
QSAR for antimycobacterial activity of β -thia adduct of chalcone and diazachalcone derivatives

Younes Abrouki*, Abdelkader Anouzla, Hayat Loukili, Ahmed Rayadh, Driss Zakarya, Mohamed Zahouily

Department of Research, Faculty of Science and Technology, University Hassan II, Mohammedia, Morocco

Email address:

younesabrouki@gmail.com (Y. Abrouki)

To cite this article:

Younes Abrouki, Abdelkader Anouzla, Hayat Loukili, Ahmed Rayadh, Driss Zakarya, Mohamed Zahouily. QSAR for Antimycobacterial Activity of β -Thia Adduct of Chalcone and Diazachalcone Derivatives. *International Journal of Computational and Theoretical Chemistry*. Vol. 2, No. 3, 2014, pp. 20-25. doi: 10.11648/j.ijctc.20140203.11

Abstract: Quantitative structure-activity antimycobacterial relationships have been studied for a series of β -thia adduct of chalcone and diazachalcone derivatives by means of multiple linear regression (MLR) and artificial neural networks (ANN). The antimycobacterial activity against *M. tuberculosis* H37Rv of the compounds studied was well correlated with descriptors encoding the chemical structure. Using the pertinent descriptors revealed by a stepwise procedure in the multiple linear regression technique, a correlation coefficient of 0.9798 ($s=0.0869$) for the training set was obtained for the ANN model in a [3-3-1] configuration. The results show that the antimycobacterial activity of these compounds is strongly dependent on hydrogen-bonding donors, molecular refraction and also molecular connectivity indices for 2nd order.

Keywords: QSAR, MLR, ANN, Antimycobacterial, β -Thia Adduct of Chalcone, Diazachalcones

1. Introduction

Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis* [1]. It has become a major health problem being the second leading cause of death worldwide, after human immunodeficiency virus (HIV) [2].

According to the Global Tuberculosis Report, 2012 by World Health Organization, 8.7 million new cases of TB, 13% of which were co-infected with HIV and 1.4 million deaths from TB were estimated in 2011. TB is most prevalent in Asia and Africa with India and China alone accounting for about 40% of the global cases [3].

Today many drugs are available for clinical use to treat TB, but the current treatment lasts for six to nine months. During the course of treatment, the pathogen develops resistance against these drugs which results in Multi-Drug resistant Tuberculosis and eventually lead to untreatable extensively drug resistant Tuberculosis [4].

To overcome the problem of growing drug resistance, identification of new targets which are essential for survival and replication of the pathogen has become an urgent need. For the purpose of finding drugs against novel targets we require fast and reliable computational techniques for cost-effective evaluation of large virtual databases of chemical compounds in order to identify a limited set of

candidates which can be synthesized and examined experimentally for their biological activity [5].

Quantitative structure activity relationship (QSAR) is a powerful approach being used to establish a correlation between the physicochemical properties of the chemical compounds and their biological activity to obtain a reliable statistical model [6].

This model serves as a valuable tool for the design of new chemical entities and to predict their activity.

The QSAR [7] model so developed facilitates identification of promising lead candidates, thus decreasing the number of compounds required to be synthesized and tested *in vitro*.

In the present work, we use a combination of Multiple Linear Regression (MLR) [8] and Artificial Neural Network (ANN) [9] techniques for modeling the observed antimycobacterial activity for a series of β -thia adduct of chalcone and diazachalcone derivatives (Fig. 1).

The pertinent variable descriptors [10] selected by the first method are introduced as input neurons in the ANN architecture to optimize their non-linear combinations.

