

2D, 3D and Hybrid QSAR Studies of Nostoclode Analogues as Inhibitors of the Photosystem II

Pedro O. M. de Carvalho^a and Márcia M. C. Ferreira^{✉*,a}

^aInstituto de Química, Universidade Estadual de Campinas, 13083-970 Campinas-SP, Brazil

Synthetic herbicides are widely used for weed control in crops. Continuous application of chemicals induces the adaptation of weeds, leading to the development of resistance. Therefore, research for novel synthetic herbicides plays an important role in crop protection and food production. Within this context, it is important to understand the relationship between the structures of a set of compounds endowed with herbicide activity and their biological response. In this work, a quantitative structure-activity relationship (QSAR) study of 34 nostoclode analogues was carried out in order to analyze their ability to inhibit the photosystem-II. Compounds were optimized using the DFT/B3LYP/Def2-TZVPP method. 2D and 3D (Lennard-Jones and electrostatic potential energies) molecular descriptors were calculated at the same theory level. QSAR models showed major importance of electrostatic ligand-acceptor interactions and indicated the presence of aromatic interaction in the benzyl group. A photoelectron transfer via hydrogen-bond was proposed to occur between His215 residue and the carbonyl group.

Keywords: QSAR, DFT, herbicides, photosystem II, nostocclides

Introduction

Human growing population and prosperity around the world is driving an increased demand for food. Therefore, agriculture has developed to achieve higher yields and better-quality products. To yield high crop production, it is important to protect the cultures from pests and diseases.^{1,2} Weeds produce the highest potential loss overall among the pests, which can be avoided by applying physical, biological and chemical measures.² Crop protection by chemical measures spreads worldwide since the commercialization of the two phenoxyacid herbicides: 2,4-dichlorophenoxyacetic acid (usually called 2,4-D) and the 2-methyl-4-chlorophenoxyacetic acid (MCPA).^{2,3} The use of chemicals in crops has become the main method of weed control because of its low cost and high efficiency, encouraging reliance on them.⁴ However, frequent use of chemicals with similar biological mechanism creates a selective pressure on weeds, leading to the development of resistance.^{1,4,5} Herbicide resistance was firstly reported in 1970 and confirmed in 495 unique cases until 2017.^{5,6} Usually, the resistance acquired by a weed species is a result of a mutation in the active site. The development of herbicides with a different mechanism of action is

desirable, because it will generally be able to eliminate the herbicide-resistant weed.⁷ However, Duke,⁷ in 2012, stated that “no new major herbicide mode of action has been introduced in a commercial herbicide active ingredient in the last 20 years”. Moreover, environmental and health hazards concerns related to pesticides have grown since 1960’, which led to stringent regulation of pesticides.⁴ It is imperative to find new substances with herbicide properties to fight weed resistance and that should be in agreement with the environmental protection agencies regulations.

There are different experimental strategies in the search for novel herbicides. They usually involve the measurement and analysis of herbicide activity of sets of analogue compounds. These experimental activities could be further exploited using quantitative structure-activity relationship (QSAR) studies. For that, molecular descriptors, that could be related to biological properties based on the chemical structure of the set of compounds, are calculated and mathematical models are built to predict the biological property using the desirable molecular descriptors. The interpretation of such models could provide information about the ligand-receptor interactions and the mechanism of action of the set of molecules. QSAR models can be also used to predict the biological activity of compounds with similar structures.⁸⁻¹¹ Therefore, QSAR study is an important research tool when applied to the development

