Journal of Molecular Graphics and Modelling 54 (2014) 19–31

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

# Modeling structure–activity relationships of prodiginines with antimalarial activity using GA/MLR and OPS/PLS

Luana Janaína de Campos, Eduardo Borges de Melo*

*Theoretical Medicinal and Environmental Chemistry Laboratory (LQMAT), Department of Pharmacy, Western Paraná State University (UNIOESTE), Cascavel, PR, Brazil*

ARTICLE INFO

*Article history:*
Accepted 25 August 2014
Available online 6 September 2014

*Keywords:*
Prodiginines
Malaria
QSAR
PLS
OPS
Genetic algorithm

ABSTRACT

In the present study, we performed a multivariate quantitative structure-activity relationship (QSAR) analysis of 52 prodiginines with antimalarial activity. Variable selection was based on the genetic algorithm (GA) and ordered predictor selection (OPS) approaches, and the models were built using the multiple linear regression (MLR) and partial least squares (PLS) regression methods. The leave-N-out crossvalidation and ***y***-randomization tests showed that the models were robust and free from chance correlation. The mechanistic interpretation of the results was supported by earlier findings. In addition, the comparison of our models with those previously described indicated that the OPS/PLS-based model had a higher quality of external prediction. Thus, this study provides a comprehensive approach to the evaluation of the antimalarial activity of prodiginines, which may be used as a support tool in designing new therapeutic agents for malaria.



## 1. Introduction

Malaria is a mosquito-borne infectious disease caused by parasitic protozoa, which is considered the most prevalent parasitic disease in the world. Approximately one-third of the world population lives in malaria-endemic areas. About 250 million people in more than 109 countries are affected, with 90% of deaths occurring in Africa (80% in sub-Saharan Africa). The disease has a great impact on the public health and financial situation in these countries; the economic loss in affected African countries is estimated to be US$ 12 billion every year, thus complicating therapeutic intervention. In addition, malaria may look unattractive to private pharmaceutical industry, mainly because of low purchasing power of the affected population, which emphasizes the necessity to develop cost-effective approaches to diagnose and treat malaria [1–9].

Malaria is caused by the parasites of *Plasmodium* spp. (*P. vivax*, *P. falciparum*, *P. ovale*, and *P. malariae*) transmitted primarily by mosquitoes of the genus *Anopheles*. *P. vivax* and *P. falciparum* are responsible for 80% of human cases [10,11]. The disease is characterized by intermittent fever occurring every 2 or 3 days, headache, body aches, anemia, jaundice, and swelling of the liver and spleen [12]. Treatment is very complex and often based on two or three different drugs used in combined mode [10,13]. The most common chemotherapy involves chloroquine and sulfadoxine–pyrimethamine. In the recent years, the derivatives of artemisinin, a natural product extracted from a Chinese plant *Artemisia annua* have been introduced and are currently the treatment of choice in sub-Saharan Africa [14].

The development of resistance to antimalarial agents, including artemisinin derivatives, is one of the main factors underlying the need for the development of new antimalarial drugs. This need is reinforced by inadequate pharmacokinetic properties, adverse effects, toxicity and high cost of current antimalarial agents [10,13–15]. In this context, prodiginines (Fig. 1) gained attention as natural products with antimalarial activity. Prodiginines are a class of red-pigmented secondary metabolites produced by actinomycetes and other eubacteria [16,17]. These compounds have been described to have multiple activities, including antibacterial, anticancer, and immunosuppressive effects. The antimalarial activity was first described by Gerber [18] and Papireddy et al. [16], who showed that prodiginines exhibited in vitro activity against *Plasmodium* species at lower concentrations than other agents did. The chemical structure of prodiginines has been a focus of attention because the first prodiginine derivative with potential therapeutic activity, 2-(2-((3,5-dimethyl-1*H*-pyrrol-2-yl)methylene)-3-methoxy-2*H*-pyrrol-5-yl)-1*H*-indole (GX15-070; Obatoclax®) [20] (Fig. 1), is currently being tested in clinical trials for the treatment of lung cancer, leukemia, and other types of

* Corresponding author at: Department of Pharmacy, UNIOESTE, 2069 Universitária St., 85819110 Cascavel, PR, Brazil. Tel.: +55 45 3220 3256.
*E-mail address:* eduardo.b.de.melo@gmail.com (E.B. de Melo).

http://dx.doi.org/10.1016/j.jmgm.2014.08.004
1093-3263/

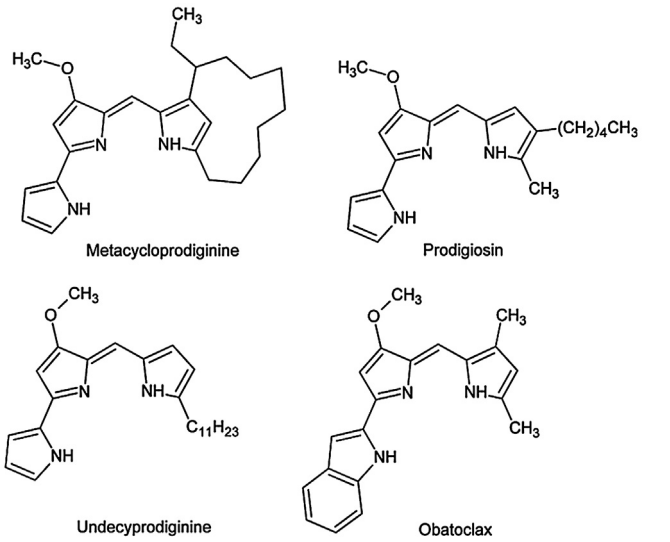


**Fig. 1.** Structures of naturally occurring prodiginines and commercial Obatoclax®.

malignancies [19], indicating that the compounds with this chemical structure may be safely used in clinical practice.

