3rd International Electronic Conference on Medicinal Chemistry
1-30 November 2017
chaired by Dr. Jean Jacques Vanden Eynde
sponsored by pharmaceuticals

SAR studies for the *in silico* prediction of HIV-1 inhibitors

Ismail Hdoufane [1] *, Imane Bjij [1,2], Mahmoud E. S. Soliman [2], Alia Tadjer [3], Didier Villemin [4], Jane Bogdanov [5] and Driss Cherqaoui [1]

[1] Department of Chemistry, Faculty of Sciences Semlalia BP 2390 Marrakech, Morocco.

[2] School of Health Sciences, University of KwaZulu-Natal, Westville, Durban 4000, South Africa.

[3] Sofia University "ST.KLIMENT OHRIDSKI" Faculty of Chemistry and Pharmacy, 1 James Bourchier Avenue 1164 Sofia, Bulgaria.

[4] Ecole Nationale Supérieure d'Ingénieurs (E.N.S.I.) I. S. M. R. A., LCMT, UMR CNRS n° 6507, 6 boulevard Maréchal Juin, 14050 Caen France.

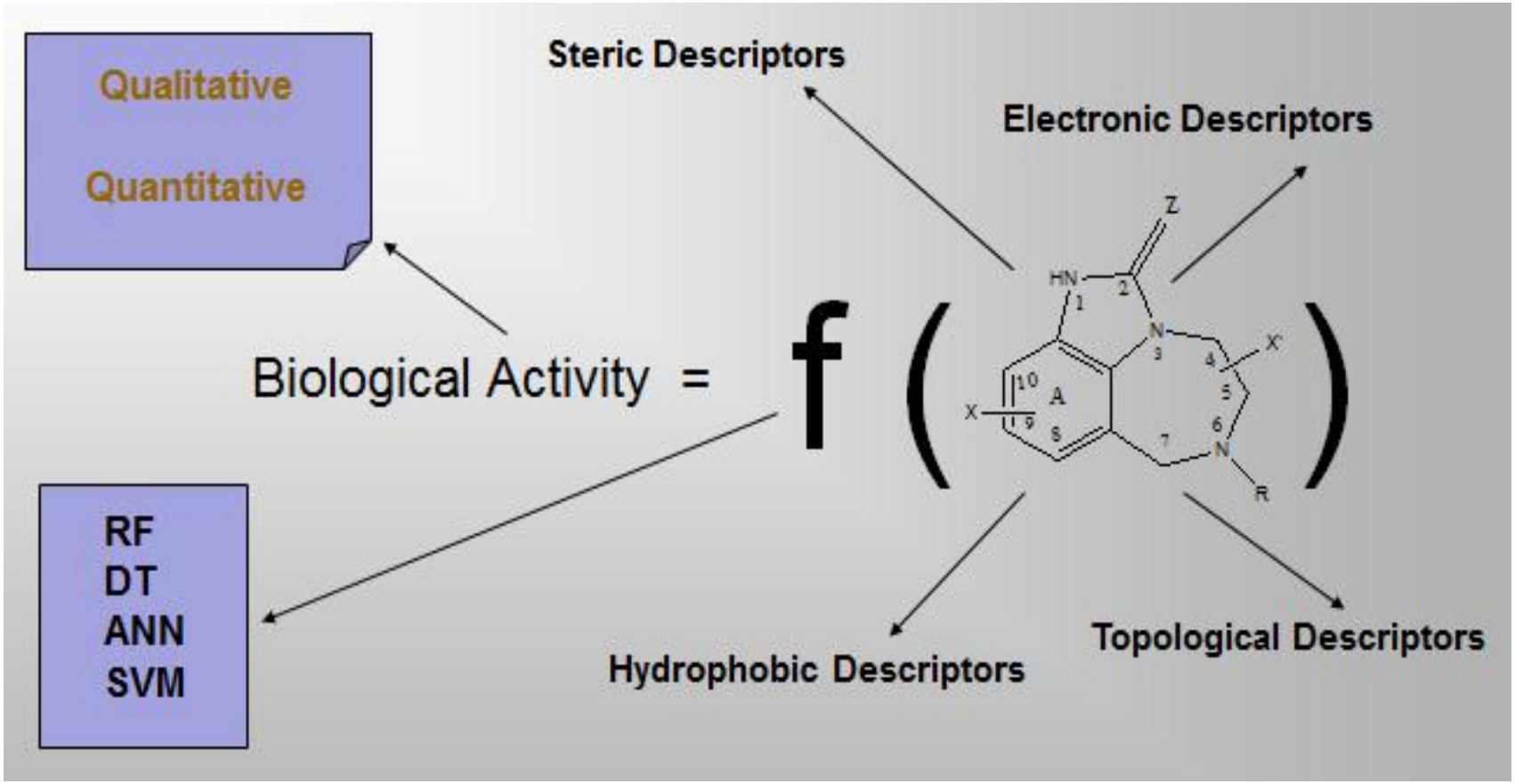
[5] Institute of Chemistry, Faculty of Natural Science and Mathematics, Ss. Cyril and Methodius University, Skopje, Macedonia

