Research in Biotechnology, 2: 01-06, 2011 **ISSN: 2229-791X**
www.researchinbiotechnology.com

Regular Article

Classification of Anti-HIV Compounds using QSAR Studies

Vignesh and Arnold Emerson I*
School of Bio Sciences & Technology, VIT University,
Vellore 632 014, Tamil Nadu, India

The Acquired Immunodeficiency Syndrome (AIDS) is one the most fatal disorders for which there have been no complete or successful chemotherapy yet. The pandemic spread of this disease has prompted an unprecedented scientific clinical effort to understand and propagate ways to combat it. Quantitative Structure–Activity Relationship (QSAR) study has been reportedly used to study topological indices, physiochemical and indicator parameters on a series of HEPT analogues for understanding HIV reverse transcriptase inhibitor activity (Shovanlal Gayen et al, 2004). The aim of their study was to correlate the identified molecular surface properties with the inhibitory activity of HIV-1 RT using QSAR. The docking energies were used to categorize anti-HIV compounds as classified and non classified multiple linear regression models. Results indicate that classified MLR had a better predictability than the non classified models. Thus this method of classification proves to be a good predictability model.

Keywords : AIDS, Compounds, Docking, Multiple linear regression, QSAR

The Human Immunodeficiency Virus subtype 1 (HIV-1) is a retrovirus of the lent virus family and is also the causal agent of this disease. Various partial drugs have been created that would treat the HIV infection at various stages but there has been no drug found yet to cure. For this we need to comprehend the chemicals and mathematical models that could be applied as an extrapolation model to study the desired features of an anti HIV drug. The best mathematical model that can quantitatively relate the anti HIV activity with the structural descriptors is the QSAR (Quantitative Structure Activity Relationship). The QSAR analysis has been done for various groups of compounds and also for diverse sets of anti HIV compounds (Rajni Garg et al., 1999).

Principal Component Factor Analysis (FA) was used as the data-preprocessing step to identify predictor variables contributing to the response variable. This is to avoid collinearities among them. The generated Multiple Linear Regression (MLR) equations were statistically validated using leave-one-out technique. The Genetic Function Approximation (GFA) was also used on the same data set to generate QSAR equations, which then would produce the best equations as obtained with FA-MLR.

Computer-aided predictions of the new anti-HIV compounds, derived from substructures of 2-amino-6-arylsulfonyl-benzonitriles and cyclotriazadisulfonamide analogues has been reported by Julia R.

***Corresponding author : i_arnoldemerson@yahoo.com**

Pinheiro et al. (2008). A ligand-based approach (MIA-QSAR) along with docking evaluation were used to model the title compounds, macrocycles containing a trisubstituted benzene moiety. According to the MIA-QSAR method, predicted potencies for proposed compounds were seven times higher than that of the experimentally obtained active compounds of the training set. Moreover using the docking approaches the binding orientations were studied. Thus the binding affinities of these compounds in CD4 receptors could be predicted.

A series of 2-(2,6-dihalophenyl)-3-(substituted pyrimidinyl)-1,3-thiazolidin-4-ones was designed based on the prediction of QSAR studies. This was evaluated as HIV-1 reverse transcriptase inhibitors as presented by Ravindra K. Rawal et al. (2007), The aim of the study was to correlate the identified molecular surface features for modeling the HIV-1 RT inhibitory activity. This resulted in some statistically significant QSAR models with good predictive ability. The derived models collectively suggest that the compounds should be compact without bulky substitution on its peripheries for better HIV-1 RT inhibitory activity. These models also indicate a preference for hydrophobic compounds due to their inhibitory activity.

Materials and Methods

Datasets

Various anti-HIV compounds containing the physiochemical property values namely IC50 along with the structures constitute the dataset. This is the concentration of agent that reduces viral growth by 50% over a 3 day assay period. All the QSAR datasets were obtained from

i) NCI database.
ii) Enhanced NCI database.
iii) Chemi-informatics anti-HIV QSAR datasets. (Along with the DRAGON molecular descriptors).