In this scenario, methods of quantitative structure–activity relationships (QSAR) should be useful tools for the development of new drugs. The approach is based on the assumption that the behavior of a set of structurally similar compounds in a biological system (in vitro or in vivo) can be quantitatively described by mathematical models, which can predict the activity of structural analogs not yet synthesized. The success of the QSAR methodology is a great assistance in reducing overhead costs, decreasing the time of obtaining positive results, reducing the use of laboratory animals as well as chemical and biological waste during drug development [21–25].

This study was aimed to obtain QSAR models with multiple linear regression (MLR) and partial least squares (PLS) based on a set of prodiginine derivatives described by Papireddy et al. as antimalarial agents [16]. The models were based on classical molecular descriptors and was constructed, with the aid of variable selection using genetic algorithms (GA) and ordered predictors selection (OPS), respectively [24,26,27].

## 2. Materials and methods

### 2.1. Softwares

Molecular modeling step was performed using the HyperChem software 7 [28] (structural design and optimization in molecular mechanics and semi empirical levels), Gaussian 09 [29] (optimizations in Hartree–Fock and density functional theory levels), Open Babel 2.3.1 [30] (conversion of file formats), Gauss View 5 [31] (visualization of structures and obtaining the electronic descriptors), Dragon 6 [32] (to obtain the other descriptors), QSARINS 1.1 [33] (variable selection by systematic search and genetic algorithm; regression using MLR; selection of the set of external validation; and validation of by MLR), QSAR Modeling [34] (variable selection by OPS; regression by PLS; internal validation of PLS models and robustness checks and random correlation for all the models), and Pirouette 4 [35] (for the refinement and external validation of PLS models). The $r^2_m$ metrics were obtained using the online server RmSquare Calculator (http://aptsoftware.co.in/rmsquare/). Additional test sets were obtained using the Dataset Division GUI 1.0. The validation was also performed with an "in-house" spreadsheet to calculate some statistical parameters of internal and external validation steps. The softwares

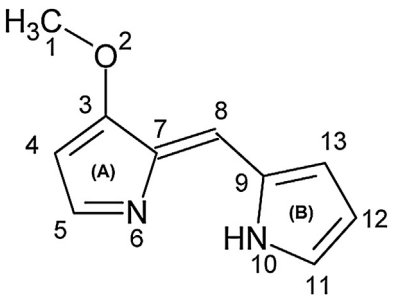


**Fig. 2.** The basic structure of novel prodiginine derivatives. Rings A and B are also identified.

QSARINS, QSAR Modeling and Dataset Division GUI 1.0 may be downloaded in http://www.qsar.it, http://lqta.iqm.unicamp.br, and http://dtclab.webs.com/software-tools, respectively.

### 2.2. Dataset

Papireddy et al. [16] synthesized 52 new prodiginines derivatives (Table 1) and tested their antimalarial activity against the chloroquine resistant D6 strain of *P. falciparum*. The antimalarial activity was measured as the concentration (nM) required to kill 50% of parasites ($EC_{50}$) using the methodology described by Smilkstein et al. [36] and Burgess et al. [37]. The observed $EC_{50}$ values were converted to the corresponding $-\log EC_{50}$ (or $pEC_{50}$), resulting in vector ***y*** with a range of 4.34 logarithmic units (from 4.71 to 9.05). The dataset was divided into training (45 compounds) and test sets (compounds **26**, **28**, **36**, **41**, **49**, **60**, and **65**). The test set was randomly selected using the function of the QSARINS 1.1 [33], but it was verified a posteriori if the selected compounds represented adequately the $pEC_{50}$ range as well as structural variations of the dataset. To ensure the quality of the external prediction, the original auxiliary models [38] was splitted in 14 different training and test sets, in a similar approach to that used recently that Kar et al. [39], using the Kennard–Stone and Euclidean Distance approaches [40]. The adopted identification code of each compound is the same used in the original reference [16].

### 2.3. Molecular modeling

The dataset was built in the tautomeric form #1 (four possible), as described by Masand et al. [41], using as base the crystallographic structure of a synthetic prodiginine available in the support info. of García-Valverde et al. [42] (file jo301008c_si_003.cif). The basic structure of dataset is presented in Fig. 2. All molecules were initially optimized using molecular mechanics (MM+). In this step, the optimizations were alternated with cycles of molecular dynamics (1 ps, 300 K), until the energy obtained in MM+ did not vary more, indicating the obtention of a possible minimum energy structure. Next, simple optimizations were performed at Austin Model 1 (AM1), then in Hartree–Fock (HF/6-31G(d)), and finally density functional theory (DFT) level using the functional Becke, three-parameter, Lee–Yang–Parr (B3LYP), with the basis 6-311G(d,p). The optimization process was carried out in this sequence and steps to reduce the computation time required to obtain the optimized geometries at the level DFT/B3LYP, that was chosen because it has been reported to lead to satisfactory results when molecular geometries and energies are considered [43,44].

### 2.4. Molecular descriptors

Based on three-dimensional structures obtained by molecular modeling at DFT level, 29 electronic descriptors were obtained: Mulliken and Natural Bond Orders (NBO) partial charges for