•Corresponding author: i.hdoufane@gmail.com

MARRAKECH
جامعة القاضي عياض
UNIVERSITÉ CADI AYYAD

كليـــــــة العلوم السملالية - مراكش
FACULTÉ DES SCIENCES SEMLALIA - MARRAKECH

SAR studies for the *in silico* prediction of HIV-1 inhibitors



Philosophy of Classification-SAR

Abstract:

Tetrahydroimidazo[4,5,1jk][1,4]benzodiazepine (TIBO), as non-nucleoside analogues, constitute potent inhibitors of HIV-1 reverse transcriptase. In the present study, classification structure-activity relationship (C-SAR) models are developed to distinguish between high and low anti-HIV-1 inhibitors of these compounds. Different classifiers, such as support vector machines, artificial neural networks, random forests and decision trees have been established by using ten molecular descriptors. All models were validated using several strategies: internal validation, Y-randomization, and external validation. The correct classification rate ranges from **97%** to **100%** and from **70%** to **90%** for the training and test sets, respectively. A comparison between all methods was done in order to evaluate their performances. The contribution of each descriptor was evaluated to understand the forces governing the activity of this class of compounds.

Keywords: SAR, TIBO, SVM, ANN, DT.

Abbreviations

TIBO	Tetrahydroimidazo[4,5,1jk][1,4]benzodiazepine
C-SAR	Classification Structure Activity Relationship
HIV	Human immunodeficiency virus
RT	Reverse Transcriptase
SVM	Support Vector Machines
ANN	Artificial Neural Networks
RT	Random Forests
DT	Decision Trees

Outlines

1- Introduction

2- Classification-SAR (C-SAR) of TIBO derivatives

- Data set used
- Description of Molecular Structure

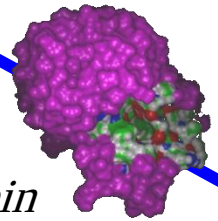
3- Results and Discussion

- Computational Methods Used in C-SAR
- C-SAR Validation

4- Conclusion

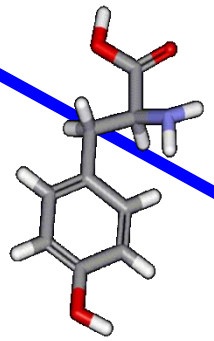
Introduction

Identify disease



Isolate protein involved in disease (2-5 years)

Drug Design
- Molecular Modeling
- Virtual Screening



Find a drug effective against disease protein (2-5 years)

Preclinical testing (1-3 years)

Scale-up

File IND

Human clinical trials (2 -10 years)

Formulation

File NDA

FDA approval (2-3 years)

IND : Investigational New Drug Application

NDA : New Drug Application

Introduction

Creating new medicines requires :

- Enormous investment in terms of *time and money*

- Large team of scientists with training in many different scientific disciplines including various areas of chemistry, biology, engineering, informatics and medicine.

