Ecotoxicology and Environmental Safety 153 (2018) 175–180

Contents lists available at ScienceDirect

Ecotoxicology and Environmental Safety

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

In silico study toward the identification of new and safe potential inhibitors of photosynthetic electron transport

Taisa Pereira Piacentini Ribeiro[a], Flávia Giovana Manarin[b], Eduardo Borges de Melo[a,*]

[a] Dept. of Pharmacy, Western Paraná State University, UNIOESTE, Cascavel, PR, Brazil
[b] Dept. of Chemistry, Western Paraná State University, UNIOESTE, Toledo, PR, Brazil

ARTICLE INFO

Keywords:
Quantitative structure-activity relationship
Virtual screening
In silico toxicology
New herbicides
Quinolines
Naphthalenes

ABSTRACT

To address the rising global demand for food, it is necessary to search for new herbicides that can control resistant weeds. We performed a 2D-quantitative structure-activity relationship (QSAR) study to predict compounds with photosynthesis-inhibitory activity. A data set of 44 compounds (quinolines and naphthalenes), which are described as photosynthetic electron transport (PET) inhibitors, was used. The obtained model was approved in internal and external validation tests. 2D Similarity-based virtual screening was performed and 64 compounds were selected from the ZINC database. By using the VEGA QSAR software, 48 compounds were shown to have potential toxic effects (mutagenicity and carcinogenicity). Therefore, the model was also tested using a set of 16 molecules obtained by a similarity search of the ZINC database. Six compounds showed good predicted inhibition of PET. The obtained model shows potential utility in the design of new PET inhibitors, and the hit compounds found by virtual screening are novel bicyclic scaffolds of this class.

1. Introduction

Agricultural pests and weeds interfere with crop yields. When used in no-tillage practices, chemical agents with herbicidal activity contribute to reduction in soil erosion and increase in nutrient flow, and assist in water conservation. The use of chemical agents requires less labor than mechanical control methods. Therefore, there has been an increase in the annual growth rate of global herbicide market (Green, 2014). However, a highly intensive nature in the use of these substances resulted in the widespread pollution of pesticides in the environment. Added to this is the fact that pesticides have been shown to cause various kinds of organ toxicities, including some types of cancers (Wasi et al., 2013).

Among herbicidal agents, those that act as photosystem II (PSII) inhibitors are the most commonly used agents in agriculture. These agents inhibit the photochemical phase of photosynthesis and consequently NADPH and ATP production, leading to the interruption of carbon fixation by plants. Because electrons cannot store chemical energy, they form free radicals, which lead to lipid peroxidation of the membrane, resulting in necrosis and death of weeds (Oliveira et al., 2011; Hess, 2000). Several herbicides that act by this mechanism, for instance, atrazine, diuron, and metribuzin, are currently available. However, environmental and safety issues, which have resulted in the discontinuation of some herbicides, and the evolution of herbicide-resistant weeds, combined with the fact that no new herbicides exhibiting beneficial effects via new mechanisms of action have been available in the last few decades, have led to the need for development of new chemical agents as herbicides (Duke, 2012).

In medicinal chemistry, the concept of privileged structures refers to the idea that certain structural features produce biological effects more often than others (Polanski et al., 2012). These structures include quinoline (Hussaini, 2016) and naphthalene (Horton et al., 2003) scaffolds. Computer-aided molecular design tools are currently very important in the rational designing of new biologically active chemicals and can be used for the development of new molecules based on privileged structures.

Among these tools, quantitative structure-activity relationship (QSAR) model describes how a given biological activity can vary as a function of molecular structure in a set of chemical compounds (Csizmadia and Enriz, 2000). Although these tools are widely used in drug development and in several studies on environmental toxicology, they are still little explored in the development of new herbicidal agents. QSAR models can also be used to predict the response of new herbicide candidates (González et al., 2003; Zuo et al., 2016; Sharma, 2016). Therefore, we performed a multivariate QSAR study based on a set of 44 derivatives (Musiol et al., 2007; Gonec et al., 2013), with the objective of obtaining models that can be helpful as support tools for designing new PSII inhibitors. A 2D similarity-based virtual screening

* Correspondence to: Theoretical Medicinal and Environmental Chemistry Laboratory (LQMAT), Dept. of Pharmacy, 2069 Universitária St, 85819110, Cascavel, PR, Brazil.
E-mail address: eduardo.b.de.melo@gmail.com (E. Borges de Melo).

https://doi.org/10.1016/j.ecoenv.2018.02.016
Received 5 November 2017; Received in revised form 31 January 2018; Accepted 2 February 2018

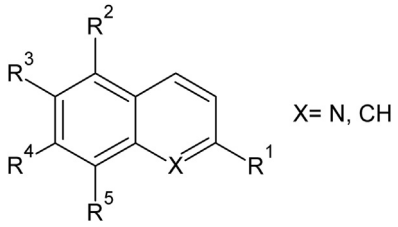

Fig. 1. Basic structure of quinolines and naphthalenes used as data set.

was also performed, with the objective of identifying a related scaffold for the future synthesis of new derivatives.

2. Methods

2.1. Data set

The dataset for this study consisted of 44 selected derivatives, including 26 quinolines and 18 naphthalenes (Fig. 1 and Supplementary material, Table S1), capable of 50% inhibition of photosynthesis in spinach chloroplasts (IC_{50}, in μmol/L) (Musiol et al., 2007; Gonec et al., 2013). The observed values were converted into their corresponding $-\log IC_{50}$ (or pIC_{50}). The activities were distributed within the range of 2.329 log units (pIC_{50} from 2.796 to 5.125). For the validation step, a training set consisting of 37 compounds and a test set consisting of seven compounds were used. The test set was selected to adequately represent the structural variability and biological activity range of the dataset.