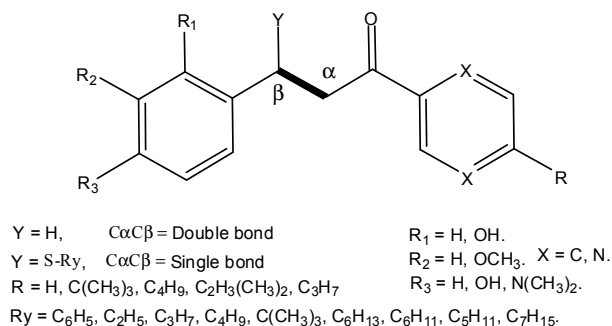


Figure 1. General structure of β -thia adduct of chalcone and diazachalcone derivatives.

The objectives of our study were both to provide supplementary information concerning the behavior of these compounds and further define the criteria necessary for the rational design of a new generation of anti-tuberculosis drug.

2. Material and Methods

2.1. Statistical Methods

2.1.1. Multiple Linear Regression

The MLR [11^a] was used to generate linear models between the antimycobacterial activity and the molecular descriptors. Because of the large number of descriptors considered, a stepwise MLR procedure based on the forward selection and backward-elimination methods was used to select the powerful variable descriptors.

In order to avoid all difficulties in the interpretation of the resulting models, pairs of variables with a correlation coefficient greater than 0.8 were classified as intercorrelated, and only one of these was included in the screened model. The validity of the model was proven by the multiple correlation coefficients (r), the standard deviation (s) and the F-test value. The reliability of the model was indicated in terms of predictive Q .

2.1.2. Artificial Neural Network

As biological phenomena are considered non-linear by nature, it therefore appears very interesting to study the present series of compounds with the ANN technique [11^b], in order to discover the possible existence of nonlinear relationships between antimycobacterial activity and molecular descriptors that appeared pertinent for the linear model.

The ANN was trained by the back-propagation of errors algorithm and had the following architecture:

- An input layer including the pertinent descriptors of MLR.
- A hidden layer for which the ratio of the number of data points in the training set and the number of connections controlled by the network, ρ , is critical to the predictive power of the neural net. The range $1.8 < \rho < 2.2$ [$\rho = (\text{number of data points in the training set})/(\text{number of$

adjustable weights controlled by the network)], was used as a guideline for an acceptable number of neurons in the hidden layer. It is claimed that, for $\rho \ll 1.0$, the network simply memorises the data, whereas for $\rho \gg 3.0$, the network loses its ability to generalize.

- An output layer of one neuron, representing the antimycobacterial activity. The input and output values were normalized.

After this step, the learning rate was varied from 0.01 to 0.9 and for each learning rate the momentum was examined from 0.1 to 0.9.

The number of the neurons in the hidden layer with the use of optimized momentum and learning rate was determined.

Finally, to preclude training we have studied the variation of the root mean-squares error versus number of iteration and we have used two strategies for testing the validity of the selected ANN model.

2.2. Molecular Descriptors

Computational chemistry resides in parameterising the variation in chemical structure. It is evident that in the context of pharmaceutical research, the molecular descriptors should be candidates since they have been successfully used in the modeling. In this study a set of descriptors related to physicochemical geometries, energies and associated properties of the molecules was used in order to study their influence on the inhibition activity of these compounds.

These include the octanol/water partition coefficient ($\log P$) used as a descriptor of the hydrophobic molecular properties, hydrogen-bonding donors (HBD), hydrogen-bonding acceptors (HBA), Molar refractivity (MR), Molecular connectivity indices (IC₀ for 0th order, IC₁ for 1st order, IC₂ for 2nd order), Van der Waals volume (V), Molecular weight (MW), Electronegativity and other molecular descriptors, All these descriptors were calculated with the software MMP Molecular Modeling Program [11^c].

2.3. Biological Data

The biological data for β -thia adduct of chalcone[12] and diazachalcone derivatives[13] used in this study were obtained from various studies.

All the derivatives were tested for antimycobacterial activity in vitro against *Mycobacterium tuberculosis* H₃₇Rv.

The value of $\log(\%IT)$ was used for subsequent QSAR analysis as dependent variable in which $\%IT$ represents the percent inhibition of *Mycobacterium tuberculosis* H₃₇Rv using the minimum inhibitory concentration of the product.