Drug Design (Rational Drug Design or Computer-Aided Drug Design)

Structure-based (SBDD) and ligand-based (LBDD) drug design are extremely important and active areas of research

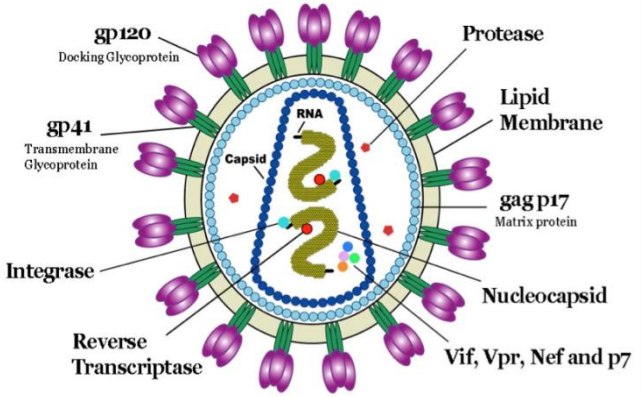
C-SAR (Classification Structure-Activity Relationship)

Introduction

Acquired Immunodeficiency Syndrome (AIDS) has become the center of interest of several studies due to its massive spread all over the world

↓

Reverse transcriptase (RT) is one of the most important enzymes that plays a key role in the replicative cycle of HIV



↓

Non-nucleoside reverse transcriptase inhibitors (NNRTI) are compounds that show great promise in the therapy of HIV infection

↓

4,5,6,7-Tetrahydro-5-méthylimidazo[4,5,1-jk][1,4]benzodiazepin-2(1H)-ones (TIBO)

Goal of this study

Our objective is to propose classifiers that can be able to classify TIBO compounds into two groups: high "*H*" and low "*L*" active compounds, and then to find the variables responsible for this classification.

Use of ANN, SVM, DT and RF in a C-SAR Study of Anti-HIV-1 inhibitors

Data set

The data set in this work consists of a set of 89 TIBO derivatives. The common structure of the compounds used is given in Figure 1.

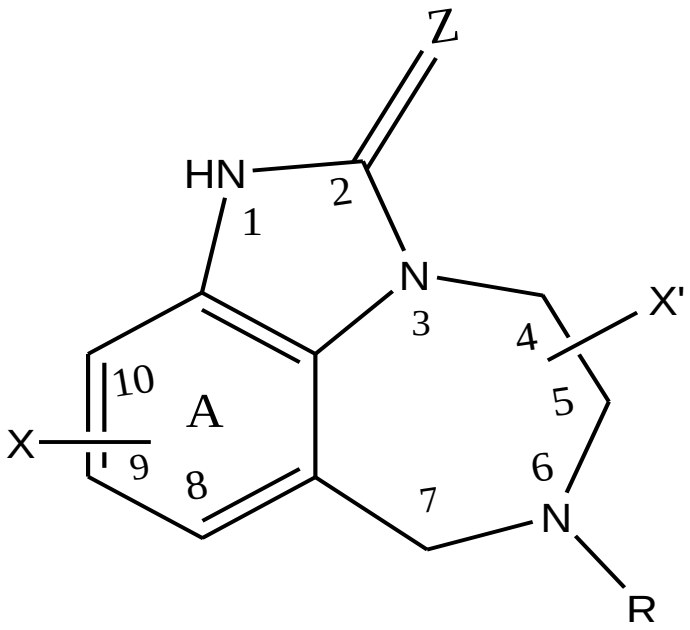


Figure 1: TIBO derivatives

Biological activity

IC_{50} is the effective concentration and refers to the concentration required to achieve 50 % inhibition of the enzyme (RT).

The logarithm of the inverse of this parameter has been used as biological end points (log $1/IC_{50}$) in the C-SAR studies

Since it is a classification (qualitative) study, the original dependent variable (log ($1/IC_{50}$)) was divided into two classes:

- Class H includes compounds with high activities
- Class L contains compounds with low activities

Molecular Descriptors

* Many descriptors were calculated

* Stepwise multiple regression procedure based on the forward-selection and backward-elimination methods for inclusion or rejection of descriptors in the screened models

7 molecular descriptors

MD1	MD2	MD3	MD4	MD5	MD6	MD7

Description of Molecular Structure

Molecular Descriptors

Table 1: List of the selected molecular descriptors and their physical–chemical meanings