*e-mail: marcia@iqm.unicamp.br

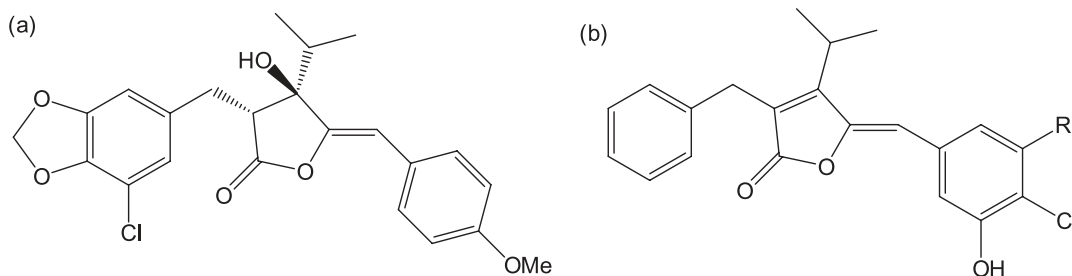


Figure 1. Structure of (a) cyanobacterin; and (b) nostoclides (nostoclides I: R = Cl; nostoclides II: R = H).

of chemicals with biological activity, such as herbicides.^{8,12}

A large number of QSAR studies applied to novel herbicides was published in last few years. Karacan *et al.*¹² reported a QSAR analysis on perfluoroisopropylidinitrobenzene derivatives, a type of dinitrophenolic compound, which have been used in crop protection. A 2D-QSAR model of a series of 1*H*-1,4-benzodiazepine-2,5-dione analogues was proposed by Banjare *et al.*¹³ and used to generate a library of new compounds with comparable and better expected activity. Teixeira *et al.*¹⁴ used photosynthetic inhibition activity data (IC_{50}) for 19 nostoclides analogues to build and analyze a 2D-QSAR model, which indicated that their activity is associated with their polarity. A series of rubrolide analogues were synthesized by Barbosa *et al.*¹⁵ and 2D-QSAR models were built and analyzed using their photosynthetic inhibition activity, leading to insights about the most effective substituents among the set. Funar-Timofei *et al.*¹⁶ published a QSAR study in combination with molecular docking applied to fused heterocyclic herbicides that act as inhibitors of the D1 protein of photosystem II. A 3D-QSAR study of series of 3-(pyridin-2-yl)benzenesulfonamide derivatives was made by Xie *et al.*¹⁷ based on CoMFA and CoMSIA models, which were then used to successfully design new derivatives with superior herbicide activity. In recent work, Teixeira *et al.*¹⁸ synthesized and measured the photosynthetic inhibitory activity of a set of trifluoromethyl arylamides, which were also used to build a 2D-QSAR model.

Herbicides that target the photosystem II compete with the plastoquinone Q_B for its binding site and block the electron transfer to plastoquinone Q_A , thus inhibiting the photosynthetic electron transport.¹⁹ Trebst²⁰ proposed the classification of the herbicides that act on photosystem II into two families: the serine family, whose herbicides bind oriented towards the D1-Ser264 residue; and the histidine family, whose herbicides bind oriented towards the D1-His215 residue.

Cyanobacterin (Figure 1a) is a natural occurring lactone that has the property of specifically inhibit photosynthetic electron transport.^{21,22} Nostoclides (Figure 1b) are another

natural occurring lactones that resembles cyanobacterin, but their herbicide properties were not fully investigated yet.¹ Teixeira *et al.*¹ synthesized and measured the photosynthetic inhibitory activity of a series of nostoclides analogues to investigate their potential phytotoxicity.

In the present study, the reported experimental results were used to build QSAR models to predict the photosynthetic inhibitory activity of 34 nostoclides analogues.¹ The best models were interpreted in order to understand the ligand-receptor interactions and provide some insight on the mechanism of action of this class of compounds.

Methodology

Data set

A set of 34 3-benzyl-5-(arylmethylene)furan-2(5*H*)-ones (Figure 2) and their photosynthetic inhibitory activity were used. The aromatic ring of the arylidene portion presented different substitution patterns. The percentages of photosynthetic inhibitory activity (PI) are relative to untreated control for an analogue concentration of $10 \mu\text{mol L}^{-1}$. Table 1 shows the nostoclides analogues and the respective PI values. The molecules that presented PI lower than 5% were originally considered ineffective.¹ In this work, such compounds were included in the QSAR study considering their PI as 5%.

