

STC 2006 EFMI Special Topic Conference
Reichert/Mihalaş/Stoicu-Tivadar/Schulz/Engelbrecht (Eds.)
AKA, Berlin (2006)

Molecular Descriptors Family on Structure-Activity Relationships on anti-HIV-1 potencies of HEPT and TIBO derivatives

Sorana BOLBOACĂ¹, Stefan ȚIGAN¹, and Lorentz JÄNTSCHI²

¹*Iuliu Hatieganu University of Medicine and Pharmacy, 13 Emil Isac, 400023 Cluj-Napoca, Romania, sbolboaca@umfcluj.ro*

²*Technical University of Cluj-Napoca, 15 Constantin Daicoviciu, 400020 Cluj - Napoca, Romania*

Abstract

A new developed methodology of Structure-Activity Relationships (SAR) was applied on a set of 57 compounds with known inhibition activity of immunodeficiency virus type 1. The methodology uses an original family of molecular structure descriptors called Molecular Descriptors Family. With a set of multiple linear regression analysis programs, the whole set of MDF members were crossed in order to find the best SAR model. The obtained model allows making important remarks on structure-activity links. The disadvantage of time consuming to analyze the entire set of descriptors is compensated by better structure-activity relationships.

Keywords: HIV-1 inhibitors, Structure Activity Relationships (SAR), Molecular Descriptors Family (MDF)

Introduction

Two different types of human immunodeficiency viruses (HIV-1 and HIV-2) differing in nucleotide and amino-acid sequences are responsible by the acquired immunodeficiency syndrome, but the HIV-1 type is most predominant [1].

A previous study analyzed the HEPT and TIBO derivatives potencies on HIV-1 [2] using quantitative structure-activity relationships methodology. The results obtained by Toporov & all are:

$n = 57$; $r = 0.9397$; $s = 0.520$; $F = 416$ (all compounds)

$n = 37$; $r = 0.9426$; $s = 0.513$; $F = 279$ (training set)

(1)

$n = 20$; $r = 0.9408$; $s = 0.547$; $F = 139$ (test set)

where n is size of the sample; r is the correlation coefficient; s is standard error and F is Fisher parameter.

Starting with the integration of complex structure information of HEPTA and TIBO derivatives, the aim of the research was to evaluate the ability of molecular

descriptor family structure-activity relationships in modeling of the inhibition effectiveness against HIV-1.

Material and Method

A number of nineteen HEPT derivatives and thirty-eight TIBO derivatives with inhibition properties on HIV-1 were included into the study. The effectiveness in inhibiting HIV-1 of HEPT and TIBO derivatives (two groups of reverse transcriptase inhibitors) was taken from a previous paper [3] and is expressed as the concentration of compound required to achieve 50% protection of MT-4 cells against the virus (called $\log(10^6/C_{50})$).

The use of a new original set of molecular descriptors, called Molecular Descriptors Family (MDF) into a Quantitative Structure-Activity Relationship was applied in order to study the inhibiting HIV-1 activity of 19 HEPT and 38 TIBO compounds. The steps of molecular descriptor family on structure activity relationship (MDF SAR) modeling are [4]:

- Step I: Sketch of HEPT and TIBO compounds by the use of HyperChem software [5];
- Step II: Create the file with measured inhibiting HIV-1 activity (Y_{C50}) of HEPT and TIBO compounds;
- Step III: Generate the MDF members based on topological and geometrical representations of the compounds. There were identified 296965 MDF members with real and not identical values from which only 95277 were distinct each from other. More, considering also the withdrawing of planar dependencies (one descriptor is dependent on other two) it remains only 84408.
- Step IV: Finding the SAR models for HEPT and TIBO compounds. The selected members enter into multiple linear regression analysis. Mono-varied and multi-varied models were applied. At the end of all pair's computations the best QSAR models were selected and presented here. Note that for bi-varied model, 3562313028 pairs enter into bi-varied regression model and a multiple of enters into tri- and more varied models.
- Step V: Validation of the obtained SAR models were performed through computing the cross-validation leave-one-out correlation score [6], and the difference between this parameter and the squared correlation coefficient.
- Step VI: Analyze the selected SAR model and comparing it with previous reported model.

Results

The best performing SAR (five-varied model) was selected and is presented here. The selection of the best performing five-varied model was made first after the greatest squared correlation coefficient and then after the greatest values of cross-validation leave-one-out (loo) score ($r_{cv(loo)}^2$).

The models have the following equation:

$$\hat{Y} = 17.7 - 7.11 \cdot \ln MdTHg - 1.23 \cdot IFDMwEt + 8.36 \cdot AiMrKQt + 6.59 \cdot 10^5 \cdot ImDMtQt - 5.98 \cdot IIMdEMg \quad (2)$$

where \hat{Y} is predictor of measured inhibition activity (Y_{C50}) and $\ln MdTHg$, $IFDMwEt$, $AiMrKQt$, $ImDMtQt$, and $IIMdEMg$ are molecular descriptors.