Descriptors	Chemical meaning
MD1	logP: Octanol/Water partition coefficient calculated for the whole molecule
MD2	Average nucleophyl reaction index for a N atom
MD3	Minimum total interaction for a H-N bond
MD4	Minimum (>0.1) bond order of a N atom
MD5	ESP-HBSA H-bonding surface area
MD6	Maximum atomic state energy for a N atom
MD7	$^{3}\chi$: molecular connectivity index to the third order

Three other descriptors (MD8 = I_R, MD9 = I_Z and MD10 = I_X) have been added

- I_R = 1 if R = 3, 3-dimethyallyl and I_R = 0 for others (see figure 1)
- I_Z = 1 if Z = Sulphur and I_Z = 0 if Z = Oxygen (see figure 1)
- I_X = 1 for position 8, I_X = 0.5 for position 9 and I_X = 0 for position 10 (see figure 1)

Results and Discussion

Table 2: Classification results of the training and the test sets for the all methods.

Methods	Training set			Test set		
	Total accuracy %	High samples %	Low samples %	Total accuracy %	High samples %	Low samples %
ANN	98.60	96.43	100.00	90.00	83.3	100.00
DT	97.10	96.43	97.56	70.00	66.7	75.00
SVM	100.00	100.00	100.00	85.00	84.62	85.71
RF	100.00	100.00	100.00	75.00	75.00	75.00

Table 2 shows good classification for all established models.

Table 3: Results of randomization test for the developed models using LOO-CV.

Methods	Total accuracy (%)	
	Real models	Random models
ANN	94.20	53.62
DT	92.80	49.28
SVM	92.80	60.87
RF	95.65	52.17

⟹ the total accuracy of randomization test is lower than the corresponding one obtained for the real models and thus it excluded the possibility of chance correlation

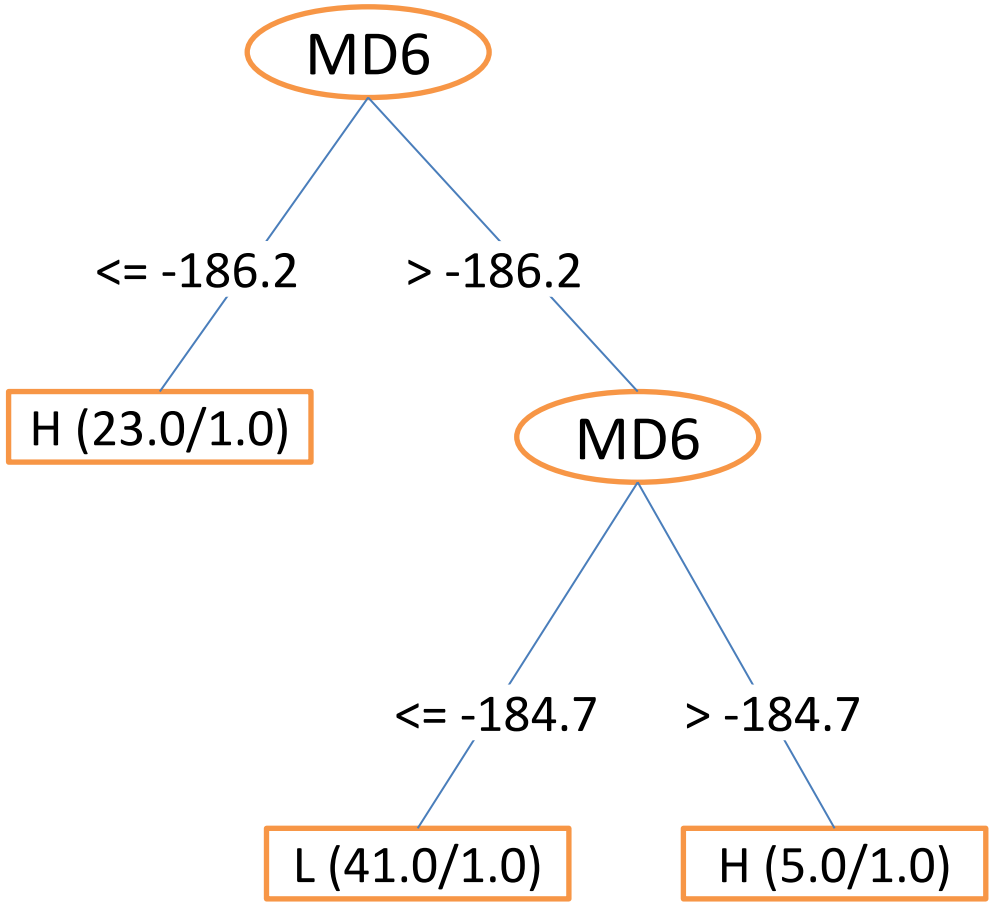
Anti-HIV Study of TIBO Using ANN, SVM, DT and RF

