ASIAN JOURNAL OF PHARMACEUTICAL AND CLINICAL RESEARCH

INNOVARE ACADEMIC SCIENCES Knowledge to Innovation

Vol 7, Issue 4, 2014

ISSN - 0974-2441

Research Article

QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP STUDY FOR THE PREDICTION OF INHIBITORY CONCENTRATION 50 FOR 5-N-ACETYL-BETA-D-NEURAMINIC ACID STRUCTURALLY SIMILAR COMPOUNDS USING NEURAL NET

PONMARY PUSHPA LATHA D[1], JOSEPH PUSHPA RAJ D[2]

[1]Department of Computer Applications, Karunya University, Coimbatore - 641 114, Tamil Nadu, India. [2]Department of Information Technology, Francis Xavier Engineering College, Tirunelveli - 627 003, Tamil Nadu, India. Email: ponmarymca@gmail.com

Received: 26 June 2014, Revised and Accepted: 24 July 2014

ABSTRACT

Quantitative structure activity relationship (QSAR) study has been developed for structurally similar to 5-N-acetyl-Beta-D-Neuraminic acid as inhibitors for *Clostridium tetani* causing targets using neural network. QSAR models for biological activity of half-maximal inhibitory concentration 50 (IC_{50}) were created with 110 training compounds, 50 test compounds, and 16 different descriptors. The predictive capability of the QSAR models was evaluated by r^2, $q^2_{LMO\ (TestSet)}$, $q^2_{LOO(TestSet)}$, $q^2_{BOOT(TestSet)}$. The comparison of various external validation reveals identical $q^2_{LMO(TestSet)}$, $q^2_{LOO(TestSet)}$ and $q^2_{BOOT(TestSet)}$ for IC_{50} (0.9) which demonstrates the high robustness and real predictive power of IC_{50} model.

Keywords: Quantitative structure activity relationship, Neural network, Inhibitory concentration 50, Leave-many-out, Leave-one-out, BOOT.

INTRODUCTION

Quantitative structure activity relationship (QSAR) describes how a known biological activity can differ as a function of molecular descriptors derived from the chemical structure of a set of molecules. Many physiological activities of a molecule can be associated with their composition and structure. Molecular descriptors, which are numerical depictions of the molecular structures, are used for performing QSAR analysis. 5-N-acetyl-Beta-D-Neuraminic acid represents the most important class of biologically active compounds as inhibitors of *Clostridium tetani* [1,2]. The half maximal inhibitory concentration 50 (IC_{50}) is the concentration of an inhibitor that is necessary for 50% inhibition of an enzyme *in vitro* [3]. In the present study, QSAR studies have been carried out for 5-N-acetyl-Beta-D-Neuraminic acid and its structurally similar compounds with (>95%). We have developed the IC_{50} QSAR [4,5] models for 5-N-acetyl-Beta-D-Neuraminic acid and its structurally similar compounds with (>95%) using neural network by the rapid miner software [6].

MATERIALS AND METHODS

Data set

Training set of 110 compounds and test set of 50 compounds related to 5-N-acetyl-beta-D-neuraminic acid (Fig. 1) which is available in *C. tetani* were collected from pubchem [7]. The Dataset is in the form of smiles notation, which are given as supporting material, the smiles notation are given to QikProp [8] program to calculate the molecular descriptors. The molecular descriptors are converted into tabular form, and it was given as input to rapid miner software to predict the model using neural network. In order to get the efficient model, 69% of the dataset are taken as a training set and the remaining dataset are considered for test set.

Molecular descriptors

Theoretical molecular descriptors are calculated using QikProp [8] program. The following descriptors are procured into consideration for developing the model: (1) Molecular weight (MW), (2) hydrophobic SASA (HAS), (3) hydrophilic SASA (HLSA), (4) molecular volume (MV), (5) vdW polar SA (PSA), (6) number of rotatable bonds (RB), (7) donor - hydrogen bonds (DHB), (8) acceptor - hydrogen bonds (AHB), (9) ionization potential (IP), (10) electron affinity (EA), (11) log P for octanol/water (Log P), (12) log S for aqueous solubility (AS), (13) human oral absorption (HOA), (14) lipinski rule (LR), (15) half-maximal IC_{50}, (16) number of ring (NR).