The basic structures of these compounds are shown in figure 1.

The $\log(\%IT)$ values of these compounds are listed in table 1.

Table 1. Chemical structure of β -thia adduct of chalcone and diazachalcone derivatives and observed antimycobacterial activities

N°	R ₁	R ₂	R ₃	R	X	Y	(C _α C _β) [*]	log (%IT)
01	H	H	H	H	C	SC ₆ H ₅	1	1.897627091
02	H	H	H	H	C	SC ₂ H ₅	1	1.897627091
03	H	H	H	H	C	SC ₃ H ₇	1	1.897627091
04	H	H	H	H	C	SC ₄ H ₉	1	1.924279286
05	H	H	H	H	C	SC(CH ₃) ₃	1	1.770852012
06	H	H	H	H	C	SC ₆ H ₁₃	1	1.86332286
07	H	H	H	H	C	SC ₆ H ₁₁	1	1.924279286
08	H	H	H	H	C	SC ₅ H ₁₁	1	1.944482672
09	H	H	H	H	C	SC ₇ H ₁₅	1	1.748188027
10	H	OCH ₃	OH	H	N	H	2	1.544068044
11	H	OCH ₃	OH	C(CH ₃) ₃	N	H	2	1.886490725
12	H	OCH ₃	OH	C ₂ H ₅ (CH ₃) ₂	N	H	2	1.72427587
13	H	OCH ₃	OH	C ₄ H ₉	N	H	2	1.653212514
14	H	OCH ₃	OH	C ₃ H ₇	N	H	2	1.301029996
15	OH	H	H	H	N	H	2	1.740362689
16	OH	H	H	C(CH ₃) ₃	N	H	2	1.973127854
17	OH	H	H	C ₂ H ₅ (CH ₃) ₂	N	H	2	1.301029996
18	OH	H	H	C ₄ H ₉	N	H	2	1.51851394
19	OH	H	H	C ₃ H ₇	N	H	2	1.698970004
20	H	H	OH	H	N	H	2	1.62324929
21	H	H	OH	C(CH ₃) ₃	N	H	2	1.857332496
22	H	H	OH	C ₂ H ₅ (CH ₃) ₂	N	H	2	1.799340549
23	H	H	OH	C ₄ H ₉	N	H	2	1.707570176
24	H	H	OH	C ₃ H ₇	N	H	2	1.544068044
25	H	H	N(CH ₃) ₂	H	N	H	2	1.414973348
26	H	H	N(CH ₃) ₂	C(CH ₃) ₃	N	H	2	0.84509804
27	H	H	N(CH ₃) ₂	C ₂ H ₅ (CH ₃) ₂	N	H	2	1
28	H	H	N(CH ₃) ₂	C ₄ H ₉	N	H	2	0
29	H	H	N(CH ₃) ₂	C ₃ H ₇	N	H	2	0.903089987

* (1 = Single bond; 2 = Double bond).

3. Results and Discussion

3.1. Multiple Linear Regression Analysis

After variable selection, we have centered and reduced the values of pertinent descriptors obtained in order to have homogeneity in our data.

MLR was performed on all 29 molecules of the training set for the model generation, a few suitable models were obtained, and the best model Eq. (1) with minimal complexity was selected.

$$\log (\%IT) = 1.5829 - 0.2764 * MR + 0.1663 * IC2 - 0.4427 * HBD \quad (1)$$

$$n=29 \quad r=0.857 \quad s=0.237 \quad F=23.19$$

To remain close to the experimental error (5%), we take away all molecules with error ($di > 2s/\text{Mean}$).

Consequently, a new regression model was derived using 28 molecules Eq. (2):

$$\log (\%IT) = 1.6161 + 0.1049 * IC2 - 0.1774 * MR - 0.3450 * HBD \quad (2)$$

$$n=28 \quad r=0.861 \quad s=0.171 \quad F=22.85$$

This equation (2) shows a slight increase in the correlation coefficient and a decrease in the standard deviation value.