Table 4: Misclassified samples by ANN, DT, SVM and RF

Methods	Sets	The number of misclassified compounds
ANN	Training set	89
	Test set	30, 83
DT	Training set	60,65
	Test set	26,30,32,66,67,81
SVM	Training set	
	Test set	26,30,32
RF	Training set	
	Test set	26,30,32,81,83

Regarding the misclassified compounds: compound 30 is common to all methods and compounds 26, 30 and 32 are common to DT, SVM and RF.

According to the results shown in table 4 we remark that compounds 26, 30 and 32 cannot be correctly classified. It is difficult to find a reason for why the model failed to predict them accurately. We think that the values of inhibitory activity of these compounds, which are close to that taken as a reference, can explain the inability of the models to accurately predict their classes



From this tree we can observe that only one descriptor is enough to build this classification model using the descriptor MD6 (*Maximum atomic state energy for a N atom*).

Figure 2: Decision tree

Anti-HIV Study of TIBO Using ANN, SVM, DT and RF

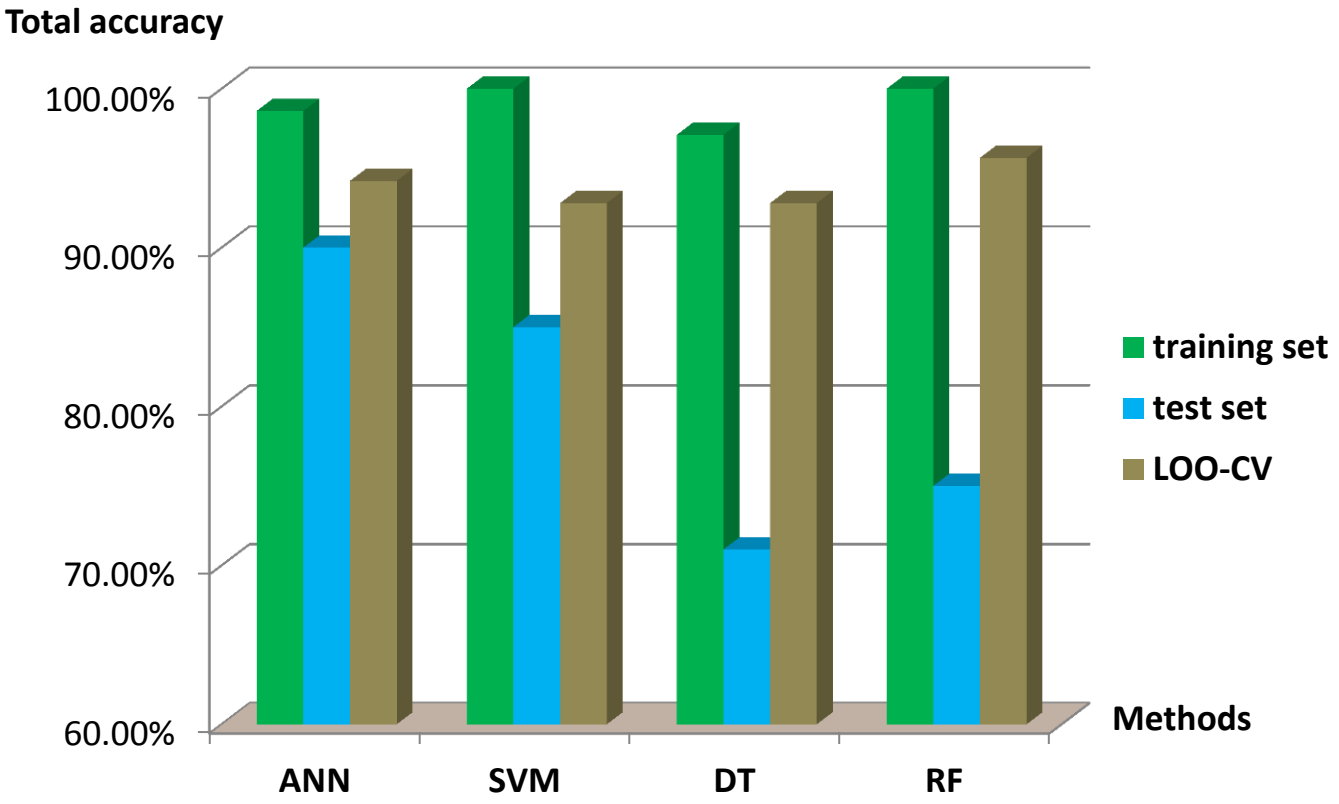


Figure 3: Comparison of the accuracy for three data sets by ANN, DT, SVM and RF

Comparison between all methods by using an external test set demonstrates that the performance of ANN model is better than that of SVM, DT and RF.

Figure 3 indicates that all methods give similar good results in the LOO-CV (92.80 % - 95.65 %) procedure and in the training set (97.1 % - 100 %). However, for the external validation test set, we can notice that the ANN model gives better results than the ones obtained by SVM, DT and RF.

ANN can handle problems involving imprecise or “ noisy ” data as well as problems that are highly non-linear and complex.