Neural net

An artificial neural network (ANN) is an information processing paradigm that is enthused, by the way, biological nervous systems such as the brain, process information [9]. The key constituent of this model is the novel structure of the information processing system. It is composed of a large number of greatly interrelated processing elements (neurons) operational in unison to resolve exact problems. This ANN operator learns a model by means of a feed-forward neural network trained by a back propagation algorithm (multilayer perception) (Fig. 2).

A feed forward neural network is a biologically stirred classification algorithm. It contains a large number of simple neuron-like processing units, structured in layers. Each unit in a layer is related with all the units in the previous layer. These connections are not all equal; each connection may have different strength or weight. The weights on these

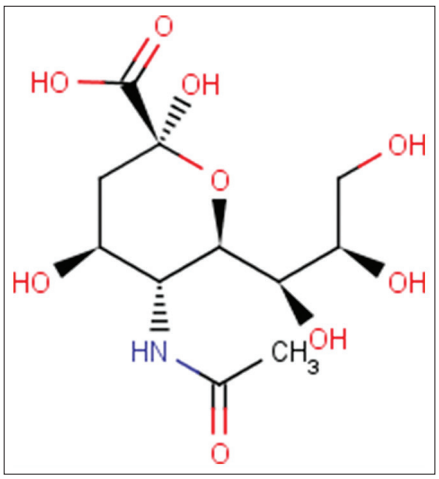

Fig. 1: 5-n-acetyl-beta-d-neuraminic acid

connections encode the knowledge of the network. Often the units in a neural network are also called nodes. Data enter at the inputs and pass all the way through the network, layer by layer, until it arrive at the outputs. During normal operation that is when it acts as a classifier, there is no feedback between layers. This is why they are called feed-forward neural networks. Propagation algorithm propagates inputs forward in the usual way, i.e., 1. All outputs are calculated via sigmoid thresholding of the inner product of an equivalent weight and input vectors. All outputs at stage n are connected to all the inputs at stage n+1.2. It propagates the errors backwards by apportioning them to each unit according to the amount of this error the unit is responsible for. q^2 is calculated using the following formula. y_i is the actual experimental activity,

$$q^2=1-\Sigma(y_i-j_i)^2/\Sigma(y_i-y_i)^2$$

Where, y_i- the average actual experimental activity and y_i- the predicted activity of the compound *i* are computed by the predicted model. The robustness and internal predictivity of the models were evaluated by both leave-one-out (LOO) cross-validation ($q^2_{LOO\ (TestSet)}$) and leave-many-out (LMO) cross-validation ($q^2_{LMO\ (TestSet)}$) [10-21]. The in-house computer programs are created in Java programming to do the following cross-validation techniques: Leave-some-out, LOO, and bootstrapping. In LMO, the data set was split into the sequence of nine set of compounds (45, 40, 35, 30, 25, 20, 15, 10, 5) and the cross-validation was performed. The average of q^2 LMO was calculated as follows: IC_{50} (0.9610). LOO cross-validation is as follows:

1. Assign total compound n=50, compound i=1
2. Leave compound i
3. Calculate q^2_i
4. i=i+1
5. Repeat step 2-5 till i≤50
6. Find the average of $q^2_{i=1..n.}$

q^2 LOO (Test set) for IC_{50} is 0.8669. Bootstrap cross validation is computed as follows:

1. Generate n random number R_i within the range of 1-50 where i=1..n
2. i=1
3. Remove R_i compounds
4. Calculate q^2_i
5. i=i+1
6. Repeat step 3-5 till i<=n
7. Find the average of $q^2_{i=1..n.}$

The average of q^2 BOOT was calculated as follows: IC_{50} (0.9581). Table 1 shows the different cross-validation of IC_{50} (Fig. 3) and Table 2 represents the observed and predicted values which were found to be a small deviation.