It is noteworthy that there is no significant intercorrelation between the descriptors that appear in the model selected, as seen in table 2.

Table 2. Inter correlation matrix

	Constant	MR	IC2	HBD
HBD	-0.1320	+0.5765	-0.5424	1
IC2	-0.3847	-0.7780	1	
MR	-0.2700	1		
Constant	1			

Statistical criteria of the model are fairly good. Indeed we have a model with 74.13% of the total variance and a standard deviation lower than associated with the mean value of log (%IT), as seen in table 3.

Table 3. Regression variance analysis for the model

	Coefficient	Standard error	Sum of squares	v	F exp.	Prob.
HBD	-2.1278	0.2675	3.5405	1	-7.9518	0.0000
IC2	0.1817	0.0791	0.3225	1	2.2966	0.0303
MR	-0.0266	0.0071	0.0324	1	-3.7125	0.0010
Constant	2.9330	0.4531	-	1	6.4731	0.0000
Regression	-	-	1.9999	3	22.85	0.0040
Residue	-	-	0.7002	24	-	-
Total	-	-	2.7002	27	-	-

The descriptors contributions to this equation calculated according to the Gore method, also justified this result, as seen in table 4.

Table 4. Descriptors contribution

Descriptors	Contribution (%)
MR	34
IC2	19
HBD	47

The results summarized in Table 4, show a large contribution to the hydrogen-bonding donors. The contribution of molecular refraction reflects the strong influence of both the size and polarizability of the antimycobacterial activity of our series. These two descriptors negatively contribute to this activity. The Molecular connectivity indices for 2nd order, which reflects the influence of molecular form, contributes positively to the antimycobacterial activity.

To better assess the correlation between the values calculated by the model and experimental data, we plotted the curve calculated values based on observed values from Equation 2. The results are shown in Figure 2, which shows a significant correlation between the calculated and observed values of $\log(\%IT)$.

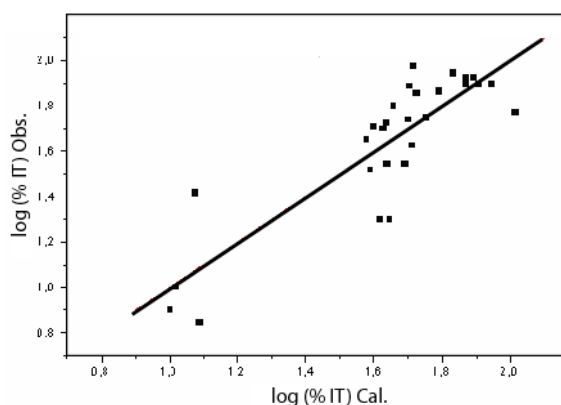


Figure 2. Values of $\log\%IT$ observed and calculated by MLR.

In the cross-validation phase, 28 subsets were created according to the leave-one-out method and the output of the removed compound was predicted for each subset [14].

A cross-validation coefficient (Q) was calculated according the following equation (3):

$$Q = 1 - (\text{PRESS}/\text{Variance}) \quad (3)$$

Where, PRESS is the predictive residual sum of squares. They yielded a $Q = 0.68$, indicating good predictive quality of the model, according to Wold ($Q > 0.6$).

3.2. Artificial Neural Network Analysis

The ANN was generated by using the pertinent descriptors appearing in the MLR model as input.

A (3-3-1) neural network architecture was developed with the optimum momentum and learning rate of 0.9 and 0.3,

respectively and with 50.000 iterations (Fig. 3).

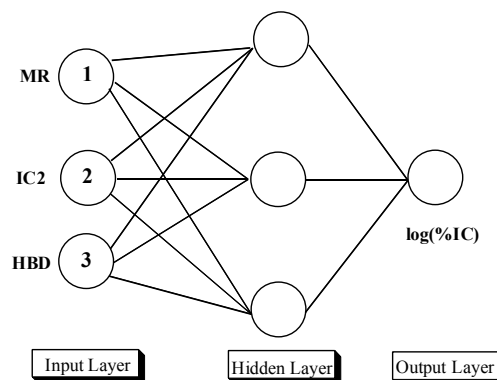


Figure 3. Typical architecture of a three layer artificial neural network.

