Journal of Molecular Graphics and Modelling 28 (2009) 28–36

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

QSAR analysis for quinoxaline-2-carboxylate 1,4-di-*N*-oxides as anti-mycobacterial agents

Esther Vicente [a], Pablo R. Duchowicz [b,*], Eduardo A. Castro [a], Antonio Monge [a]

[a] Unidad de Investigación y Desarrollo de Medicamentos, Centro de Investigación en Farmacobiología Aplicada (CIFA), University of Navarra, C/Irunlarrea s/n, 31008 Pamplona, Spain
[b] Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (UNLP, CCT La Plata-CONICET), Diag. 113 y 64, C.C. 16, Suc.4, (1900) La Plata, Argentina

ARTICLE INFO

Article history:
Received 24 January 2009
Received in revised form 18 March 2009
Accepted 21 March 2009
Available online 31 March 2009

Keywords:
QSAR
Multivariable linear regression
Replacement method
Quinoxaline oxides
Tuberculosis

ABSTRACT

In a continuing effort of our research group to identify new active compounds against *Mycobacterium tuberculosis*, we resort to the quantitative structure–activity relationships (QSARs) theory. For this purpose, we employ certain parameters of potency, cytotoxicity and selectivity as given by the Tuberculosis Antimicrobial Acquisition & Coordinating Facility (TAACF) program. The molecular structure of 43 quinoxaline-2-carboxylate 1,4-di-*N*-oxide derivatives is appropriately represented by 1497 DRAGON type of theoretical descriptors, and the best linear regression models established in this work are demonstrated to result predictive. The application of the QSAR equations developed now serves as a rational guide for the proposal of new candidate structures that still do not have experimentally assigned biological data.

1. Introduction

Tuberculosis (TB), caused by *Mycobacterium tuberculosis*, is a respiratory transmitted disease affecting nearly 32% of the world's population, more than any other infectious disease. There were an estimated 8.8 million new TB cases in 2005. A total of 1.6 million people died of TB, including 195,000 patients infected with HIV [1]. The World Health Organization (WHO) estimates that within the next 20 years about 30 million people will be infected with the bacillus [2]. The development of resistance by *M. tuberculosis* to commonly used anti-tuberculosis drugs, the increasing incidences of disease in immuno-compromised patients, and longer durations of therapy that are required as a result of resistance development, highlights the need for new drugs to extend the range of effective TB treatment options [3].

The quinoxaline is described as a bioisoster of quinoline, naphthalene, benzothiophene and other aromatic rings such as pyridine and pyrazine [4]. Because of the similarity between some antitubercular drugs and quinoxaline (Fig. 1), as well as the presence of the quinoxaline moiety in some broad spectrum antibiotics, it was hoped that quinoxaline analogs would exhibit antitubercular activity.

As a result of the anti-tuberculosis research project, our group published several studies in which the synthesis and biological evaluation of a large amount of quinoxaline and quinoxaline 1,4-di-*N*-oxide derivatives have been described [5–15]. From these studies, several quinoxaline-2-carboxylate 1,4-*N*-oxide derivatives, with different patterns of substituents at quinoxaline nucleus, were prepared and they showed important antitubercular activity *in vitro*, **1**–**43** (Fig. 2) [12,14,15]. We also observed that the lack of the two *N*-oxide groups generally led to the loss of the antimycobacterial activity [5,8]. It has been established that a widely useful anti-tubercular drug must be inexpensive so that it is routinely available to populations in need in developing countries. For this reason, our research group seeks new available and cheap reagents in our continuing efforts to identify new inexpensive anti-TB drug candidates.

Nowadays, the systematization of available information following its subsequent interpretation is considered of crucial importance by the scientific community. An alternative way for overcoming the absence of experimental measurements for biological systems is based on the ability to formulate quantitative structure–activity relationships (QSARs) [16,17]. This theory has as ultimate role, the proposal of a model capable of estimating the activities of compounds, by relying on the assumption that these resulting effects are a consequence of the molecular structure. Since the pioneer studies of Hansch the use of QSAR has become helpful in understanding chemical–biological interactions in drug and pesticide research, as well as in various areas of toxicology. Therefore, the structure is translated into the so-called molecular

* Corresponding author. Tel.: +54 221 425 7430/7291; fax: +54 221 425 4642.
E-mail addresses: pabloducho@gmail.com, duchow@inifta.unlp.edu.ar (P.R. Duchowicz).

1093-3263/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2009.03.004

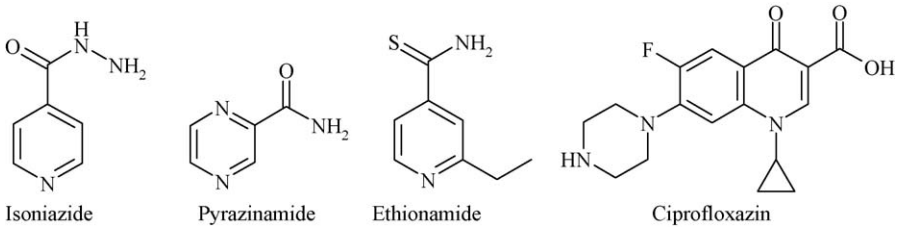

Fig. 1. Structures of certain drugs for treating tuberculosis.

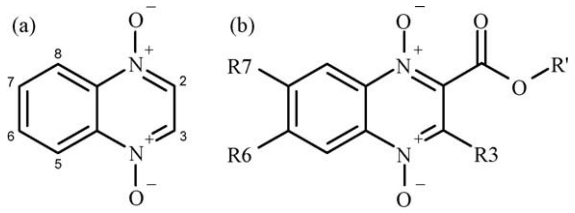

Fig. 2. (a) Numbered quinoxaline 1,4-di-*N*-oxide ring; (b) general structure for quinoxaline-2-carboxylate 1,4-di-*N*-oxide derivatives in Tables 1 and 3.

descriptors, describing different relevant features of the compounds, through mathematical formula obtained from the chemical graph theory, information theory, quantum mechanics, etc. [18] More than thousand available descriptors exist and are reported in the literature, and one has to decide how to select those that characterize the property under consideration in the best possible manner.

Although a large number of structure–activity relationships (SARs) were reported previously for analyzing the biological properties displayed by quinoxaline derivatives, none of these included QSAR results [5–15,19–22]. A main reason of this can be attributed to the fact that in past decades few experimental biological data on quinoxaline compounds were available for building a quantitative model, but nowadays this drawback is continuously being surmounted by different research groups that register new data into the literature. In present study, we developed a further exploration of the biological profile exhibited by quinoxaline-2-carboxylate 1,4-di-*N*-oxide derivatives resorting to the QSAR formalism, establishing predictive models on the current available data for five different biological properties.

2. Methods

2.1. Data set

The experimental information of the quinoxaline compounds appearing in Table 1 was collected from the literature [12,15]. The reported *in vitro* evaluation of the antituberculosis activity was carried out at the GWL Hansen's Disease Center within the Tuberculosis Antimicrobial Acquisition & Coordinating Facility (TAACF) screening program for the discovery of novel drugs for the treatment of tuberculosis. Under the direction of the U.S. National Institute of Allergy and Infectious Disease (NIAID), the Southern Research Institute coordinates the overall program [23]. The Level 1 screen is conducted at 6.25 μg/mL against *M. tuberculosis* H37Rv using the Microplate Alamar Blue Assay. Compounds demonstrating at least 90% inhibition in the Level 1 screen are retested at lower concentrations against *M. tuberculosis* H37Rv to determine the actual minimum inhibitory concentration (MIC). Concurrent with the determination of MICs, compounds are tested for cytotoxicity (IC_{50}) in VERO cells. The selectivity index (SI = IC_{50}/MIC) was also determined; it was considered significant when SI >10. Once the Level 2 screens are completed, selected compounds are considered for further evaluation in the macrophage assay; compounds are tested for killing of *M. tuberculosis* strain Erdman inside mouse bone marrow macrophages. EC_{90} and EC_{99} are determined as the lowest concentration effecting 90% and 99% reduction, respectively, in colony forming units compared to drug-free controls. Compounds with EC_{90}/MIC <16 are considered for further evaluation. At this stage a medical chemistry team reviews the data, considering characteristics such as SAR, solubility, and log *p* values, to determine if evaluation in *in vivo* testing models is warranted [22].

