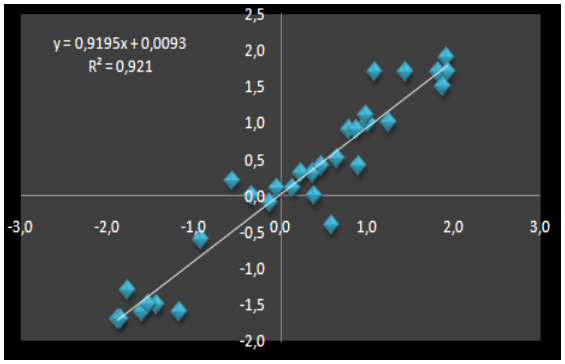


Fig. 4 The correlation of predicted logIC50 (logIC50-NN) for Inhibitory activities against MMP-2 calculated using Neural networks and observed logIC50 (logIC50-obs).

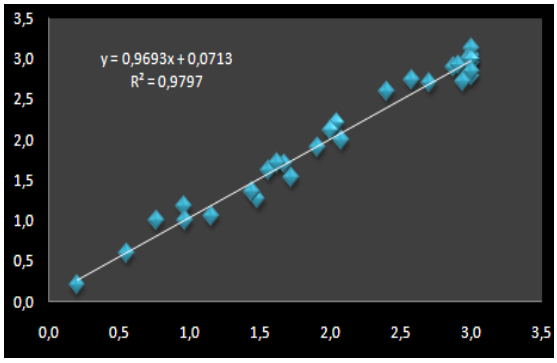


Fig. 5 The correlation of predicted logIC50 (logIC50-NN) for Inhibitory activities against HDAC-8 calculated using Neural networks and observed logIC50 (logIC50-obs).

C. Cross-Validation technique

Before using a QSAR model to predict the activity of new compounds, we should validate it using a validation method. In this paper we validated our model with cross validation using LOO procedure. The correlation of the observed activities with the CV calculated ones are illustrated in figure 6 and figure 7.

Fig. 6 and Fig. 7 display a plot between observed activity (logIC50 (Obs)) and predicted activity (logIC50 (CV))

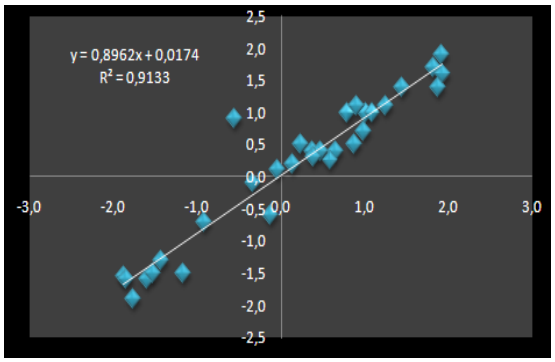


Fig. 6 The correlation of predicted logIC50 (logIC50-CV) for Inhibitory activities against MMP-2 calculated using Cross-Validation and observed logIC50 (logIC50-obs)

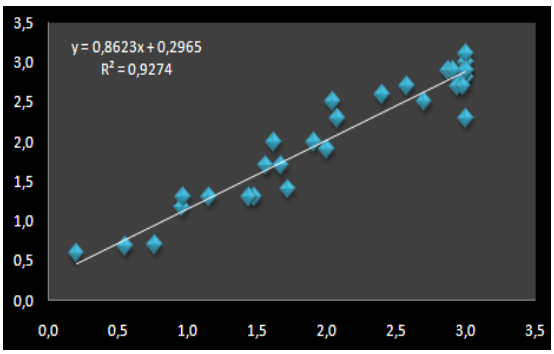


Fig.7 The correlation of predicted logIC50 (logIC50-CV) for Inhibitory activities against HDAC-8 calculated using Cross-Validation and observed logIC50 (logIC50-obs).

The good correlation obtained with cross validation (rcv=0.920 for Inhibitory activities against MMP-2 and r = 0.950 for Inhibitory activities against HDAC-8) confirm that the predictive power of this two models is very significant.

V. CONCLUSION

In this paper, we have explored a set of 31 L-tyrosine compounds inhibiting matrix metalloproteinases 2 (MMP-2) and histone deacetylase 8 (HDAC-8) using QSAR tools. Multiple linear regression and NN methodologies were employed to perform QSAR models. The analysis of the Inhibitory activities against MMP-2 model suggests that the descriptors representing high interest for activity are Partition Coefficient and van der Waals volume. However, inhibitory activities aganist HDAC-8 depend on the Repulsion Energy and Bend Energy. Both QSAR models proposed in this work (rcv MMP-2=0.920, rcv HDAC-8 = 0.950) are expected to be a useful tool in the conception of novel active molecules.

REFERENCES

[1] M. Bertalmio, G. Sapiro, C. Ballester and V. Caselles, SIAM News, vol 35, pp. 417–424, 2000.
[2] A. Criminisi, P. Perez, and K. Toyama, IEEE Transactions on Image Processing, vol 13, pp. 1200–1212, 2004.
[3] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, IEEE Transactions on Image Processing, vol 12, pp. 8, 2003
[4] Y. Hasan, and J. Karam, IEEE Transactions On Image Processing, vol 9, pp. 11, 2000.
[5] A .Eftychios, Pnevmatikakis, and P. Maragos, IEEE trans, vol 9, pp. 978, 2008.
[6] S.Bhuvaneswari, and T.Subashini, J. Computer Applications ,vol 6, pp. 975 – 8887, 2013.
[7] A. Pezeshk , and L. Tutwiler, IEEE Transactions on geosciences and remote sensing, vol 49, pp. 12, 2011.
[8] X. Liu and J. Samarabandu, IEEE Trans, 1424403677, 2006.
[9] N. Ezaki, M. Bulacu , and Schomaker , IEEE Computer Society, vol 2 pp. 683-686, 2004.
[10] H. Rajesh , and N. Mohammed, J. Futuristic Science Engineering and Technology, vol 1, pp. 2320 – 4486, 2013.
[11] U. Modha, and P. Dave, J. Engineering Research and Applications (IJERA), Vol 2, pp. 2248-9622, 2012.
[12] S. Muthukumar, .N. Krishnan,P. Pasupathi, and S. Deepa, J. Computer Applications, vol 9, pp. 0975 – 8887, 2010.
[13] Molecular Modeling Pro Plus, Version 6.2.5. www.chemistry-software.com.
[14] P. Gramatica, J. QSAR Comb. Sci, vol 26, pp. 694-701, 2007.
[15] A. Tropsha, J. Mol. Inf.pp. 476-488, 2010.
[16] S. Richards, G. J. Med. Chem, pp. 35-3207, 1992.
[17] B. Efron, J.Am.Stat.Assoc,vol 9, pp.78-316,1983.
[18] A.Wilf, Eds. Wiley. NewYork. 1960.
[19] D. W. Osten, J.Chemom,vol 2, pp.39, 1998.