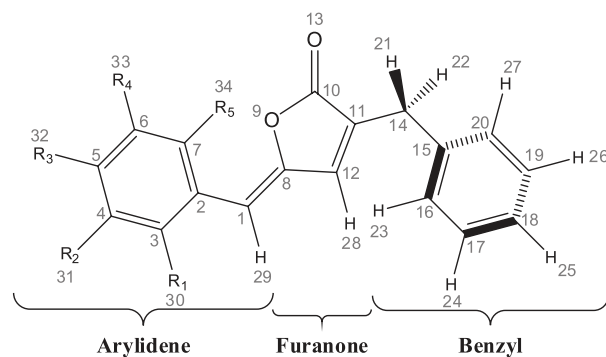


Figure 2. General structure and atom indices of the nostoclides analogues studied. The substituents R_1 - R_5 for each analogue is presented in Table 1.

Table 1. Substituents and photosynthetic inhibitory activity of the studied nostoclide analogues^a

Molecule	R ₁	R ₂	R ₃	R ₄	R ₅	PI ^b / %
m01	-H	-H	-H	-H	-H	29.2
m02	-CH ₃	-H	-H	-H	-H	26.0
m02i	-H	-H	-H	-H	-CH ₃	16.7
m03	-H	-CH ₃	-H	-H	-H	8.0
m03i	-H	-H	-H	-CH ₃	-H	50.7
m04	-H	-H	-CH ₃	-H	-H	NE ^c
m05	-H	-H	-CH ₂ CH ₃	-H	-H	NE ^c
m06	-H	-CN	-H	-H	-H	44.9
m06i	-H	-H	-H	-CN	-H	49.9
m07	-H	-H	-CN	-H	-H	55.5
m08	-CF ₃	-H	-H	-H	-H	NE ^c
m08i	-H	-H	-H	-H	-CF ₃	38.6
m09	-H	-CF ₃	-H	-H	-H	7.1
m09i	-H	-H	-H	-CF ₃	-H	25.4
m10	-H	-H	-CF ₃	-H	-H	39.5
m11	-H	-H	-Ph	-H	-H	57.8
m12	-H	-N(CH ₃) ₂	-H	-H	-H	NE ^c
m12i	-H	-H	-H	-N(CH ₃) ₂	-H	39.5
m13	-H	-H	-N(CH ₃) ₂	-H	-H	57.8
m14	-Cl	-H	-N(CH ₃) ₂	-H	-H	6.9
m14i	-H	-H	-N(CH ₃) ₂	-H	-Cl	6.5
m15	-H	-NO ₂	-H	-H	-H	NE ^c
m15i	-H	-H	-H	-NO ₂	-H	6.9
m16	-H	-H	-NO ₂	-H	-H	6.5
m17	-OH	-H	-H	-H	-H	NE ^c
m17i	-H	-H	-H	-H	-OH	6.9
m18	-H	-OH	-H	-H	-H	6.5
m18i	-H	-H	-H	-OH	-H	NE ^c
m19	-H	-H	-OH	-H	-H	6.9
m20	-H	-OH	-OCH ₃	-H	-H	6.5
m20i	-H	-H	-OCH ₃	-OH	-H	NE ^c
m21	-H	-H	-OCH ₃	-H	-H	15.0
m22	-OCH ₃	-H	-H	-OCH ₃	-H	43.5
m22i	-H	-OCH ₃	-H	-H	-OCH ₃	43.6
m23	-OCH ₃	-H	-OCH ₃	-H	-OCH ₃	NE ^c
m24	-H	-OCH ₃ O-	-H	-H	-H	NE ^c
m24i	-H	-H	-OCH ₃ O-	-H	-H	49.5
m25	-F	-H	-H	-H	-H	28.8
m25i	-H	-H	-H	-H	-F	25.6
m26	-H	-F	-H	-H	-H	6.6
m26i	-H	-H	-H	-F	-H	30.8
m27	-H	-H	-F	-H	-H	20.3
m28	-F	-F	-F	-F	-F	5.8
m29	-Cl	-H	-H	-H	-H	21.8
m29i	-H	-H	-H	-H	-Cl	22.6
m30	-H	-Cl	-H	-H	-H	NE ^c
m30i	-H	-H	-H	-Cl	-H	21.8
m31	-H	-H	-Cl	-H	-H	5.8
m32	-Br	-H	-H	-H	-H	21.8
m32i	-H	-H	-H	-H	-Br	22.6
m33	-H	-Br	-H	-H	-H	22.6
m33i	-H	-H	-H	-Br	-H	NE ^c
m34	-H	-H	-Br	-H	-H	NE ^c