The characteristics associated with the above-described models are in table 1 and is graphically represented in figure 1.

Table 1. Statistics associated with the five-varied model

Characteristic	Notation	SAR Model (eq. 2)
Correlation coefficient	r	0.9579
Squared correlation coefficient	r^2	0.9175
Adjusted squared correlation coefficient	r^2_{adj}	0.9094
Standard error of estimated	S_{est}	0.4521
Fisher parameter	F_{est}	113
Probability of wrong model	$P_{est}(\%)$	$2.14 \cdot 10^{-24}$
t parameter for intercept	t_{int}	22.33
p-values	$P_{tint}(\%)$	$5.52 \cdot 10^{-26}$
95% CI (confidence interval) [lower 95%; upper 95%]	$95\% CI_{int}$	[16.13, 19.32]
t parameter for InMdTHg descriptor	$t_{InMdTHg}$	-9.27
Associated p-value	$P_{InMdTHg}(\%)$	$1.63 \cdot 10^{-10}$
95% CI [lower 95%; upper 95%]	$95\% CI_{InMdTHg}$	[-8.65, -5.57]
t parameter for IFDMwEt descriptor	$t_{IFDMwEt}$	-12.43
Associated p-value	$P_{IFDMwEt}(\%)$	$4.75 \cdot 10^{-15}$
95% CI [lower 95%; upper 95%]	$95\% CI_{IFDMwEt}$	[-1.43, -1.04]
t parameter for AiMrKQt descriptor	$t_{AiMrKQt}$	9.58
Associated p-value	$P_{AiMrKQt}(\%)$	$5.43 \cdot 10^{-11}$
95% CI [lower 95%; upper 95%]	$95\% CI_{AiMrKQt}$	[6.61, 10.11]
t parameter for ImDMtQt descriptor	$t_{ImDMtQt}$	6.86
Associated p-value	$P_{ImDMtQt}(\%)$	$9.22 \cdot 10^{-7}$
95% CI [lower 95%; upper 95%]	$95\% CI_{ImDMtQt}$	$[4.66 \cdot 10^5, 8.52 \cdot 10^5]$
t parameter for IIMdEMg descriptor	$t_{IIMdEMg}$	-7.07
Associated p-value	$P_{IIMdEMg}(\%)$	$4.15 \cdot 10^{-7}$
95% CI [lower 95%; upper 95%]	$95\% CI_{IIMdEMg}$	[-7.68, -4.29]
Cross-validation leave-one-out (loo) score	$r^2_{cv(loo)}$	0.8997
Fisher parameter for loo analysis	F_{pred}	91
Probability of wrong model for loo analysis	$P_{pred}(\%)$	$< 10^{-17}$
Standard error for loo analysis	S_{loo}	0.4987
The difference between r^2 and $r^2_{cv(loo)}$	$r^2 - r^2_{cv(loo)}$	0.0178

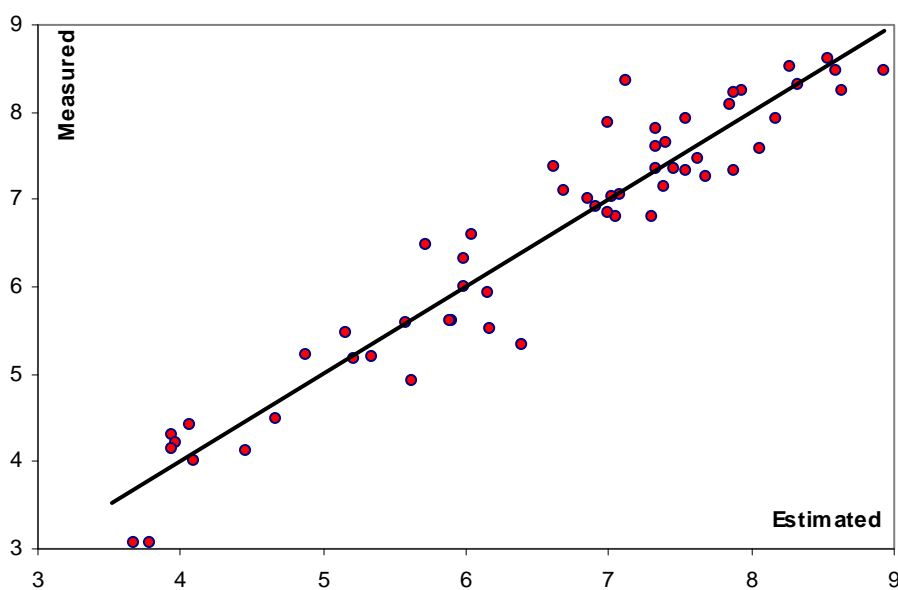


Figure 1. Measured inhibition activity (Measured) vs. estimated (Estimated) with five-varied SAR model

Assessment of the MDF SAR model was performed by applying a correlated correlation analysis, which took into consideration the five-varied SAR models and compared it with the best performing (model with four variables, $r = 0.9397$, $n = 57$ – equation 1) previous reported model [2] by the use of Steiger's Z test. The results of comparison are in table 2.