2.2. Molecular descriptors and model search

For the purpose of computing the theoretical structural descriptors for quinoxaline derivatives, we first pre-optimized the structures of the compounds with the molecular mechanics force field (MM+) procedure included in the Hyperchem 6.03 package [24] and after that, we refined the resulting geometries by means of the semiempirical method PM3 (Parametric Method-3) using the Polak-Ribiere algorithm and a gradient norm limit of 0.01 kcal $Å^{-1}$. The molecular descriptors were calculated using the software DRAGON [25] resulting in a pool containing $D = 1497$ numerical variables.

We used the replacement method (RM) [26–29] as variable subset selection approach, an algorithm that generates multivariable linear regression models by minimizing its standard deviation (*S*), and whose results are quite close to the ones obtained with an exact (combinatorial) search of molecular descriptors although requiring much less computational work. The RM gives models with similar statistical parameters than genetic algorithms [30]. In all our calculations, we used the computer system Matlab 7.0 [31].

In all the designed QSAR equations, *N* is the number of data points, *R* is the correlation coefficient, *S* the model's standard deviation, *p* is the significance of the model. *FIT* is the Kubinyi function [32,33], a statistical parameter that closely relates to the Fisher ratio (*F*), but avoids the main disadvantage of the latter that is too sensitive to changes in small *d* (number of descriptors) values and poorly sensitive to changes in large *d* values. For those quinoxaline compounds that exhibited a censored data of the biological property, and that were experimentally reported as exceeding a certain limiting value (>value), an experience-based number was manually assigned with the aim of guiding the model's performance on these missing data. The various decision criteria that were simultaneously analyzed for determining the model's size of the five training sets investigated, such as, the determination of the optimal *d* to be included in each QSAR, are the following ones: (a) lowest value for the *S* parameter; (b) lowest *S*(*l*-*n*%-*o*) value; (c) highest *FIT* parameter; (d) lowest number of outlier quinoxalines exceeding 2*S*, 2.5*S* and 3*S*; (e) lowest value for the maximal inter-correlation between descriptors in the model (R_{ij}^{max}); and (f) favourable distribution of the percentual relative prediction errors.

We employed the orthogonalization procedure introduced several years ago by Randic [34] as a way of improving the statistical interpretation of the models built by interrelated

Table 1
Experimental and predicted biological data of quinoxaline-2-carboxylate 1,4-di-*N*-oxides according to derived QSAR equations (1)–(5).[a]

#	R′	R3	R6	R7	MIC[b] exp	MIC pred	IC_{50}[c] exp	IC_{50} pred	SI[d] exp (IC_{50}/MIC)	SI pred (IC_{50}/MIC)	EC_{90}[e] exp	EC_{90} pred	EC_{90}/MIC[f] exp	EC_{90}/MIC exp
1	CH_3	CH_3	H	H	1.56	0.86	>10.00	23.41	>6.41	10.11			Nonselective	
2	CH_3	CH_3	H	CH_3	1.56	1.48	>62.50	39.08	>40.06	26.28	1.94	1.30	1.24	0.72
3	CH_3	CH_3	H	OCH_3	1.56	1.64	52.58	38.85	33.71	14.18	ND[g]	0.20	ND	0.14
4	CH_3	CH_3	F	F	0.39	0.37	1.48	0.91	3.79	3.41			Nonselective	
5	CH_3	CH_3	Cl	Cl	0.20	0.28	1.68	2.45	8.40	9.56			Nonselective	
6	CH_3	CH_3	CH_3	CH_3	3.13*	5.78	>62.50	49.47	>20.0	18.55	1.64	3.12	0.52	0.47
7	CH_2CH_3	CH_3	H	H	1.56	1.73	47.00	43.90	30.13	28.51	0.88	0.87	0.56	1.16
8	CH_2CH_3	CH_3	H	Cl	0.20	0.81	>10.00	11.91	>50.00	13.39	0.46	0.43	2.30	2.08
9	CH_2CH_3	CH_3	H	CH_3	1.56	2.24	>62.50*	53.06	>40.06*	23.81	2.18	3.83	1.40	1.12
10	CH_2CH_3	CH_3	H	OCH_3	6.25	2.62	39.95	65.81	6.39	5.51			Nonselective	
11	CH_2CH_3	CH_3	F	F	1.56	1.15	1.20	1.41	0.77	3.68			Nonselective	
12	CH_2CH_3	CH_3	Cl	Cl	<0.20*	0.56	4.00	3.90	>20.00	10.48	0.63	0.87	>3.15	3.08
13	CH_2CH_3	CH_3	CH_3	CH_3	6.25	8.63	>62.50	63.21	>10.00	10.90	2.89	1.31	0.46	0.87
14	$C(CH_3)_3$	CH_3	H	H	>6.25	>12.5					Inactive			
15	$C(CH_3)_3$	CH_3	H/Cl	Cl	>6.25	>12.5					Inactive			
16	$C(CH_3)_3$	CH_3	Cl	Cl	>6.25	8.57					Inactive			
17	$C(CH_3)_3$	CH_3	CH_3	CH_3	>6.25	>12.5					Inactive			
18	$CH_2CH{=}CH_2$	CH_3	Cl	Cl	6.25	5.62	5.80	5.06	0.93	1.03			Nonselective	
19	$CH_2CH{=}CH_2$	CH_3	CH_3	CH_3	>6.25	10.21					Inactive			
20	$(CH_2)_2OCH_3$	CH_3	Cl	Cl	0.39	0.28	4.50	3.65	11.54	17.47	>1.56	1.47	>4.00	4.72
21	CH_2Ph	CH_3	H	H	0.10	0.12	47.00	28.77	470.00	486.67	0.15	0.20	1.50	1.05
22	CH_2Ph	CH_3	H	Cl	0.10	0.09	7.69	10.58	76.90	73.65	0.0005	0.0005	0.005	0.006
23	CH_2Ph	CH_3	H	CH_3	0.39*	0.67	56.05	47.44	143.72	265.19	1.13	1.03	2.90	1.63
24	CH_2Ph	CH_3	H	OCH_3	0.78	0.61	>62.50	57.28	>80.13	74.21	ND	0.05	ND	0.18
25	CH_2Ph	CH_3	F	F	0.05	0.11	1.00	1.23	20.00	8.93	ND	>100	ND	>32
26	CH_2Ph	CH_3	Cl	Cl	0.20	0.11	5.80	3.84	29.00	21.28	2.87	1.85	14.35	18.49
27	CH_2Ph	CH_3	CH_3	CH_3	3.13	5.03	ND	64.46	ND	57.37	ND	0.46	ND	1.66
28	CH_2CH_3	CH_2–COOEt	H	Cl	1.56	0.68	1.89	3.78	1.21	5.06			Nonselective	
29	CH_2CH_3	CH_2–COOEt	Cl	Cl	0.20	0.32	1.22	1.13	6.10	1.68			Nonselective	
30	CH_2CH_3	Ph	H	H	1.56	1.16	24.76	20.59	15.87	14.72	ND	0.11	ND	1.34
31	CH_2CH_3	Ph	H	Cl	1.56	0.92	3.33	4.16	2.13	5.46			Nonselective	
32	CH_2CH_3	Ph	H	CH_3	1.56	1.69	18.61	14.72	11.93	16.48	ND	0.060	ND	1.37
33	CH_2CH_3	Ph	H	OCH_3	6.25*	1.22	17.78	17.25	2.84	4.00			Nonselective	
34	CH_2CH_3	Ph	F	F	1.56	1.31	0.38	0.43	0.24	0.62			Nonselective	
35	CH_2CH_3	Ph	Cl	Cl	0.20	0.23	ND	1.23	ND	1.62	ND	0.0002	ND	0.10
36	CH_2CH_3	Ph	CH_3	CH_3	6.25	3.12	>10.00	13.43	>1.60	1.57			Nonselective	
37	CH_2CH_3	4′-F–Ph	H	H	0.39	1.24	8.90	6.20	22.82	14.53	ND	0.024	ND	0.89
38	CH_2CH_3	4′-Cl–Ph	H	H	0.39*	0.54	4.80*	3.97	12.31*	22.18	ND	1.46	ND	5.75
39	CH_2CH_3	4′-Br–Ph	H	H	0.39	0.46	6.12	2.91	15.69	14.26	ND	>100	ND	7.76
40	CH_2CH_3	4′-CH_3–Ph	H	H	0.78	0.89	11.29	15.30	14.47	23.02	ND	0.067	ND	1.08
41	CH_2CH_3	4′-CF_3–Ph	H	H	1.56	0.81	2.24	2.17	1.44	0.65			Nonselective	
42	CH_2CH_3	4′-OCH_3–Ph	H	H	1.56	1.71	7.65	12.55	4.90	5.42			Nonselective	
43	CH_2CH_3	4′-NO_2–Ph	H	H	0.78	0.81	2.26	2.30	2.90	1.97			Nonselective	