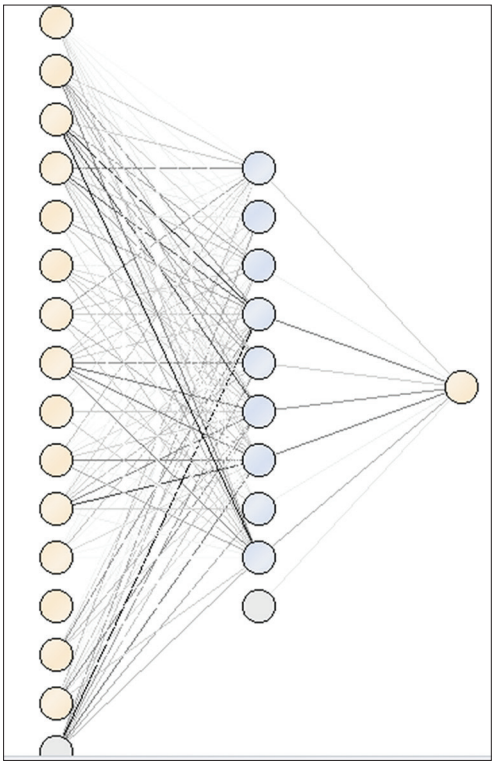

Fig. 2: Improved neural net

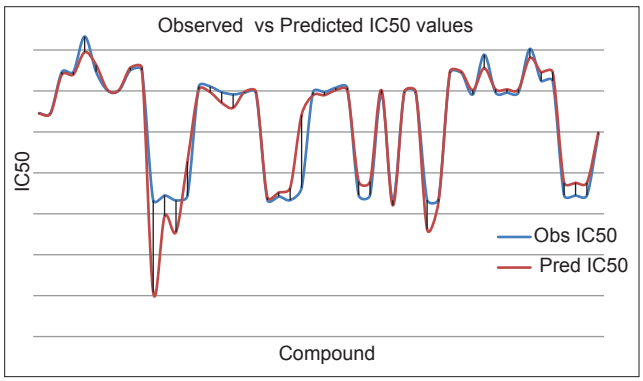

Fig. 3: Observed versus predicted inhibitory concentration 50 for testset

RESULTS AND DISCUSSION

In practice, neurons normally do not produce an output unless their total input goes over a threshold value. The total input for each neuron

Hidden layer	
Node 1 (sigmoid)	
RB	−0.131
MW	−0.707
HAS	0.747
HLSA	1.707
MV	−0.163
DHB	−0.330
AHB	1.354
Log P	−0.721
IP	0.432
EA	0.907
HOA	0.229
RF	−0.217
Ring A	0.035
AS	0.004
PSA	0.933
Threshold	0.241
Node 2 (sigmoid)	
RB	0.238
MW	0.346
HAS	0.134
HLSA	0.304
MV	0.280
DHB	−0.021
AHB	0.271
Log P	−0.534
IP	−0.110
EA	−0.223
HOA	0.713
RF	0.094
Ring A	0.019
AS	−0.192
PSA	0.198
Threshold	−1.287
Node 3 (sigmoid)	
RB	0.275

Contd...

Contd...

MW	−0.688
HAS	1.308
HLSA	0.841
MV	0.347
DHB	−0.252
AHB	0.683
Log P	−0.823
IP	−0.005
EA	0.098
HOA	0.321
RF	−0.231
Ring A	0.006
AS	0.590
PSA	−0.137
Threshold	−0.889
Node 4 (sigmoid)	
RB	0.158
MW	1.603
HAS	−3.288
HLSA	−2.445
MV	−1.225
DHB	0.530
AHB	0.745
Log P	0.674
IP	−0.071
EA	−0.201
HOA	0.686
RF	0.963
Ring A	0.002
AS	1.594
PSA	−0.824
Threshold	−3.616
Node 5 (sigmoid)	
RB	−0.324
MW	−1.046
HAS	1.286
HLSA	1.484
MV	0.220
DHB	−0.781
AHB	0.268
Log P	−1.615
IP	0.029
EA	0.477
HOA	0.976
RF	−0.243
Ring A	−0.048
AS	0.302
PSA	−0.399
Threshold	−1.487
Node 6 (sigmoid)	
RB	0.404
MW	−1.204
HAS	2.705
HLSA	0.469
MV	0.863
DHB	−0.674
AHB	0.055
Log P	−1.950
IP	0.725
EA	0.669
HOA	1.784
RF	−0.070
Ring A	−0.028
AS	1.067
PSA	−1.768
Threshold	−0.890
Node 7 (sigmoid)	
RB	0.342
MW	0.555
HAS	0.641
HLSA	−0.707
MV	0.088
DHB	−0.898

Contd...

Contd...