The structures of the data-set compounds were built using HyperChem 7 (Hyper Co.) from crystallographic structures (CIF codes: 2201734, 7213893, 2208696, and 2218249) obtained from the Crystallography Open Database (Gražulis et al., 2012) (http://www.crystallography.net/cod). All structures were optimized using molecular mechanics and quantum mechanics strategies until the energy obtained no longer varied, indicating a possible minimum energy structure. The compounds were then optimized using quantum mechanics in Gaussian 09 (http://www.gaussian.com) by applying the Austin Model 1 (AM1) and Hartree-Fock (HF/6-31Gd,p). In the final step, density functional theory (DFT) (B3LYP/6–311 G + + d,p) was used.

2.2. Molecular descriptors

The following electronic descriptors were obtained in the GaussView 5 program (http://www.gaussian.com): Mulliken's partial charges of structure common to all derivatives, total energy (E_T), total dipole moment (D) in *x* (D_X), *y* (D_Y), and *z* (D_Z) axes, and the energies of the two highest occupied molecular orbitals (E_{HOMO-1} and E_{HOMO}) and two lowest unoccupied molecular orbitals (E_{LUMO} and E_{LUMO+1}). In addition, electrophilicity index (ω), electrophilicity index in the ground state (ω_{gs}), molecular electronegativity (c), molecular hardness (h) and softness (S), ionization potential (IP), activation energy index (AEI), electronic affinity (EA), difference between E_{HOMO} and E_{LUMO} (GAP), and the fraction of E_{HOMO}/E_{LUMO} energy ($f_{(H/L)}$) were calculated. These descriptors were obtained using the equations described by Todeschini and Consonni (2009). Moreover, 4855 molecular descriptors (divided into constitutional, topological, geometric, molecular, and mixed) were calculated in the Dragon 6 program (http://www.talete.mi.it/index.htm).

Next, a matrix with all descriptors was treated with variable reduction filters (also in Dragon 6) to eliminate descriptors that did not present information relevant to the model. The filters were used to eliminate: (i) descriptors with constant values; (ii) descriptors with constant and near-constant variables; (iii) descriptors with a standard deviation of less than 0.001; (iv) descriptors with at least one missing value; and (v) descriptors with correlation to another descriptor larger than or equal to 0.90.

A manual reduction was also performed to remove variables that still showed minor variation. The final reduction step was performed using the QSAR modeling software LQTA-QSAR (Martins and Ferreira, 2009) (http://lqta.iqm.unicamp.br), wherein descriptors that had absolute correlation with biological activity ($|r|$) below 0.2 and did not have relevant information for model construction were excluded. In the end, a matrix with 337 descriptors was obtained.

2.3. Variable selection and construction of models

In QSAR study, the variable selection process is usually conducted in an automated manner owing to the large number of descriptors available. This was done in the QSAR modeling software using ordered predictors selection (OPS) (Teófilo et al., 2009). This method uses partial least squares regression (PLS) (Liu and Long, 2009) to assign importance to each descriptor based on three possible informative vectors: correlation vector, regression vector, and the product between them. The final models were also constructed using PLS regression.

2.4. Validation of models

Validation methods are used to check the quality of QSAR models, thus providing a measure of their capability to perform reliable predictions (Gaudio and Zandonade, 2001). The quality of the obtained model was tested through two validation steps: internal and external. To be internally validated, the model must present a good degree of fit, significance, and predictability. These criteria can be evaluated through: (i) coefficient of determination (R^2), which must be greater than 0.6 (i.e., must be able to explain at least 60% of variability of the observed values of biological activity); (ii) *F* test—for correlating the variability explained by the model (R^2) and the variability that remains unexplained (root mean square error of calibration, *RMSEC*), which should have the highest possible values in relation to a tabulated critical value; (iii) leave-one-out (LOO) cross-validation—a procedure by which a compound is excluded from the model. The model is then reconstructed to calculate the value of the excluded object to obtain the coefficient of determination of cross-validation (Q^2_{LOO}), which must be able to predict at least 50% of variability of the observed values of biological activity; (iv) RmSquare metrics [average r^2_m(LOO)-scaled and Δr^2_m(LOO)-scaled] of cross-validation that aids in confirming the predictability expressed by Q^2_{LOO}, because in some cases, a large value for this parameter does not necessarily indicate a good predictability. The $r_m{}^2$ metric is the result of a correlation between observed and predicted values without (r^2) and with ($r^2{}_0$) the prediction values centered on the origin. The same criterion can be applied in the external validation step; (v) ***y***-randomization, which aims to evaluate whether the variabilities explained and predicted by the model are due to chance. In this process, the significance of R^2 and Q^2_{LOO} values is estimated by the development of parallel models, maintaining the values of original descriptors (matrix X), and scrambling the values of the dependent variable (vector ***y***) between the samples. These new models may be necessarily worse or there is a possibility that the data fit is mainly due to spurious correlations. It is expected that the values of these two parameters together will be considerably lower than the original values (without permutation), and this quality is expressed by the values of the intercepts of the new models ($Q^2 < 0.05$ and $R^2 < 0.3$) (Kiralj and Ferreira, 2009); and (vi) evaluation of the robustness of the model, which aims to verify the ability of the model to resist small and deliberate variations. For this purpose, leave-*N*-out (LNO) cross-validation is used, which aims to evaluate whether the model has the capacity to resist small and deliberate variations in its composition (Kiralj and Ferreira, 2009). The commonly used N value is 25–30% of the total number of training set samples (Ferreira et al., 2002; Lang et al., 2014; Roy and Mitra, 2012; Golbraikh and Tropsha, 2002).