[a] For a general structure and definition of R′, R3, R6 and R7, see Fig. 2.
[b] Minimum inhibitory concentration against H37Rv strain of *Mycobacterium tuberculosis* (μg/mL).
[c] Measurement of cytotoxicity in VERO cells: 50% inhibitory concentrations (μg/mL).
[d] Selectivity index (*in vitro*): IC_{50} in VERO cells/MIC against *M. tuberculosis*.
[e] Measurement of efficacy inside mouse bone marrow macrophages: 90% efficacy concentration (μg/mL).
[f] Efficacy:activity ratio: EC_{90} in macrophages/MIC against *M. tuberculosis*.
[g] Not determined.
* Data used as test set.

indices. From our point of view, the co-linearity of the molecular descriptors should be as low as possible, because the interrelatedness among the different descriptors can lead to highly unstable regression coefficients, which makes it impossible to know the relative importance of an index and underestimates the utility of the regression coefficients of the model. The crucial step of the orthogonalization process is the choice of an appropriate order of orthogonalization, which in present analysis is the order that maximises the correlation between each orthogonal descriptor and the observed biological property. The *y*-randomization technique [35] consists in scrambling the experimental property in such a way that activities do not correspond to the respective compounds. After analyzing 5,000,000 cases of *y*-randomization for each developed QSAR, the smallest *S* value obtained using this procedure turned out to be a poorer value when compared to the one found when considering the true calibration. The theoretical validation practiced over each linear model is based on the leave-more-out cross-validation procedure (*l*-*n*%-*o*) [36], with *n*% representing the number of molecules removed from the training set. The number of cases for random data removal analyzed in every *l*-*n*%-*o* was 5,000,000, while *n*% = 10% was considered.

3. Results and discussion

For modeling purposes, the biological data of the 43 quinoxaline derivatives listed in Table 1 were converted into logarithm units. We searched for the best linear regression models on the five antimycobacterial parameters by resorting to the Replacement Method variable subset selection approach. Thus, the most relevant and representative structural descriptors were selected by using this efficient optimization tool on each training set of quinoxalines, exploring a pool containing more than a thousand numerical variables computed by means of the program DRAGON. All the predictive linear QSAR models reported below are able to capture the essential structural features of the quinoxaline

Table 2
Notation for molecular descriptors involved in QSAR models for biological data of quinoxaline 1,4-di-*N*-oxide derivatives.

Type	Molecular descriptor	Description
BCUT	*BELm6*	Lowest eigen value no. 6 of Burden matrix/weighted by atomic masses
2D-Autocorrelations	*MATS7p*	Moran autocorrelation of lag 7/weighted by atomic polarizabilities
	MATS2m	Moran autocorrelation of lag 2/weighted by atomic masses
	GATS3m	Geary autocorrelation of lag 3/weighted by atomic masses
	GATS7v	Geary autocorrelation of lag 7/weighted by atomic volumes
GETAWAY	$R2u^+$	R maximal autocorrelation of lag 2/unweighted
	R3p	R autocorrelation of lag 3/weighted by atomic polarizabilities
	$R7p^+$	R maximal autocorrelation of lag 7/weighted by atomic polarizabilities
	RTe^+	R maximal index/weighted by atomic Sanderson electronegativities
Galvez topological charge indices	*GGI8*	Topological charge index of order 8
3D-MoRSE	*Mor02e*	3D-MoRSE-signal 02/weighted by atomic Sanderson electronegativities
	Mor19u	3D-MoRSE-signal 19/unweighted
Topological	*X5A*	Average connectivity index chi-5
	IVDE	Mean information content vertex degree equality
	TIE	E-state topological parameter
	T(F..F)	Sum of topological distances between F..F
	S3K	3-path Kier alpha-modified shape index
Geometrical	*DISPe*	*d* COMMA2 value/weighted by atomic Sanderson electronegativities
WHIM	*G2v*	2nd component symmetry directional WHIM index/weighted by atomic van der Waals volumes

compounds that relate to their biological effects. A brief description for each molecular descriptor, that was orthogonalized when included in each model, is supplied by Table 2.

Minimum inhibitory concentration (MIC) against M. tuberculosis H37Rv:

$$\log_{10}(\mathrm{MIC}) = -0.144(\pm0.04) - 3.492(\pm0.3)\cdot\mathbf{\Omega}_{MATS7p} + 8.465(\pm1)\cdot\mathbf{\Omega}_{R2u^+} + 3.109(\pm0.5)\cdot\mathbf{\Omega}_{BELm6} - 1.987(\pm0.4)\cdot\mathbf{\Omega}_{R3p} + 67.114(\pm21)\cdot\mathbf{\Omega}_{R7p^+} \quad (1)$$

$N = 38$, $R = 0.946$, $S = 0.229$, $FIT = 4.289$, $p < 10^{-5}$, $R_{loo} = 0.928$, $S_{loo} = 0.263$, $R_{l\text{-}10\%\text{-}o} = 0.909$, $S_{l\text{-}10\%\text{-}o} = 0.299$, range in MIC: 0.05 to >6.25 μg/mL, outliers(>2*S*): **8**, **37**.

50% inhibitory concentration (IC_{50}) in VERO cells:

$$\log_{10}(\mathrm{IC}_{50}) = 43.160(\pm3) - 43.093(\pm3)\cdot\mathbf{\Omega}_{MATS2m} - 3.340\times(\pm0.3)\cdot\mathbf{\Omega}_{GGI8} - 0.285(\pm0.03)\cdot\mathbf{\Omega}_{T(F..F)} + 0.070(\pm0.009)\cdot\mathbf{\Omega}_{Mor02e} \quad (2)$$

$N = 34$, $R = 0.969$, $S = 0.165$, $FIT = 8.992$, $p < 10^{-5}$, $R_{loo} = 0.960$, $S_{loo} = 0.189$, $R_{l\text{-}10\%\text{-}o} = 0.942$, $S_{l\text{-}10\%\text{-}o} = 0.234$, range in IC_{50}: 0.38 to >62.5 μg/mL, outliers(>2*S*): **1**.

Selectivity index (calculated as IC_{50}/MIC):

$$\log_{10}\left(\frac{\mathrm{IC}_{50}}{\mathrm{MIC}}\right) = -13.910(\pm2) + 221.107(\pm23)\cdot\mathbf{\Omega}_{X5A} - 0.915(\pm0.2)\cdot\mathbf{\Omega}_{DISPe} - 8.298(\pm2)\cdot\mathbf{\Omega}_{RTe^+} - 4.259(\pm1)\cdot\mathbf{\Omega}_{IVDE} - 1.373(\pm0.4)\cdot\mathbf{\Omega}_{Mor19u} \quad (3)$$

$N = 34$, $R = 0.916$, $S = 0.308$, $FIT = 2.461$, $p < 10^{-5}$, $R_{loo} = 0.871$, $S_{loo} = 0.379$, $R_{l\text{-}10\%\text{-}o} = 0.794$, $S_{l\text{-}10\%\text{-}o} = 0.489$, range in SI: 0.24–470, outliers(>2*S*): **11**, **28**.