Table 2. The Steiger's Z test results

Characteristic	Value
$r(Y_{C50}, \hat{Y}_{SAR})$	0.9579
$r(Y_{C50}, \hat{Y}_{Previous})$	0.9399
$r(\hat{Y}_{SAR}, \hat{Y}_{Previous})$	0.9252
Steiger's Z test parameter	1.3462
$p_{Steiger's Z} (\%)$	0.0891

Discussions

The selected best found five-varied SAR model of HEPT and TIBO QSAR HIV-1 inhibiting activity shows that atoms mass and attached hydrogen's has significance on activity behavior using geometrical model of the molecule (InMdTHg and IIMdEMg). Partial charge has significance on activity behavior using strictly topological model. More, two descriptors use the **Qt** association in the selected best found five-varied model (AiMrKQt and ImDMtQt).

The atomic and group electronegativity, as a composed property tends with increasing of number of descriptors to be replaced by more accurate properties: attached hydrogen's, partial charge and mass. Thus, if bi-varied model has only atomic and group electronegativity as atomic descriptors, in tri-varied model disappear one electronegativity based descriptor and appear one attached hydrogen's based and one partial charge based, and for five-varied are two descriptors based on partial charge, one based on attached hydrogen's and one based on atomic mass.

Looking at the five-varied model, we can say that the inhibitory activity it is of molecular topology as well as molecular geometry and depend on partial change of molecule, molecular mass and number of bounded hydrogen's. The values of squared correlation coefficient ($r^2 = 0.9175$), the student parameter, associated p-values and 95% confidence intervals (see table 1) demonstrate the goodness of fit of the five-varied MDF SAR model. The power of the five-varied model in prediction of the inhibitory activity of compounds is demonstrate by the cross-validation leave-one-out correlation score ($r^2_{cv(loo)} = 0.8997$). The stability of the best performing five-varied MDF SAR model is give by the difference between the squared correlation coefficient and the cross-validation leave-one-out correlation score ($r^2 - r^2_{cv(loo)} = 0.0178$).

Comparing with previous reported model (equations (1)) [2], our model (equations (2)) is better ($r^2 = 0.918$ – see table 1 and equation 1). At modeling level, our approach is more software independent than previous reported. We use software dependent procedures only for constructing a basic geometrical model of the molecules and compute the partial charge distribution inside the molecules. We do not "optimize" the geometrical shape according to an arbitrary choused model and/or algorithm.

Starting with the knowledge learned from the studied set, inhibition property of new compound from the same class can be predict by the use of an original software, which is available at the following address:

http://vl.academicdirect.org/molecular_topology/mdf_findings/sar/

Thus, the software is able to predict the inhibitory activity of new compounds from the same class with low costs.

It can be concluded that the use of MDF for SARs finding on HIV-1 potent compounds offers accurate models and allow making of important remarks about structure-activity links.

Acknowledgement

The research was supported in part through project ET36/2005 by UEFISCSU Romania.

References

- [1] Gallo RC, Reitz Jr. MS. The first human retroviruses: are there others? *Microbiological Sciences* 1985; 2(4): 97-104.
- [2] Toropov AA, Toropova AP, Nesterova IV, Nabieva O.M. Comparison of QSAR models of anti-HIV-1 potencies based on labeled hydrogen filled graph and graph of atomic orbitals. *Journal of Molecular Structure: THEOCHEM*. 2003;640(1-3):175-81.
- [3] Ganzalez OG, Murray JS, Peralta-Inga Z, Politzer P. Computed molecular surface electrostatic potentials of two groups of reverse transcriptase inhibitors: Relationships to anti-HIV-1 activities. *Int. J. Quantum Chem*. 2001;83:115–21
- [4] Jäntschi L. Molecular Descriptors Family on Structure Activity Relationships 1. The review of Methodology, *Leonardo Electronic Journal of Practices and Technologies*. AcademicDirect. 2005;6:76-98.
- [5] ***, HyperChem, Molecular Modelling System [Internet page]; ©2003, Hypercube, Inc., available at: <http://hyper.com/products/>
- [6] ***, Leave-one-out Analysis. ©2005, Virtual Library of Free Software, available at: http://vl.academicdirect.org/molecular_topology/mdf_findings/loo/