The three hidden neurons were chosen to maintain ρ between 1.8 and 2.2. Under these conditions, we obtained a good correlation ($r = 0.9798$).

To evaluate the neural network, the correlation coefficient r of its results is compared with the r for the regression model developed in this work (Table 5).

Table 5. Comparison of model results of MLR and ANN

Model	s	r^2	r
MLR	0.1708	0.7413	0.8610
ANN	0.0869	0.9601	0.9798

The r values were 0.8610 and 0.9798 for the training set in the present MLR and the ANN, respectively. The corresponding standard error s for the two analyses was 0.1708 and 0.0869 respectively. This reveals the improvement of the MLR model.

The corresponding Q values in cross-validation methods were 0.68 and 0.726 in MLR and ANN, respectively. These results indicate that both MLR and ANN have good predictive ability.

To better compare the predictive power of our model obtained by MLR or ANN, we plotted the curves of the activities observed and predicted based on the sequence of the sample (Figure 4). The plot in Fig. 4 indicates that the results of the ANN are more efficient than the MLR.

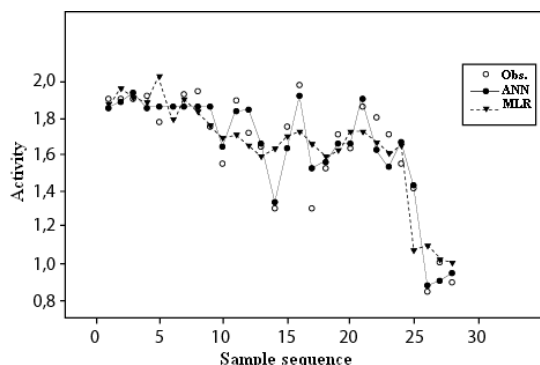


Figure 4. Comparison of values of $\log(\%IT)$ observed and calculated by the MLR and ANN.

To estimate the relative contribution of descriptors, we have chosen two different approaches: The first is to analyze the deviations between observed and calculated activities of the ANN, discarding each time a descriptor. The second is to remove each time a descriptor and then calculate its contribution using the equation Chastrette [15].

Table 6. Evaluating the impact of descriptors in MLR and ANN

Descriptors	Contribution (%)	
	MLR	ANN
MR	34	29
IC2	19	28
HBD	47	43

The results, summarized in Table 6, show that both methods give the same ranking of descriptors: HBD > MR > IC2.

The comparison of this classification with that obtained in the linear model shows a change in the percentage contribution but keeping the same classification.

We can explain this change by the possible existence of non-linear relationship between activity and Molecular connectivity indices for 2nd order. In other words, the molecular shape contributes positively to a non-linear manner, on the anti-tuberculosis activity of our series.

3.3. QSAR Model Anti-Tuberculosis Drug

These very important results clearly indicate that the descriptors selected for this study adequately describe the activity and allow us to build an efficient QSAR Model anti-tuberculosis drug (Fig. 5).

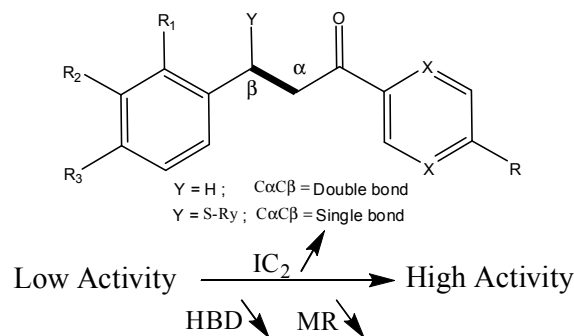


Figure 5. QSAR Model anti-tuberculosis drug.