90% effective concentration in a TB-infected macrophage model:

$$\log_{10}(\mathrm{EC}_{90}) = 2.220(\pm0.2) - 0.007(\pm0.0005)\cdot\mathbf{\Omega}_{TIE} + 650.099(\pm103)\cdot\mathbf{\Omega}_{GATS3m} - 2.923(\pm0.9)\cdot\mathbf{\Omega}_{GATS7v} \quad (4)$$

$N = 12$, $R = 0.985$, $S = 0.212$, $FIT = 12.105$, $p < 10^{-5}$, $R_{loo} = 0.964$, $S_{loo} = 0.328$, range in EC_{90}: 0.0005–2.89 μg/mL, outliers(>2*S*): –.

Effectiveness:activity ratio:

$$\log_{10}\left(\frac{\mathrm{EC}_{90}}{\mathrm{MIC}}\right) = -15.192(\pm2) - 0.009(\pm0.0007)\cdot\mathbf{\Omega}_{TIE} + 1.908(\pm0.2)\cdot\mathbf{\Omega}_{S3K} + 79.972(\pm10)\cdot\mathbf{\Omega}_{G2v} \quad (5)$$

$N = 12$, $R = 0.977$, $S = 0.210$, $FIT = 8.155$, $p < 10^{-5}$, $R_{loo} = 0.887$, $S_{loo} = 0.493$, range in EC_{90}/MIC: 0.005–14.35, outliers(>2*S*): –.

Here, Ω is the symbol for the orthogonal descriptor, *N* is the number of molecules in the training set, *R* is the correlation coefficient, *S* is the model's standard deviation, *FIT* is the Kubinyi function, *p* is the significance of the model, outliers(>2*S*) denotes the number of molecules having a residual (*res*) that exceeds two standards deviations, and *loo* and *l-10%-o* stand for the leave-one-out and leave-10%-out cross validation techniques, respectively. None of these good quality derived QSARs incorporate redundant structural information, as they include orthogonalized molecular descriptors that avoid collineality among the different numerical variables that may yield to highly unstable models.

As a next step, with the main purpose of verifying that the linear QSARs established were not only correlative but would also function similarly well for the prediction of new biological data not contemplated during the training stage of the models, it was mandatory to perform their validation in order to analyze their predictive performance [36]. Eqs. (1)–(5) satisfactorily passed the internal validation process, as evidenced by the stability of the models upon the inclusion/exclusion of compounds from the training set, measured via the commonly employed *loo* parameter (R_{loo}) together with the more severe *l-n%-o* parameter ($R_{l\text{-}10\%\text{-}o}$). These results are in the range of a validated model: $R_{l\text{-}n\%\text{-}o}$ must be greater than 0.50, according to the specialized literature [37]. Alternatively, for the case of models 1–3 that involved more data points (*N*) we performed external validations, by estimating the biological data for some fresh quinoxaline derivatives not considered during the model development. The predictions achieved on these test set compounds (denoted with asterisk in Table 1) are able to provide an insight on the variation of the experimental property for such structurally related "unknown structures", demonstrating the correct functioning of the three structure–activity relationships. Finally, as a further step to assess

the robustness of present equations, we applied the *y*-randomization procedure [35] for each of them, demonstrating that the calibrations did not result from happenstance and therefore, corresponded to valid structure–activity relationships.

Topological descriptors appearing in Eqs. (1)–(5), such as *X5A*, *IVDE*, *TIE*, *T(F..F)*, *S3K* and *GGI8*, characterize the "topological shape" describing different structural elements such as branching degree, size, symmetry, flexibility, cyclicity, centricity, etc. [38] The Topological Distance matrix (**D**), introduced by Harary in the 1960s, accounts for the "through bond" interactions of atoms in molecules; most of these descriptors characterize the distribution of the topological distances in each chemical graph. The structural variables introduced by Broto, Moreau and Geary [39] correspond to bi-dimensional autocorrelations between pairs of atoms in the molecule, and were defined in order to reflect the contribution of a considered atomic property to the dependent property under investigation. Different weighting schemes can be adopted to differentiate the nature of atoms: mass (m), polarizability (p), electronegativity (e) or volume (v). 2D-Autocorrelations (*MATS7p*, *MATS2m*, *GATS3m* and *GATS7v*) can be readily calculated, i.e.: by summing products of the atomic weights of the terminal atoms of all the paths of a prescribed length. BCUT descriptors (*BELm6*) are eigen values of the Burden matrix (**B**) [40]. As it is known, **B** corresponds to a modified Adjacency matrix (**A**) with the main diagonal elements being weighted with a certain atomic property.

Among the more elaborated descriptors, geometrical variables (*DISPe*) are 3D-descriptors obtained from moment expansions that do not require molecular superposition or alignment for the assignment of molecular similarity, incorporating information about the magnitude of the displacement between the molecular centroid (center of mass) and the polarizability-field center (center of charge) [41]. This kind of parameters were found valuable for the prediction of the electrophoretic mobilities of peptides. GETAWAY (GEometry, Topology, and Atom-Weights AssemblY) descriptors [42], such as $R2u^+$, $R3p$, $R7p^+$ and RTe^+, were specially designed with the aim of matching the 3D-molecular geometry. These are derived from the elements h_{ij} of the Molecular Influence matrix (**H**), obtained through the values of atomic Cartesian coordinates. The diagonal elements of **H** (h_{ii}) are called leverages, and are considered to represent the influence of each molecule atom in determining the whole shape of the molecule. For instance, the mantle atoms always have higher h_{ii} values than atoms near the molecule center, while each off-diagonal element h_{ij} represents the degree of accessibility of the jth atom to interactions with the ith atom. The influence/distance matrix (**R**) involves a combination of the elements of **H** matrix with those of the geometric matrix (**G**).

The 3D-MoRSE descriptors (*Mor02e*, *Mor19u*) are obtained through the molecular transform generally employed in electron diffraction studies [43]. The electron diffraction does not directly yield atomic coordinates, but provides diffraction patterns from which the atomic coordinates are derived by mathematical transformations. These codes are defined in order to reflect the contribution to the property under investigation, at a prescribed scattering angle, of a given atomic property. The weighted holistic invariant molecular (WHIM) descriptors [44] (*G2v*) are obtained as statistical indices of the atoms projected onto three principal axes (principal components) obtained from weighted covariance matrices of the atomic coordinates. The weights employed are again the atomic properties that enable differentiation among atoms. The aim of defining such indices is to capture 3D-information regarding size, shape, symmetry and atom distributions with respect to invariant reference frames.

It is worthwhile to notice that all the QSARs derived involve a combination of 2D- and 3D-type of molecular descriptors. The sequential order in which these variables appear in each model agrees with the order of relative contribution importance (in modulus), as deduced from a subsequent standardization of the orthogonalized regression coefficients [45]. The most important descriptor from each QSAR results for all cases to be of topological nature, with further 3D-variables being also incorporated into the equations so that improving the statistical fit. This result points to the fact that most quinoxaline derivatives include a nearly planar structure, and therefore, conformation dependent type of descriptors play a secondary role. The connectivity of atoms and their

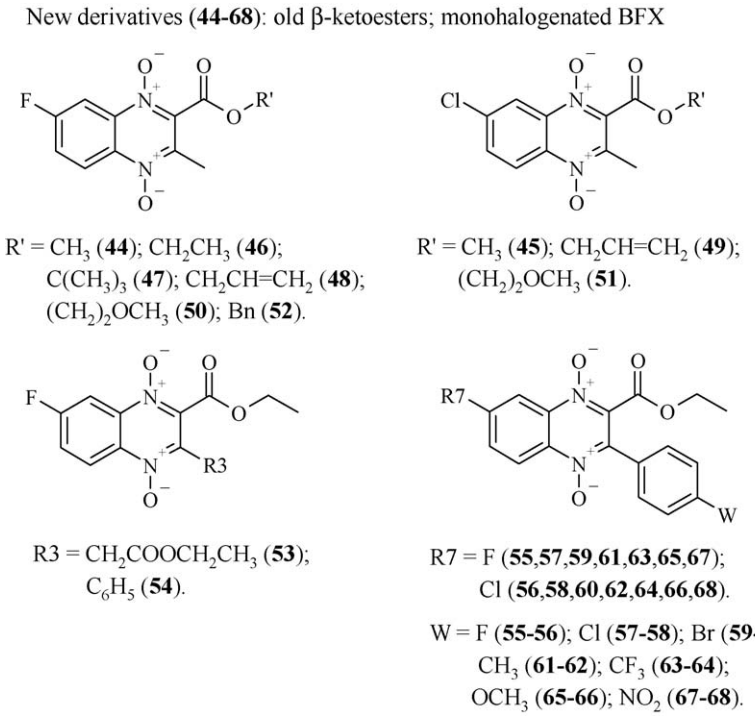

Fig. 3. Structure of new proposed compounds **44**–**68**.

identity pose the greatest influence on describing the dependent variables.

