Predictive QSAR Modeling of Cyclic Urea HIV-1 Protease Inhibitors Based on Linear and Non-Linear Regression Methods

S. Pirhadi and J. B. Ghasemi

Chemistry Department, Faculty of Sciences, K.N.Toosi University of Thechnology, Tehran, Iran

Abstract

A Quantitative Structure – Activity Relationship (QSAR) analysis of cyclic urea-based Human Immunodeficiency Virus Type 1 was carried out on a set of cyclic urea-based (HIV-1) protease inhibitors. The stepwise multiple regression analysis and genetic algorithm were employed as the feature selection and model development methods. Modeling of the relationship between selected molecular descriptors and pKi data was achieved by MLR as linear and A Levenberg–Marquardt algorithm trained feed-forward artificial neural network as nonlinear methods. A comparison between the obtained results revealed the non- linear method is better than linear method. The improvement is due to the fact that the activity of the compounds demonstrates non-linear correlations with the selected descriptors.

Introduction

Acquired immuno-deficiency syndrome (AIDS) is a leading cause of death worldwide. Intensive efforts have been madding worldwide to develop chemotherapeutic agents effective against the HIV infection. The viral-encoded protease for human immunodeficiency virus (HIV) is responsible for the processing of viral polyprotein precursors to their mature polypeptides. Since correct processing of the viral polypeptides is essential for the production of infectious virus, HIV protease represents a potential target for therapeutic agents which may prove beneficial in the treatment of AIDS [1].

One of the most potent HIV protease inhibitors has turned out to be the cyclic urea and its derivatives. Its advantages include being smaller than more common peptide – based inhibitores , having therefore better water solubility and bioavailability.

Several Linear and non-linear methods reported in literatures to model HIV-1 protease inhibitors. In this work a dataset of 55 symmetric and non-symmetric cyclic urea HIV-1 protease inhibitors was investigated [2]. MLR and ANN were used for inspection of linear and nonlinear relation between interested activity and molecular descriptors. The stepwise multiple regression analysis and genetic algorithm were employed as the feature selection methods.

Materials & Methods

In order to calculate the theoretical descriptors the 2D structures of the molecules were drawn using ChemBioOffice 11.0 and were firstly preminimized with Molecular Mechanics Force Field (MM^+) procedure and semi empirical AM1 using the Polak-Ribiere algorithm by hyperchem version 8.0.5. The gradient norm criterion 0.1 Kcal/ Å was applied in the geometrization for all structures, and then they transferred into the Dragon version 3.0 in order to obtain many different descriptors these descriptors used as input variables for variable selection by multiple linear regression analysis using the stepwise strategy in SPSS (version 16.0) software. After that the selection of the optimum number of descriptors was performed by running genetic algorithm and the highest occurrence frequency descriptors were selected (6 descriptors). Then the 20% of the total dataset to be used for the test. The multiple linear regression and artificial neural network trained with Levenberg–

Marquardt algorithm as implemented in MATLAB (version 7.6.0.324, Math Works, Inc.) were used to modeling of the relationship between selected variables and pKi (as biological activity).

Results

Once appropriate molecular descriptors are chosen, QSAR models can be constructed to predict the target property of untested chemicals. The selected descriptors are as following:a) MATS6e and b) MATS1v (among 2D autocorrelation descriptors), c) BELe5 (among BCUT descriptors), d) PCWTe, e) Mor04m (among charge and 3D-Morse descriptors, respectively), and f) RDF155v (RDF descriptors). MLR model was checked with the t-Test that showed all variables are important, F- test to test the hypothesis that there is no relationship between independent and dependent variables that is to be rejected because F exceeds the critical level, by calculation of correlation matrix indicated each descriptor encodes different aspects of molecular structure, and variance inflation factor values (VIF) showed there is no information of descriptors hidden by correlation of descriptors [3, 4].

Variables	Standardized coefficients	VIF
a	-0.478	1.214
b	0.447	1.582
c	0.272	1.621
d	0.309	1.036
e	0.232	1.049
f	-0.165	1.166

The absolute value of a standardize coefficient shows the greater the weight of the variable in the model hence order of importance of descriptors is: a>b>d>c>e>f.
In comparison with the MLR based on r^2_{ncv} and r^2_{pred}, ANN approach (with 6 nodes in hidden layer) would seem to have a greater potential for determining quantitative structure-anti-HIV-1 activity relationships.

Conclusion

A MLR and Levenberg–Marquardt algorithm trained neural network approach were used to analyze the QSAR of cyclic urea HIV-1 protease inhibitors. Based on statistical parameters of the MLR model all selected descriptors by stepwise strategy and genetic algorithm are significant.

Comparison of r^2 values in both set of training and test sets shows, the improvements of them in ANN model is due to the fact that the activity of the compounds demonstrates non-linear correlations with the selected descriptors.

References

1) A. Wlodawer, J.W. Erickson, Structure-based inhibitors of HIV-1 protease, *Annu. Rev. Biochem.*, 62, (1993) 543-585
2) P.Y.S. Lam, Y. Ru, P.K. Jadhav, P.E. Aldrich, G.V. DeLucca, C.J. Eyermann, C.-H. Chang, G. Emmett, E.R. Holler, W.F. Daneker, L. Li, P.N. Confalone, R.J. McHugh, Q. Han, R. Li, J.A. Markwalder, S.P. Seitz, T.R. Sharpe, L.T. Bacheler, M.M. Rayner, R.M. Klabe, L. Shum, D.L. Winslow, D.M. Kornhauser, D.A. Jackson, S. Erickson-Viitanen, C.N. Hodge, Cyclic HIV protease inhibitors: synthesis, conformational analysis, P2/P2′ structure-activity relationship, and molecular recognition of cyclic ureas, *J. Med. Chem.*, 39(18), (1996) 3514-3525
3) A. Afantitis, G. Melagraki, H. Sarimveis, P.A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou, A novel QSAR model for predicting induction of apoptosis by 4-aryl-4H-chromens, *Bioorg. Med. Chem.*, 14(19), (2006) 6686-6694
4) G. Choudhary, C. Karthikeyan, N.S. Moorthy Hari Narayana, S.K. Sharma, P. Trivedi, QSAR analysis of some cytotoxic thiadiazinoacridines, *Int. Elec. J. Mol. Des.*, 4(11), (2005) 793-802