AHB	1.046
Log P	1.884
IP	0.012
EA	−1.688
HOA	2.708
RF	0.527
Ring A	−0.009
AS	−0.665
PSA	0.583
Threshold	−2.067
Node 8 (sigmoid)	
RB	0.097
MW	0.035
HAS	0.399
HLSA	0.314
MV	0.103
DHB	−0.079
AHB	0.204
Log P	−0.753
IP	−0.076
EA	−0.235
HOA	0.646
RF	0.153
Ring A	0.023
AS	0.148
PSA	0.031
Threshold	−1.240
Node 9 (sigmoid)	
RB	−0.740
MW	−1.821
HAS	3.326
HLSA	1.122
MV	0.454
DHB	−0.829
AHB	0.780
Log P	−1.858
IP	−0.712
EA	−1.120
HOA	0.240
RF	0.114
Ring A	−0.013
AS	−0.059
PSA	−0.731
Threshold	1.082
Output	
Regression (linear)	
Node 1	−0.745
Node 2	0.003
Node 3	−0.282
Node 4	1.915
Node 5	−0.719
Node 6	−1.585
Node 7	1.888
Node 8	−0.251
Node 9	−1.152
Threshold	0.400

Table 1: Validation of IC_{50} model

Model	r^2	$q^2_{LMO\ (TestSet)}$	$q^2_{LOO\ (TestSet)}$	$q^2_{BOOT\ (TestSet)}$
IC_{50}	0.9655	0.9610	0.8669	0.9581

LMO: Leave-many-out, LOO: Leave-one-out, IC_{50}: Inhibitory concentration 50

is the sum of the weighted inputs to the neuron minus its threshold value. This is then passed through a sigmoid function. The following are the sigmoid and threshold values.

Neural net predicted values are more accurate than the multivariate linear regression QSAR study predicted values (Table 2) [22]. The Graph of experimental versus the predicted values for the present

Table 2: Observed versus predicted values

Observed IC_{50}	Predicted IC_{50}
0.4916	0.489698
0.4916	0.489721
0.6928	0.67989
0.6928	0.679953
0.8658	0.789687
0.6935	0.726929
0.6022	0.605415
0.6022	0.60487
0.7011	0.714114
0.7011	0.714117
0.0656	−0.38371
0.0897	−0.01185
0.0656	−0.08941
0.0874	0.255302
0.6225	0.60704
0.6212	0.593262
0.5945	0.542341
0.5833	0.517553
0.5909	0.595678
0.5909	0.595668
0.0680	0.080061
0.0847	0.104308
0.0664	0.126232
0.1265	0.482732
0.5939	0.579481
0.5939	0.579481
0.6156	0.603509
0.6156	0.603501
0.0910	0.157199
0.0910	0.157199
0.5833	0.604201
0.0663	0.041051
0.5909	0.595686
0.5909	0.595678
0.0675	−0.0761
0.0723	0.058521
0.6886	0.694463
0.6886	0.694463
0.5833	0.604126
0.7772	0.711414
0.5921	0.60771
0.5921	0.607712
0.5921	0.607712
0.8057	0.761137
0.6485	0.69177
0.6485	0.691769
0.0899	0.151601
0.0899	0.151601
0.0899	0.151601
0.3882	0.396843

IC_{50}: Inhibitory concentration 50

IC_{50} model is displayed in Fig. 3. The training compound in this study shows the range of IC_{50} between 0.0298 and 0.8439 Table 1 describes the $q^2_{LMO(TestSet)}$, $q^2_{LOO(TestSet)}$ and $q^2_{BOOT(TestSet)}$ values of neural net IC_{50} model. The $q^2_{LMO\ (TestSet)}$, $q^2_{LOO\ (TestSet)}$ and $q^2_{BOOT\ (TestSet)}$ validation values of multivariate linear regression IC_{50} model are also above 0.8. Since the values are greater than 0.8, The QSAR model may be considered.

CONCLUSION

In this study, it was possible to obtain an ANN QSAR [23,24] model of IC_{50} for a set of one hundred and ten compounds which are 95% structurally similar to 5-N-acetyl-beta-D-neuraminic acid as inhibitors for *C. tetani* neurotoxins. The LOO, LMO, and BOOT cross-validation techniques show that the model is significant, robust and has good predictability. The IC_{50} models are showing minimum deviation between observed and predicted values and also having good internal and external predictive power.

ACKNOWLEDGMENT

I thank Dr. D. Jeya Sundara Sharmila for her valuable encouragement in the field of research.