As can be observed in Table 1, models 1–5 we present here have the overall tendency of acceptably estimating the behavior of the observed biological data in each training sets. It is important to outline that these QSARs were searched in such a way they minimize the parameter *S*, and besides of this they were subjected to the condition of following the variation of the data in the best possible manner. For the case of the minimum inhibitory concentration (MIC) values, Eq. (1) involving 5 variables and 43 data points accurately predicts very potent quinoxaline inhibitors (MIC < 1 μg/mL) and the same applies for those chemicals displaying the lowest potencies $1 <$ MIC < 6.25 μg/mL. To further test the QSAR models, it is important to examine the data outliers. An outlier is normally identified by having a large residual, and there are several reasons for their occurrence in QSAR studies. There are only two compounds with residuals (in log units) exceeding two standard deviations: **8** (res = −0.605 μg/mL) and **37**

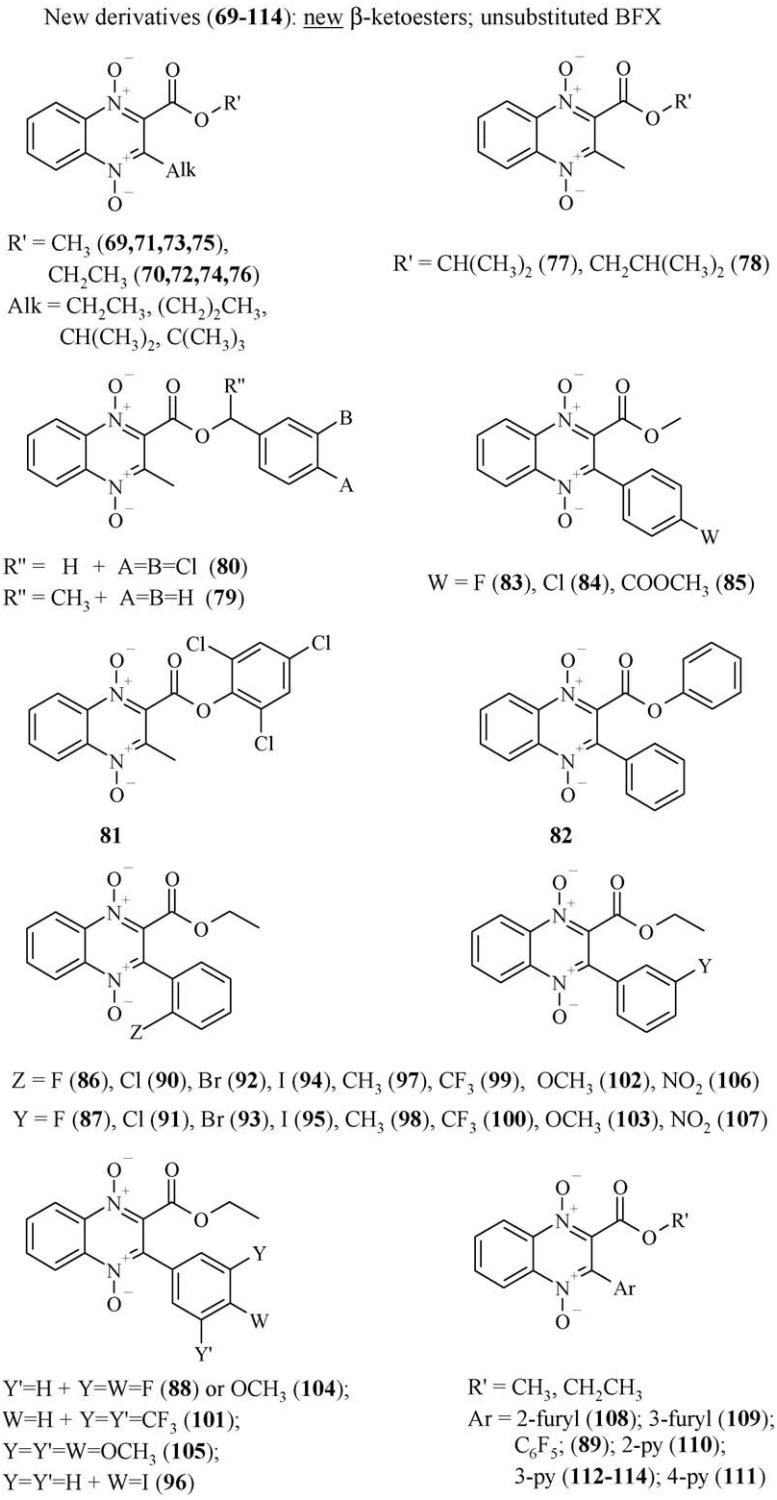

Fig. 4. Structure of new proposed compounds **69–114**.

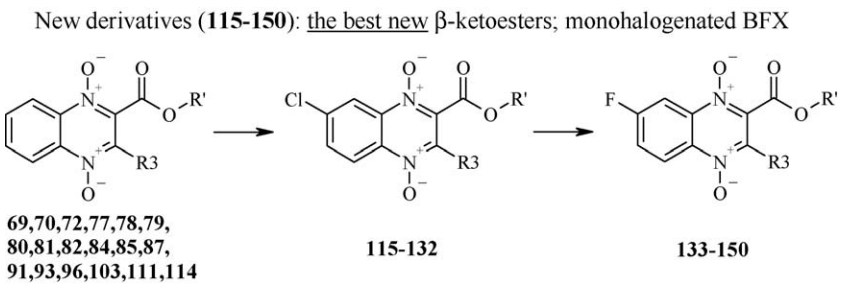

Fig. 5. Structure of new proposed compounds **115–150**.

(res = −0.502 μg/mL), although none of these exceed the value 3*S*. This fact may be attributed to the experimental error or to the selected subset of descriptors in the model, while the error resulting from the geometry optimization of the compounds can be discarded. A similar situation occurs for the estimation of toxicity, which is expressed through the 50% inhibitory concentration in VERO cells (IC_{50}). The 4-parameters QSAR model given by Eq. (2) is able to accurately distinguish the toxic effect of 34 quinoxaline derivatives, showing only a single outlier: **1** (*res* = −0.369 μg/mL) that does not exceed the 2.5*S* limit and whose relative prediction error is low (37%).

Regarding the selectivity index, which is denoted as the quotient IC_{50}/MIC, a model that is based on 5 molecular descriptors and 34 molecules is able to correctly classify those compounds that experimentally pass this test. Only two outlier molecules are present that do not exceed 3*S*: **11** (res = −0.680) and **28** (res = −0.625). Furthermore, the QSARs of Eqs. (4) and (5) were developed with only 12 molecules each one, due to the unavailability of additional experimental measurements for EC_{90} and EC_{90}/MIC, respectively. In fact, there is one data point (compound **22**) whose property value is apart from the rest ones since no experiments were conducted until now in this interval, but we decided to provide with this information to the models. Obviously, these QSARs have a narrower structural applicability domain due to the scarceness of experimental observations, although they were also cross-validated like the previous equations. Despite this inconvenience, we thought it useful to report both models as complementary decision tools to the previously developed QSARs. Eqs. (4) and (5) include three structural parameters each one and do not exhibit outlier compounds exceeding 2*S*.