Structures were manually drawn and some were obtained as individual *.mol or in marvin sketch format. Standard data files (*.sdf) containing all the structures and the IC50 or descriptor values were also used.

Softwares

i) **MDL-ISIS DRAW 2.5** was used to draw the molecules in *.skc (Sk) format.
ii) **CHEM-SKETCH 12.0** was used in the conversion of individual *.skc to *.mol file.
iii) **ARGUS LAB 4.0.1** was used in the 3-D optimization the individual *.mol files and performing molecular docking and calculation of docking energy values.
iv) **OPEN- BABEL** was used in the conversion of all the *.mol files into a cumulative *.sdf file for easy chemical file handling and manipulation.
(http://www.vcclab.org/lab/babel/start.html)
v) **POWER-MV 0.61** was used in handling *.sdf file, subset selection, table formation and various other standard file handling functions as seen in POWER- MV. This also provides Cluster analysis with K means classification and Random Forest Classification.
vi)**Tsar** is a fully integrated Quantitative Structure-Activity Relationship (QSAR) package for library design and lead optimization that helps attain information to guide decisions throughout the drug discovery process ranging from compound selection to reagent selection and library creation. It gives various model development methodologies.

Results

Docking energy based classification

1. Classification into four different classes based on the docking energy values:

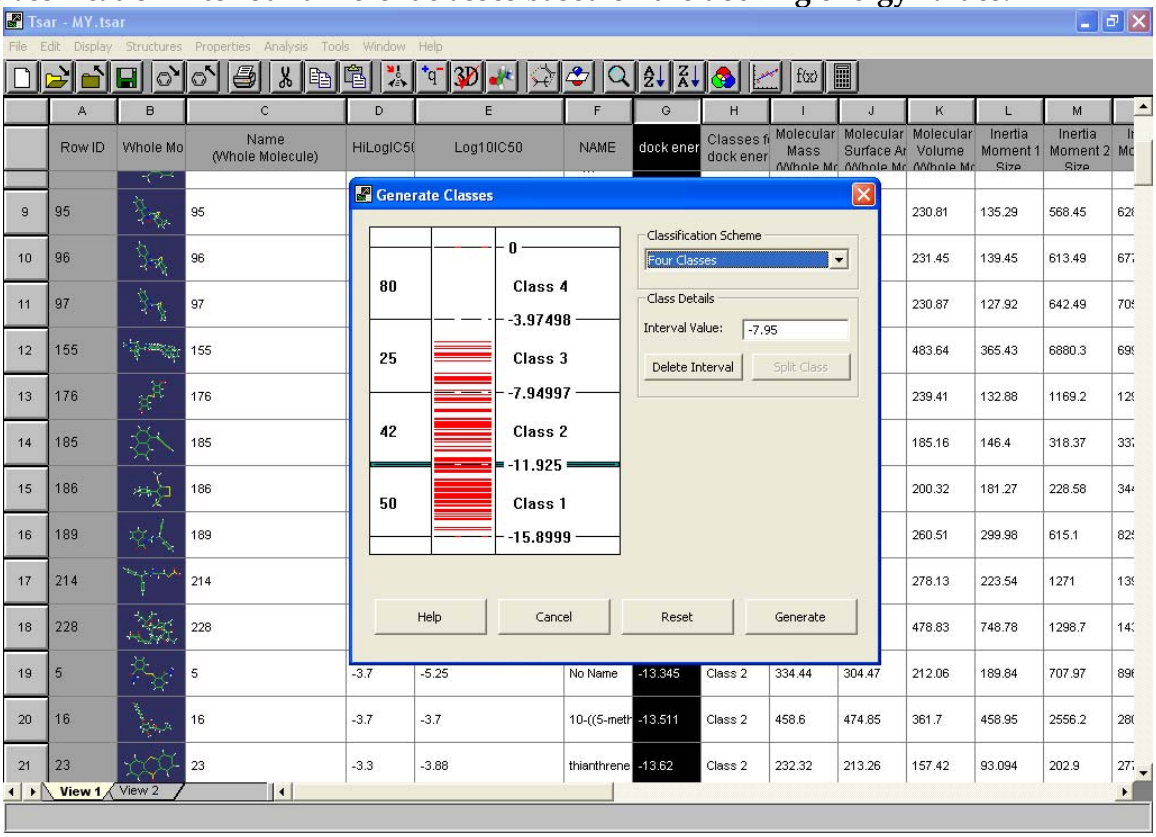

Figure 1: Classification Graph

2 Normal MLR

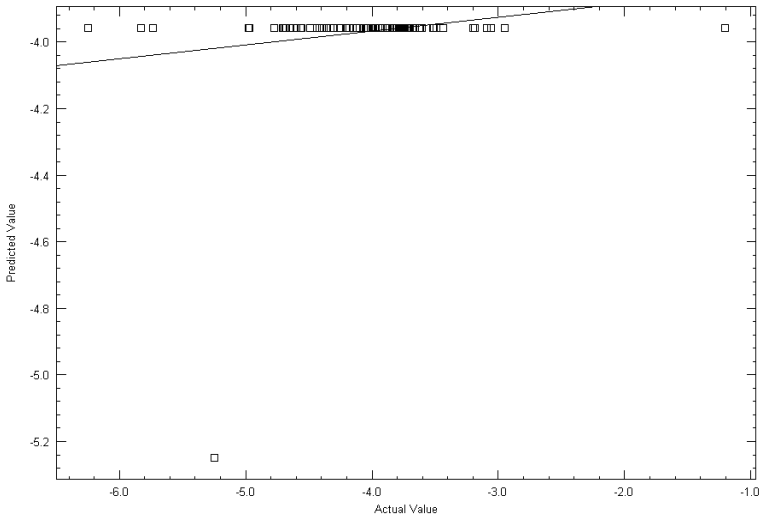

Figure 2: Overall MLR prediction graph

Equations

Original Data: Y = -0.64668876*X82 - 3.9566226

Standardized Data: Y = -0.10490677*S82 - 3.9651315

Post Classification

3. Class 1 MLR

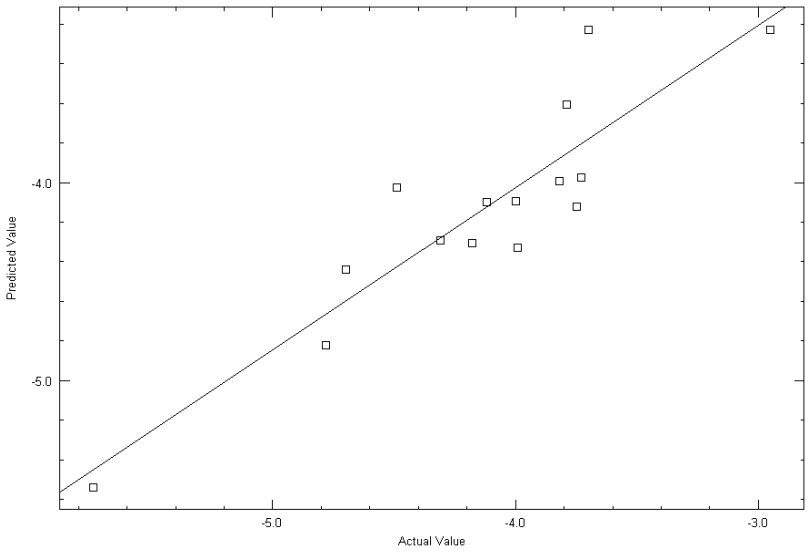

Figure 3 Class 1 MLR

Equations

Original Data : Y = -5.6283526*X21 + 0.75583827*X31 + 0.3569158*X64 - 0.30115417*X69 - 4.8714614