REFERENCES

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, *et al*. The protein data bank. Nucleic Acids Res 2000;28(1):235-42.
2. Foster J, Kane P, Speight N. The Detoxx .system: detoxification of biotoxins in chronic neurotoxic syndromes. Townsend Letter for Doctors and Patients. November 2002.
3. Cheng Y, Prusoff WH. Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. Biochem Pharmacol 1973;22(23):3099-108.
4. Available from: http://www.rapid-i.com/.
5. Pubchem Project. Available from: ttp://www.pubchem.ncbi.nlm.nih.gov/.
6. Schrodinger, Qikprop. New York, NY: LLC; 2008.
7. Yao XJ, Panaye A, Doucet JP, Zhang RS, Chen HF, Liu MC, *et al.* Comparative study of QSAR/QSPR correlations using support vector machines, radial basis function neural networks, and multiple linear regression. J Chem Inf Comput Sci 2004;44(4):1257-66.
8. Castilho C, Guido RV. Andricopulo AD. Classical and hologram: QSAR studies on a series of tacrine derivatives as butyrylcholinesterase inhibitors. Lett Drug Des Discov 2007;4:106-13.
9. Papa E, Dearden JC, Gramatica P. Linear QSAR regression models for the prediction of bioconcentration factors by physicochemical properties and structural theoretical 0molecular descriptors. Chemosphere 2007;67(2):351-8.
10. de Melo EB, Ferreira MM. Multivariate QSAR study of 4,5-dihydroxypyrimidine carboxamides as HIV-1 integrase inhibitors. Eur J Med Chem 2009;44(9):3577-83.
11. Hazuda DJ, Felock P, Witmer M, Wolfe A, Stillmock K, Grobler JA, *et al.* Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. Science 2000;287(5453):646-50.
12. Ghasemi J, Saaidpour S. QSPR prediction of aqueous solubility of drug-like organic compounds. Chem Pharm Bull (Tokyo) 2007;55(4):669-74.
13. Chung JY, Pasha FA, Cho SJ, Won M, Lee JJ, Lee K. Pharmacophore-based 3D-QSAR of HIF-1 inhibitors. Arch Pharm Res 2009;32(3):317-23.
14. Solomon KA, Sundararajan S, Abirami V. QSAR studies on N-aryl derivative activity towards Alzheimer's disease. Molecules 2009;14:1448-55.
15. Hansch C, Fujita T. A method for the correlation of biological activity and chemical structure. J Am Chem Soc 1964;86(8):1616-26.
16. Liu A, Guang H, Zhu L, Du G, Lee SM, Wang Y 3D-QSAR analysis of a new type of acetylcholinesterase inhibitors. Sci China C Life Sci 2007;50(6):726-30.
17. Pasha FA, Nam KD, Cho SJ. CoMFA based quantitative structure toxicity relationship of azo dyes. Mol Cell Toxicol 2007;3:145-9.
18. Pasha FA, Srivastava HK, Srivastava A, Singh PP. QSTR study of small organic molecules against tetrahymena pyriformis. QSAR Comb Sci 2007;26(1):69-84.
19. Pasha FA, Neaz MM, Cho SJ, Kang SB. Quantitative structure activity relationship (QSAR) study of estrogen derivatives based on descriptors of energy and softness. Chem Biol Drug Des 2007;70(6):520-9.
20. Pasha FA, Chung HW, Cho SJ, Kang SB. 3D-quantitative structure activity analysis and quantum chemical analysis of pyrido-di-indoles. Int J Quant Chem 2008;108(2):391-400.
21. Recanatini M, Cavalli A, Hansch C. A comparative QSAR analysis of acetylcholinesterase inhibitors currently studied for the treatment of Alzheimer's disease. Chem Biol Interact 1997;105(3):199-228.
22. Latha PP, Sharmila JS. QSAR study for the prediction of IC50 and Log P for 5-N-Acetyl-Beta-DNeuraminic acid structurally similar compounds using stepwise (multivariate) linear regression. Int J Chem Res 2010;2(1):32-8.
23. Devanathan R. Pharmaceutical and analytical studies on swarna makshiba bhasma - An ayurvedic formulation. Asian J Pharm Clin Res 2013;6(1):26-9.
24. Rathor S, Ram A. Floating drug delivery system as an approach to increase the gastric retention of methotrexate formulation and evaluation. Asian J Pharm Clin Res 2013;6(1):42-7.