Anti-HIV Study of TIBO Using ANN, SVM, DT and RF

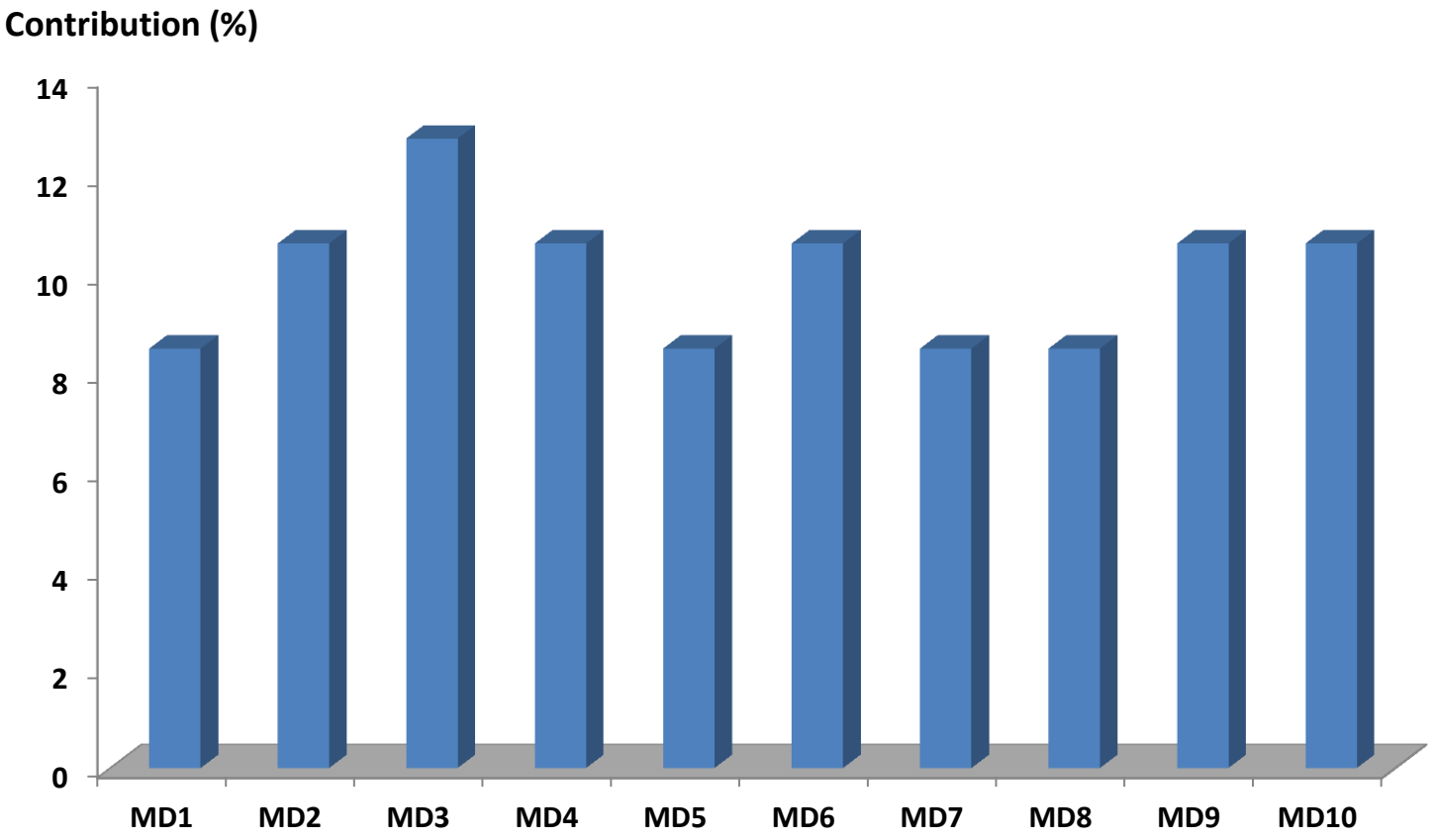


Figure 4: Contributions of molecular descriptors to C-SAR by SVM model

The contribution rate showed the relative importance of each descriptor comparing with the other descriptors. The results shown in figure 4 indicate that all descriptors contribute with an important rate. However, the descriptor MD3 (Minimum total interaction for a H-N bond) descriptor exhibits the largest contribution to inhibition effects among the ten descriptors.

Conclusion

In the present work, four methods ANN, DT, SVM and RF, were used to develop C-SAR models of anti-HIV-1 TIBO derivatives.

The established models by all methods show good classification rate ranges of the studied compounds. The comparison between these methods on the external validation test set demonstrates that the performance of ANN model is better than that of SVM, DT and RF.

The established classification models can be used in biological screening processes and in prediction of the anti-HIV activities (or other molecular properties) of untested molecules.

References

1) Exploring QSAR of Non-Nucleoside Reverse Transcriptase Inhibitors by Neural Networks
L. Douali, D. Villemin, and **D. Cherqaoui**, pp. 48–55