As part of the QSAR models application, Eqs. (4) and (5) allow predicting the effectiveness of compounds **30–43**, and other compounds without this data (**3**, **24**, **25** and **27**). All of the predicted results are good values, so these compounds would be further evaluated, with the exception of compound **25** (the only one difluorinated derivative which showed selectiveness). The predicted effectiveness of compound **39** is not clear because of the different results of models 4 and 5; this could be due to the presence of a bromine atom, not included in any other compound of the training set.

In view of the results achieved and as a further attempt to improve the biological profile of these quinoxaline di-*N*-oxide derivatives, we applied the five QSARs developed to estimate the antimycobacterial parameters of new possible antitubercular candidates that still do not present experimentally assigned biological data. The molecular structures of Figs. 2–4 were proposed on the basis that these chemicals can be synthesized from inexpensive and commercially available reagents. Non-substituted quinoxalines at positions R6/R7, together with 7-chlorinated derivatives, are considered to be the most interesting compounds from the chemical and biological points of view [10–14]. Compounds **44–68** resulted from the condensation of previous β-ketoesters (maintaining groups in R2 and R3 positions) with the corresponding monosubstituted benzofuroxans, giving rise to 7-(fluoro or chloro) quinoxaline 1,4-di-*N*-oxide derivatives (Fig. 3). With these predictions, we could determine the influence of the halogen atom at R7 position. On the contrary, compounds **69–114** were derived from the reaction of some new available commercially β-ketoesters with the non-substituted benzofuroxan skeleton (Fig. 4). This process leads to simpler compounds that do not present isomerism problems. Among the previous derivatives **69–114**, the structures with the best results were chosen and new compounds were designed, **115–150**. These analogues keep the R2/R3 positions from the corresponding β-ketoester but show a halogen atom (fluorine or chlorine) in R7 position (Fig. 5).

Table 3 summarizes the best resulting anti-tubercular structures as predicted by Eqs. (1)–(5). Several interesting results can be observed from this table, as many of the proposed compounds lead to favourable predictions for the five biological parameters. Among the first series of predicted compounds, **44–68**, the only two that unmistakably showed good values for the five models are 7-chlorinated derivatives (**45**, **51**). Regarding the substituents in the carboxylate group, *tert*-butyl (**47**) and allyl (**48**, **49**) derivatives were predicted as inactive and the benzyl derivative (**52**) was

Table 3
Predicted biological data for 12 chosen derivatives to be synthesized.[a].

#	R′	R3	R6	R7	MIC pred	IC_{50} pred	SI pred (IC_{50}/MIC)	EC_{90} pred	EC_{90}/MIC pred
45	CH_3	CH_3	H	Cl	0.62	7.46	11.11	3.13	0.07
51	$(CH_2)_2OCH_3$	CH_3	H	Cl	0.31	10.29	25.11	<0.0001	<0.001
69	CH_3	CH_2CH_3	H	H	0.91	25.95	18.48	0.45	1.32
70	CH_2CH_3	CH_2CH_3	H	H	2.58	48.26	18.74	0.27	1.57
79	$CH(CH_3)$(Ph)	CH_3	H	H	0.05	25.14	53.27	1.87	5.57
82	Ph	Ph	H	H	0.09	60.13	>500	0.87	8.98
87	CH_2CH_3	3-F–Ph	H	H	0.7	8.31	9.89	0.0015	0.09
91	CH_2CH_3	3-Cl–Ph	H	H	0.56	5.42	18.43	1.07	5.21
103	CH_2CH_3	3-OCH_3–Ph	H	H	1.12	12.34	9.87	0.15	3.42
111	CH_2CH_3	4-py	H	H	1.45	15.47	27.82	0.23	2.69
115	CH_3	CH_2CH_3	H	Cl	0.5	8.23	14.38	0.05	0.12
120	$CH(CH_3)$(Ph)	CH_3	H	Cl	0.06	8.38	15.94	0.01	0.21

[a] For a general structure and definition of R′, R3, R6 and R7, see Fig. 2.

predicted as the most active (the lowest MIC); this fact coincided with the previous experimental data. In addition, all the ethyl 3-(4′-substituted)phenylquinoxaline-2-carboxylate 1,4-di-*N*-oxide derivatives were predicted as active but nonselective; doubt remains regarding compound **56** which SI predicted to be higher than 10 but the IC_{50} in VERO cells is too low. The rest of the predicted compounds were prepared with new β-ketoesters, giving rise to new substituents in the carboxylate group and in R3 position of the quinoxaline scaffold. The introduction of bulky groups, such as *iso*-propyl and *tert*-butyl, in R3 position become inactive derivatives (**73**–**76**). Among the compounds that contain an aromatic ring linked to R3 position, only the 3′-(fluoro or chloro or methoxy) phenyl (**87**, **91**, **103**) and the 4′-pyridyl (**111**) derivatives were predicted as active, selective and effective, proving the importance of the relative position and the nature of the substituent. The introduction of a benzene ring as substituent of the carboxylate group (**79**–**82**, **120**–**123**, **138**–**141**) gave rise to the most interesting structures. Furthermore, the 7-fluorinated derivatives (**133**–**150**) were predicted, in general, as the least selective and the 7-chlorinated derivatives (**115**–**132**) did not improve the results to make their synthesis worthwhile due to problems of isomer separation.

A total of 107 different structures (**44**–**150**) were predicted as antimycobacterial candidates using 5 models for each evaluated parameter, but only 12 (**45**, **51**, **69**, **70**, **79**, **82**, **87**, **91**, **103**, **111**, **115**, **120**) of them showed results that were good enough to justify their synthesis (Table 3). This fact means that, based on the predictions, 90% of the synthesis and biological evaluation can be avoided. Table 4 includes the numerical values for the theoretical descriptors appearing in all the developed QSPR.

Table 4
Numerical values for structural descriptors of quinoxaline-2-carboxylate 1,4-di-*N*-oxides appearing in QSAR equations (1)–(5).

