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

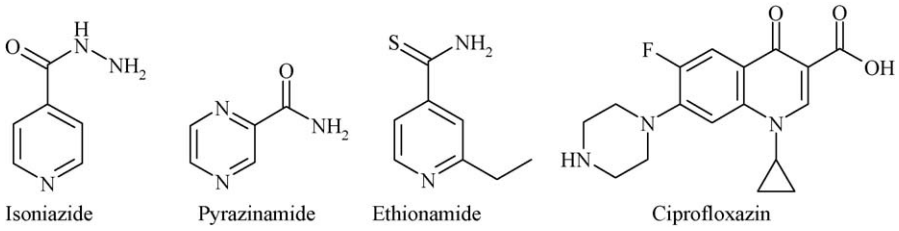

Fig. 1. Structures of certain drugs for treating tuberculosis.

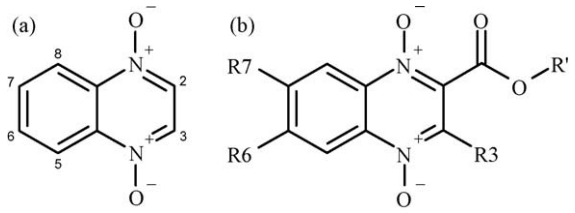

Fig. 2. (a) Numbered quinoxaline 1,4-di-*N*-oxide ring; (b) general structure for quinoxaline-2-carboxylate 1,4-di-*N*-oxide derivatives in Tables 1 and 3.

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