Standardized Data : Y = -0.74554592*S21 + 0.53089207*S31 + 0.46340185*S64 - 0.64187938*S69 - 4.1366668

4. Class 2 MLR

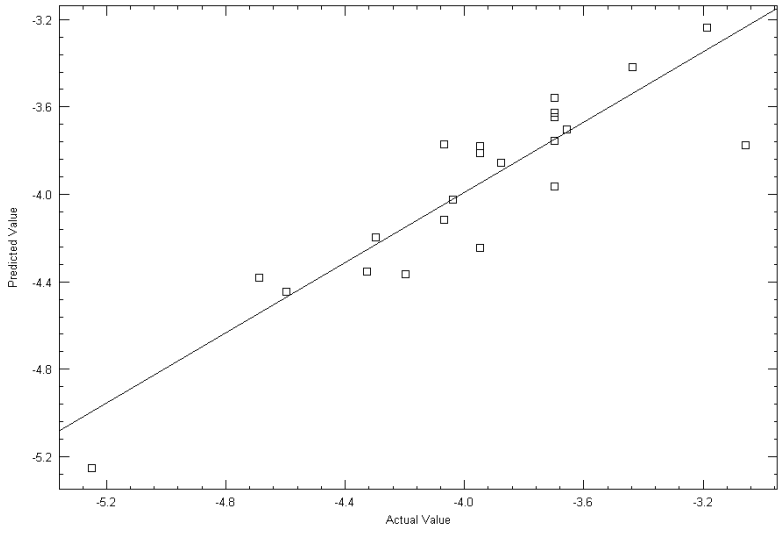

Figure 4. Class 2 MLR

Equations

Original Data : Y = -0.0059862416*X3 + 0.36194509*X26 + 0.1929584*X72 - 0.55514568*X82 - 4.4676013

Standardized Data : Y = -0.2733382*S3 + 0.41150901*S26 + 0.18949525*S72 - 0.23151176*S82 - 3.9491305

5. Class 3 MLR

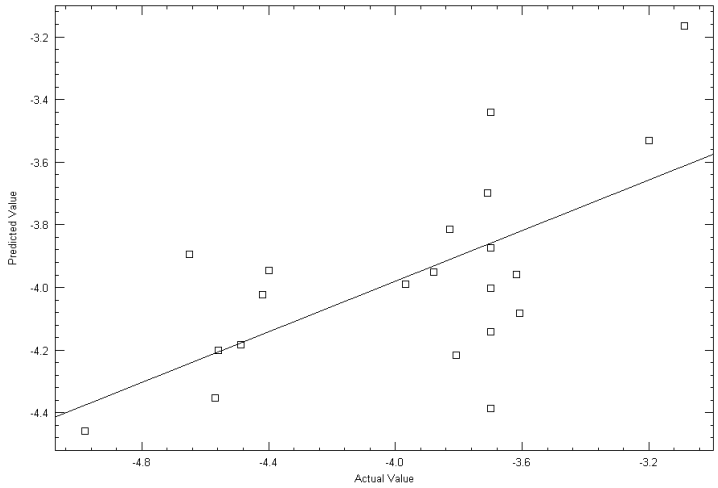

Figure 5. Class 3 MLR

Equations

Original Data : Y = -0.9000001*X54 - 0.73666692*X66 - 0.066666603*X79 - 1.1533329

Standardized Data : Y = -1.026158*S54 - 0.32944745*S66 - 0.11925685*S79 - 4.3439999