#	*MATS7p*	*R2u*$^+$	*BELm6*	*R3p*	*R7p*$^+$	*MATS2m*	*GGI8*	*T(F..F)*	*Mor02e*	*X5A*	*DISPe*	*RTe*$^+$	*IVDE*	*Mor19u*	*TIE*	*GATS3m*	*GATS7v*	*S3K*	*G2v*
1	0.174	0.202	0.882	0.419	0.012	0.979	0.020	0	23.962	0.070	0.164	0.180	1.873	0.636	228.810	0.003	0.861	1.685	0.174
2	0.010	0.157	0.982	0.432	0.009	0.979	0.020	0	27.143	0.068	0.254	0.140	1.816	0.683	235.940	0.003	0.987	1.901	0.169
3	0.119	0.157	1.048	0.377	0.009	0.980	0.045	0	28.727	0.067	0.101	0.140	1.873	0.717	346.680	0.003	0.856	2.091	0.168
4	0.166	0.201	0.778	0.403	0.012	0.992	0.045	3	24.065	0.066	0.600	0.179	1.720	0.277	63.495	0.002	0.862	1.960	0.174
5	0.116	0.201	0.748	0.567	0.016	0.999	0.045	0	22.549	0.066	0.362	0.179	1.720	0.237	267.970	0.004	1.009	2.179	0.195
6	−0.150	0.127	1.101	0.447	0.008	0.979	0.045	0	29.651	0.066	0.348	0.116	1.720	1.013	241.770	0.004	1.162	2.000	0.165
7	−0.027	0.140	0.974	0.474	0.013	0.976	0.033	0	26.677	0.069	0.107	0.124	1.925	0.908	235.460	0.003	1.049	2.008	0.169
8	0.004	0.140	0.954	0.529	0.013	0.987	0.057	0	25.933	0.067	0.317	0.124	1.873	0.721	443.770	0.005	1.036	2.312	0.188
9	−0.119	0.120	1.010	0.472	0.011	0.977	0.057	0	29.432	0.067	0.133	0.112	1.873	0.945	245.690	0.004	1.117	2.215	0.165
10	−0.024	0.127	1.051	0.416	0.011	0.977	0.057	0	30.769	0.066	0.098	0.124	1.904	1.001	476.320	0.003	1.013	2.431	0.164
11	−0.026	0.139	0.924	0.459	0.013	0.989	0.082	3	26.588	0.065	0.762	0.124	1.782	0.542	170.110	0.003	1.063	2.284	0.169
12	−0.068	0.139	0.913	0.615	0.013	0.996	0.082	0	25.245	0.065	0.523	0.124	1.782	0.499	297.500	0.004	1.146	2.518	0.169
13	−0.227	0.099	1.142	0.479	0.010	0.977	0.082	0	31.563	0.065	0.229	0.096	1.782	1.271	255.420	0.004	1.241	2.327	0.162
14	−0.211	0.110	1.263	0.656	0.013	0.973	0.091	0	31.684	0.066	0.223	0.140	1.846	1.089	254.870	0.006	1.231	2.707	0.177
15	−0.193	0.110	1.253	0.715	0.016	0.983	0.165	0	30.627	0.065	0.437	0.140	1.776	0.945	472.920	0.007	1.210	3.038	0.162
16	−0.239	0.110	1.216	0.804	0.016	0.992	0.165	0	29.619	0.064	0.647	0.139	1.677	0.777	304.840	0.006	1.272	3.242	0.162
17	−0.310	0.102	1.266	0.651	0.009	0.973	0.165	0	34.673	0.064	0.153	0.132	1.677	1.449	267.870	0.006	1.334	3.027	0.156
18	−0.041	0.186	1.043	0.615	0.014	0.991	0.103	0	24.850	0.067	0.548	0.166	1.874	0.604	345.360	0.004	1.078	2.768	0.168
19	−0.175	0.082	1.183	0.581	0.018	0.973	0.103	0	30.198	0.067	0.252	0.122	1.874	1.468	338.080	0.004	1.179	2.567	0.161
20	−0.074	0.093	1.047	0.751	0.013	0.992	0.140	0	24.950	0.067	0.499	0.124	1.839	0.524	383.940	0.005	1.074	3.227	0.164
21	0.108	0.081	1.105	0.726	0.014	0.969	0.206	0	27.239	0.074	0.269	0.120	1.826	1.024	345.740	0.003	0.861	2.832	0.161
22	0.091	0.082	1.091	0.779	0.014	0.979	0.206	0	26.799	0.072	0.491	0.123	1.865	0.851	671.300	0.004	0.862	3.167	0.161
23	0.015	0.078	1.213	0.712	0.014	0.970	0.206	0	30.921	0.072	0.123	0.109	1.865	1.104	347.170	0.004	0.936	3.065	0.158
24	0.097	0.083	1.228	0.666	0.014	0.971	0.206	0	32.667	0.071	0.246	0.108	1.903	1.088	490.910	0.004	0.852	3.269	0.157
25	0.111	0.080	1.091	0.711	0.014	0.981	0.230	3	27.310	0.070	0.917	0.116	1.874	0.750	85.926	0.003	0.868	3.108	0.175
26	−0.007	0.081	1.091	0.863	0.014	0.987	0.230	0	26.130	0.070	0.690	0.120	1.874	0.687	317.610	0.004	0.959	3.355	0.161
27	−0.093	0.069	1.375	0.705	0.013	0.970	0.230	0	33.826	0.070	0.114	0.100	1.874	1.374	347.000	0.004	1.056	3.153	0.156
28	−0.142	0.075	1.062	0.668	0.014	0.985	0.205	0	23.836	0.067	0.416	0.087	1.896	1.407	471.870	0.005	1.114	3.774	0.173
29	−0.124	0.074	1.056	0.756	0.014	0.993	0.254	0	23.010	0.065	0.610	0.088	1.829	1.381	497.660	0.004	1.113	3.954	0.173
30	−0.122	0.070	1.091	0.631	0.016	0.969	0.098	0	20.653	0.067	0.183	0.132	1.826	0.876	283.230	0.003	1.174	2.598	0.161
31	−0.133	0.070	1.091	0.682	0.016	0.979	0.172	0	19.593	0.067	0.398	0.132	1.865	0.809	2589.800	0.004	1.176	2.917	0.161
32	−0.167	0.070	1.091	0.626	0.016	0.970	0.172	0	22.237	0.067	0.106	0.125	1.865	0.919	303.230	0.003	1.196	2.821	0.158
33	−0.095	0.068	1.112	0.587	0.015	0.971	0.172	0	23.798	0.066	0.224	0.124	1.903	0.954	418.530	0.003	1.129	3.019	0.170
34	−0.126	0.070	1.091	0.616	0.016	0.981	0.217	3	20.250	0.065	0.812	0.131	1.874	0.650	93.911	0.002	1.178	2.875	0.161
35	−0.026	0.069	1.091	0.766	0.016	0.987	0.217	0	18.503	0.065	0.592	0.130	1.874	0.628	644.800	0.004	1.101	3.107	0.175
36	−0.124	0.065	1.221	0.623	0.015	0.970	0.217	0	23.548	0.065	0.180	0.124	1.874	1.226	337.980	0.004	1.150	2.917	0.156
37	−0.123	0.070	1.091	0.621	0.016	0.975	0.192	0	20.589	0.067	0.167	0.131	1.865	0.793	346.860	0.003	1.171	2.800	0.161
38	−0.079	0.070	1.091	0.702	0.016	0.979	0.192	0	20.132	0.067	0.082	0.130	1.865	0.777	280.620	0.004	1.131	2.917	0.161
39	−0.051	0.070	1.091	0.743	0.018	0.982	0.192	0	19.928	0.067	0.100	0.130	1.865	0.895	279.860	0.011	1.095	2.982	0.161
40	−0.114	0.070	1.122	0.660	0.014	0.970	0.192	0	23.312	0.067	0.264	0.101	1.865	0.992	314.880	0.003	1.136	2.821	0.158
41	−0.115	0.086	1.091	0.643	0.012	0.978	0.485	0	28.046	0.066	0.776	0.115	1.885	1.060	316.960	0.002	1.156	3.407	0.158
42	−0.137	0.069	1.127	0.611	0.015	0.971	0.254	0	25.251	0.067	0.159	0.120	1.903	1.165	289.990	0.003	1.162	3.019	0.157
43	−0.126	0.069	1.091	0.640	0.014	0.978	0.365	0	23.407	0.066	0.351	0.144	1.870	0.820	701.320	0.002	1.181	3.114	0.159
45	0.216	0.131	0.853	0.572	0.018	0.990	0.020	0	22.213	0.068	0.192	0.197	1.816	0.309	413.270	0.005	0.853	1.991	0.174
51	−0.034	0.092	1.069	0.611	0.017	0.975	0.115	0	32.437	0.069	0.190	0.227	1.917	0.765	295.820	0.006	0.988	3.144	0.162
69	−0.190	0.111	0.911	0.498	0.014	0.976	0.020	0	23.926	0.071	0.131	0.158	1.925	0.848	206.030	0.003	1.256	1.901	0.169
70	−0.196	0.102	0.975	0.586	0.020	0.974	0.033	0	25.696	0.070	0.192	0.132	1.950	1.306	229.160	0.003	1.248	2.215	0.165
79	0.029	0.109	1.106	0.696	0.012	0.968	0.227	0	26.803	0.072	0.301	0.145	1.784	1.416	322.900	0.004	0.938	2.939	0.158
82	0.027	0.054	1.091	0.719	0.012	0.966	0.140	0	30.272	0.070	0.199	0.113	1.576	0.432	260.580	0.003	0.956	3.093	0.157
87	−0.120	0.070	1.034	0.618	0.016	0.975	0.143	0	20.541	0.066	0.174	0.171	1.865	0.704	458.580	0.003	1.172	2.800	0.161
91	−0.129	0.069	1.026	0.708	0.020	0.979	0.143	0	20.175	0.066	0.164	0.123	1.865	0.751	285.390	0.004	1.160	2.917	0.161
103	−0.086	0.084	1.132	0.588	0.016	0.971	0.164	0	26.174	0.067	0.191	0.131	1.903	1.158	291.840	0.003	1.099	3.019	0.170
111	−0.113	0.072	1.091	0.578	0.016	0.972	0.098	0	26.917	0.067	0.060	0.141	1.826	0.659	253.690	0.003	1.180	2.577	0.162
115	−0.150	0.105	0.855	0.612	0.017	0.987	0.020	0	22.092	0.069	0.297	0.144	1.873	0.789	386.520	0.004	1.240	2.196	0.169
120	0.020	0.067	1.091	0.775	0.017	0.977	0.251	0	32.824	0.070	0.458	0.118	1.817	1.060	606.040	0.005	0.936	3.254	0.158

4. Conclusions

The representation of the molecular structure of quinoxaline derivatives through more than a thousand theoretical structural descriptors, encoding its various constitutional, topological, geometrical and electronical aspects, allowed formulation of predictive linear QSAR models on five different available biological endpoints. A practical application of these QSARs enabled the proposal of 12 new quinoxaline derivatives that still do not have experimentally assigned antimycobacterial data. These predicted compounds involve quite similar structures to known quinoxalines and are quite easy to synthetize from economically accessible reactants, and many of them exhibit much more favourable estimated activities that make these compounds suitable candidates for synthesis.