*International Journal of Molecular Sciences ISSN 1422-0067 © **2004** by MDPI*

2) Support vector machines: Development of QSAR models for predicting anti-HIV-1 activity of TIBO Derivatives
Rachid Darnag, EL Mostapha Mazouz, Andreea Schmitzer, Didier Villemin, Abdellah Jarid, **Driss Cherqaoui**
European Journal of Medicinal Chemistry, **2010,** 45, 1590-1597

3) QSAR studies of TIBO derivatives using support vector machines
*R. Darnag; A. Schmitzer; Y. Belmiloud; D. Villemin; A. Jarid; A. Chait; E. Mazouz; **Driss Cherqaoui***
*SAR and QSAR in Environmental Research, **2010**,21 (3-4), 231-246.*

Acknowledgments

- *Didier Villemin (ENSI Caen France)*
- *Alia Tadjer (Faculty of Chemistry and Pharmacy, Sofia, Bulgaria)*
- *Imane Bjij (School of health Sciences, South Africa)*
- *Mahmoud E.S Soliman (School of health Sciences, South Africa)*
- *Jane Bogdanov (*Institute of Chemistry, Faculty of Natural Science and Mathematics, Macedonia
- *Driss Cherqaoui (Faculty of Sciences Semlalia Marrakech, Maroc)*

This study was part-supported by The " *Agence Universitaire de la Francophonie* " and the Scientific Research Fund of Bulgaria

Thank you

* i.hdoufane@gmail.com
Laboratory of Molecular Chemistry

Faculty of Science Semlalia
Marrakech