This model based on the molecular descriptors of the sample for a series of β -thia adduct of chalcone and diazachalcone derivatives, can help in the design of new molecules more active against *Mycobacterium tuberculosis* H₃₇Rv by varying the physico-chemical descriptors.

4. Conclusions

In this series, the antimycobacterial activity of the above compounds was investigated by means of MLR and ANN techniques.

The results of the QSAR study obtained in this work

indicate that the antimycobacterial activity of β -thia adducts of chalcone and diazachalcone derivatives depend strongly on hydrophobic character, hydrogen-bond acceptors and also steric factors.

Our QSAR model antimycobacterial can be a tool to assist in the design of new generation of anti-tuberculosis drug with a better percentage inhibition of *M. tuberculosis* H₃₇Rv.

References

- [1] Brennan, P.J., Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. Tuberculosis, 2003, 83(1), 91-97.
- [2] Greenberg, A. E., Lucas, S., Tossou, O., Coulibaly, I. M., Coulibaly, D., Kassim, S., Ackah, A., De Cock, K. M., Autopsy-proven causes of death in HIV-infected patients treated for tuberculosis in Abidjan, Cote d'Ivoire. AIDS, 1995, 9(11), 1251-1254.
- [3] Global Tuberculosis Report, 2012 by World Health Organization, <http://www.who.int>.
- [4] Gandhi NPNR, Dheda K, Schaaf HS, Zignol M, Soolingen DV, Jensen P, Bayona J., Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. Lancet 2010, 375:1830-1843.
- [5] Cole, S. T., Telenti, A., Drug resistance in *Mycobacterium tuberculosis*. European Respiratory Review, 1995, 8, 701-713.
- [6] Cramer, R., A QSAR success story. Chemtech. 1980, 10, 744-747.
- [7] Verma J KV, Coutinho EC: 3D-QSAR in drug design. Curr Top Med Chem 2010, 10(1):95-115.
- [8] Draper, N.R.; Smith, H. In: Applied Regression Analysis, John Wiley & Sons: New York, 1981.
- [9] Yap, C.W., Chen, Y.Z., Quantitative structure pharmacokinetic relationships for drug distribution properties by using general regression neural network. J. Pharm. Sci. 2005, 94, 153-168.
- [10] Karelson, M., Molecular Descriptors in QSAR/QSPR; John Wiley & Sons: New York, 2000.
- [11] (a) Unistat statistical package, version 4.0 for Excel. (b) Data pro Qnet 2000 for Windows V2 K build neural network modeling. Vesta Service, Winnetka, III. (c) MMP, molecular modeling pro-Demo (TM) Revision 301 demo published by ChemSW Software (TM).
- [12] Chlupačova, M., Kubanova, P., Opletalova, V., Buchta, V., The Importance of the Enone-moiety for Antimycobacterial and Antifungal Properties of Chalcones, Published in: Proceedings of the 3rd International Symposium on Natural Drugs, Naples, 2-4 October 2003. Borelli, F., Capasso, F., Milic, N., Russo, A. (Eds.). Universita degli Studi di Napoli Federico II, Naples - Indena, Milano 2003, pp. 133-135.
- [13] Opletalova, V., Hartl, J., Patel, A., Palat, K., Buchta, V., Ring substituted 3-phenyl-1-(2-pyrazinyl)-2-propen-1-ones as potential photosynthesis-inhibiting, antifungal and antimycobacterial agents. Il Farmaco, 2002, 57, 135-144.

- [14] Wold, S., Dunn, W. J., Multivariate quantitative structure-activity relationships (QSAR): conditions for their applicability. *J. Chem. Inf. Comput. Sci.*, 1983, 23, 6-13.
- [15] Chastrette, M., Zakarya, D., Peyraud, J. F., Structure-musk odor relationships for tetralins and indans using neural networks (on the contribution of descriptors to the classification). *Eur. J. Med. Chem.*, 1994, 29, 343-348.