Acknowledgements

Antimycobacterial data was provided by the Tuberculosis Antimicrobial Acquisition & Coordinating Facility (TAACF) through a research and development contract with the U.S. National Institute of Allergy and Infectious Diseases (NIAID). E. Vicente is grateful to the Ministerio de Educación y Ciencia (Spain) for the grant AP2003-2175. PRD is a researcher from the National Council of Scientific and Technological Research (CONICET).

References

[1] WHO Fact Sheet No. 104, 2007.
[2] WHO, Weekly Epidemiological Record No. 15, 2003, 78, 121.
[3] J.C. Palomino, S.C. Leao, V. Ritacco, Tuberculosis 2007: From Basic Science to Patient Care, http://www.tuberculosistextbook.com, 2007.
[4] L.M. Lima, E.J. Barreiro, Curr. Med. Chem. 12 (2005) 23.
[5] M.E. Montoya, Y. Sainz, M.A. Ortega, A.L. De Cerain, A. Monge, FARMACO 53 (1998) 570.
[6] Y. Sainz, M.E. Montoya, F.J. Martinez-Crespo, M.A. Ortega, A.L. De Cerain, A. Monge, Arzneimittel-Forsch. 49 (1999) 55.
[7] M.A. Ortega, Y. Sainz, M.E. Montoya, A.L. De Cerain, A. Monge, Pharmazie 54 (1999) 24.
[8] M.A. Ortega, M.E. Montoya, A. Jaso, B. Zarranz, I. Tirapu, I. Aldana, A. Monge, Pharmazie 56 (2001) 205.
[9] M.A. Ortega, Y. Sainz, M.E. Montoya, A. Jaso, B. Zarranz, I. Aldana, A. Monge, Arzneimittel-Forsch. 52 (2002) 113.
[10] B. Zarranz, A. Jaso, I. Aldana, A. Monge, Bioorg. Med. Chem. 11 (2003) 2149.
[11] A. Jaso, B. Zarranz, I. Aldana, A. Monge, Eur. J. Med. Chem. 38 (2003) 791.
[12] A. Jaso, B. Zarranz, I. Aldana, A. Monge, J. Med. Chem. 48 (2005) 2019.
[13] R. Villar, E. Vicente, B. Solano, S. Pérez-Silanes, I. Aldana, J.A. Maddry, A.J. Lenaerts, S.G. Franzblau, S.-H. Cho, A. Monge, R.C. Goldman, J. Antimicrob. Chemother. 62 (2008) 547.
[14] E. Vicente, R. Villar, A. Burguete, B. Solano, S. Pérez-Silanes, I. Aldana, J.A. Maddry, A.J. Lenaerts, S.G. Franzblau, S.-H. Cho, A. Monge, R.C. Goldman, Antimicrob. Agents Chemother. 52 (2008) 3321.
[15] E. Vicente, S. Pérez-Silanes, L.M. Lima, S. Ancizu, A. Burguete, B. Solano, R. Villar, I. Aldana, A. Monge, Bioorg. Med. Chem. 17 (2009) 385.
[16] C. Hansch, A. Leo, Exploring QSAR. Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington, DC, 1995.
[17] R.P. Verma, C. Hansch, Chembiochem 5 (2004) 1188.
[18] A.R. Katritzky, V.S. Lobanov, M. Karelson, Chem. Soc. Rev. 24 (1995) 279.
[19] A. Carta, G. Paglietti, M.E.R. Nikookar, P. Sanna, L. Sechi, S. Zanetti, Eur. J. Med. Chem. 37 (2002) 355.
[20] A. Carta, M. Loriga, G. Paglietti, A. Mattana, P.L. Fiori, P. Mollicotti, L. Sechi, S. Zanetti, Eur. J. Med. Chem. 39 (2004) 195.
[21] A. Carta, P. Corona, M. Loriga, Curr. Med. Chem. 12 (2005) 2259.
[22] R.C. Goldman, B.E. Laughon, R.C. Reynolds, J.A. Secrist III, J.A. Maddry, M.A. Guie, A.C. Poffenberger, C.A. Kwong, S. Ananthan, Infect. Disord. Drug Targets 7 (2007) 92.
[23] TAACF, http://www.taacf.org (December 2008).
[24] Hyperchem 6.03 (Hypercube), http://www.hyper.com.
[25] Milano Chemometrics and QSAR Research Group, http://michem.disat.unimib.it/chm.
[26] P.R. Duchowicz, E.A. Castro, F.M. Fernandez, M.P. González, Chem. Phys. Lett. 412 (2005) 376.
[27] P.R. Duchowicz, E.A. Castro, F.M. Fernandez, MATCH-Commun. Math. Comput. Chem. 55 (2006) 179.
[28] P.R. Duchowicz, M. Fernandez, J. Caballero, E.A. Castro, F.M. Fernandez, Bioorg. Med. Chem. 14 (2006) 5876.
[29] A.M. Helguera, P.R. Duchowicz, M.A.C. Pérez, E.A. Castro, M.N.D.S. Cordeiro, M.P. González, Chemometr. Intell. Lab. 81 (2006) 180.
[30] S.S. So, M. Karplus, J. Med. Chem. 39 (1996) 1521.
[31] Matlab 7.0, The MathWorks Inc.
[32] H. Kubinyi, Quant. Struct.-Act. Relat. 13 (1994) 285.
[33] H. Kubinyi, Quant. Struct.-Act. Relat. 13 (1994) 393.
[34] M. Randic, J. Chem. Inform. Comput. Sci. 31 (1991) 311.
[35] S. Wold, L. Eriksson, Statistical validation of QSAR results, in: H. van de Waterbeemd (Ed.), Cheometrics Methods in Molecular Design, VCH, Weinheim, 1995, p. 309.
[36] D.M. Hawkins, S.C. Basak, D. Mills, J. Chem. Inform. Comput. Sci. 43 (2003) 579.
[37] A. Golbraikh, A. Tropsha, J. Mol. Graph. Model. 20 (2002) 269.
[38] F. Harary, Graph Theory, Addison-Wesley, Reading, MA, 1969.
[39] G. Moreau, P. Broto, Nouv. J. Chim. (New J. Chem.) 4 (1980) 359.
[40] F.R. Burden, J. Chem. Inform. Comput. Sci. 29 (1989) 225.
[41] B.D. Silverman, J. Chem. Inform. Comput. Sci. 40 (2000) 1470.
[42] V. Consonni, R. Todeschini, Rational Approaches to Drug Design, Prous Science, Barcelona, 2001, pp. 235.
[43] J.H. Schuur, P. Selzer, J. Gasteiger, J. Chem. Inform. Comput. Sci. 36 (1996) 334.
[44] V. Consonni, R. Todeschini, M. Pavan, P. Gramatica, J. Chem. Inform. Comput. Sci. 42 (2002) 693.
[45] N.R. Draper, H. Smith, Applied Regression Analysis, John Wiley & Sons, New York, 1981 .