6. Class 4 MLR

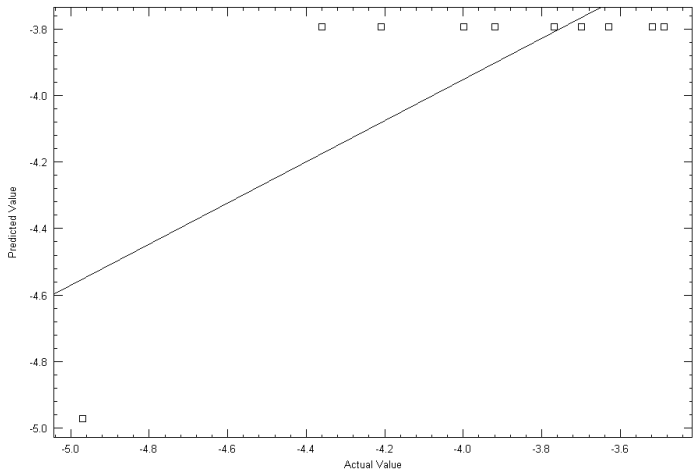

Figure 6. Class 4 MLR

Equations Original Data : Y = -0.58857131*X85 - 3.7928572

Standardized Data : Y = -0.3039369*S85 - 3.8713334

Table 1. Interpretation of classification strategy

Actual date set (Average deviation from original dataset)		Validation data set (Average deviation from original dataset)	
Normal MLR	MLR after classification based on docking energy	Normal MLR	MLR after classification based on docking energy
0.32017	**0.2885**	**0.4273**	**0.3920**

From the above tabulation it is evident that the predictability of the multiple linear regression models has increased. This can be explained as there is a decrease in the average deviation from the original dataset.

Discussion

Medicinal chemistry has widened its horizon from relying on synthetic and theoretical aspects to designing and using computational methods in chemistry. One of the important aspects in computational chemistry is in using the Quantitative Structure Activity Relationship approach which was introduced in the nineteenth century. Since then, it has rapidly become one of the most widely used tools for drug design. Over the years, this approach has applied to a wide variety of research arenas. The main success of the QSAR is in its predictive ability of new compounds. All evidence suggests that QSAR continuously makes a valuable contribution to computer assisted drug designing.

Four classified and non classified multiple linear regression models were obtained using docking energies. The 4 different models of the classified MLR had a better predictability than the non classified MLR models. To test this significance a validation set was chosen separately for which the MLR equations after classification was analyzed. This was found to be increased which is evident from the decrease in the overall deviation from the actual value. This method of classification acts a good predictability model. The descriptors with positive co-efficients bring about an increase in the IC50 values and descriptors with negative co-efficient bring about a decrease in the IC50 values. Thus the present model acts a good predictability model and an interpretational model.

Acknowledgments

The author acknowledges the computational support provided by Bioinformatics Division of VIT University.

References

Julia R. Pinheiro, Michelle Bitencourt, Elaine F. F. da Cunha, Teodorico C. Ramalho and Matheus P. Freitas. 2008. Novel anti-HIV cyclotriazadisulfonamide derivatives as modelled by ligand- and receptor-based approaches", Bioorg. Med. chem. doi:10.1016/j.bmc.2007.11.020.

Kunal Roy and J. Thomas Leonard. 2005. QSAR by LFER model of cytotoxicity data of anti-HIV 5-phenyl-1-phenylamino-1*H*-imidazole derivatives using principal component factor analysis and genetic function approximation" , Bioorganic & Medicinal Chemistry 13, 2967–2973.

Rajni Garg, Satya P. Gupta, Hua Gao, Mekapati Suresh Babu, Asim Kumar Debnath, and Corwin Hansch, 1999. Comparative Quantitative Structure-Activity Relationship Studies on Anti-HIV Drugs, Chem. Rev. 99, 3525-3601.

Ravindra K. Rawal, Rajkamal Tripathi, S. B. Katti, Christophe Pannecouquec and Erik De Clercq. 2007. Synthesis and evaluation of 2-(2,6-dihalophenyl)-3-pyrimidinyl-1,3-thiazolidin-4-one analogues as anti-HIV-1 agents", Bioorganic & Medicinal Chemistry 15 3134–3142.

Shovanlal Gayen, Bikash Debnath, Soma Samanta and Tarun Jha. 2004. QSAR study on some anti-HIV HEPT analogues using physicochemical and topological parameters, Bioorganic & Medicinal Chemistry 12, 1493–1503.