^aThe analogues highlighted in grey are those used to build the final QSAR models. The substituents R₁, R₂, R₃, R₄ and R₅ are related to Figure 2; ^bpercentage of photosynthetic inhibitory activity; ^cnot effective: PI < 5%.

Molecular descriptor calculation

The crystallographic structure of compound **m03** (Table 1) was retrieved from Cambridge Structural Database (CSD) (code CONPEB).²³ Geometry optimization was carried out in Gaussian 09 software (G09Rev-D.01) using DFT/B3LYP method and Def2-TZVPP basis set.²⁴⁻²⁷

Some molecules presented substituents at the *meta* and *ortho* positions of the aromatic ring of the arylidene portion. In these cases, several conformations arise from rotation around C1-C2 bond (Figure 2) and there are two preferential ones (Figure 3). Both conformations were optimized and one of them was selected for further analysis, as described in the next sub-sections.

The 2D molecular descriptors were calculated at the same theory level in Gaussian 09 software (G09Rev-D.01)²⁴ for the optimized structures. They were: total and molecular orbital energies; molecular orbitals energy gaps; dipole and quadrupole moments; mean polarizability; Mulliken and CHELPG (charges from electrostatic potentials using a grid based method) partial atomic charges; and sum of squares of the wave functions' molecular orbital coefficients (for the frontier orbitals ranging from highest occupied molecular orbital (HOMO)-4 to lowest unoccupied molecular orbital (LUMO)+4). In addition, the sums of partial atomic charges for certain sets of atoms were also calculated.

The 3D molecular interaction field descriptors were calculated by LQTAGrid software⁸ using a $21 \times 15 \times 9 \text{ \AA}$ grid with 1 \AA increment and a NH_3^+ probe. The alignment was made using the optimized geometries. The reference structure was compound **m01** and the atoms aligned were C8, O9, C10, C11, C12 and O13 (Figure 2). For each grid

point, electrostatic and Lennard-Jones potential energies were calculated.⁸

Conformer selection

From the 34 compounds, 19 assume two different preferential conformations by rotation around the C1-C2 bond, resulting in 19 pairs of conformers. However, only one conformer of each pair was used in the final QSAR models. The selection of a specific conformer in each pair was carried out in two steps. Firstly, a conformational equilibrium at $25 \text{ }^\circ\text{C}$ was considered to calculate the population of each conformer according to equation 1. This equation is derived from Boltzmann distribution, where *P* is the population percentage for each conformer.²⁸ *E* and *E*_{count.} are the optimization energies of the conformer and its counterpart, respectively; *k* is Boltzmann's constant and *T* is the temperature. The same procedure was applied to each one of the 19 pairs of conformers.

$$P(\%) = \left[\exp\left(\frac{E - E_{\text{count.}}}{kT}\right) + 1 \right]^{-1} \times 100 \quad (1)$$

The conformers with *P* < 1% compared to its counterparts were eliminated. For example, the compound **mXXi** would be excluded if it presented *P* < 1% compared to its counterpart, **mXX**.

Considering the remaining 11 pairs of conformers, one of each was excluded and the remaining compounds were used to build a partial least squares (PLS) regression model with the 2D descriptors. This procedure was repeated exchanging the excluded conformers to test all possible combinations. These regression models were internally validated by leave-one-out cross-validation (LOOCV) and

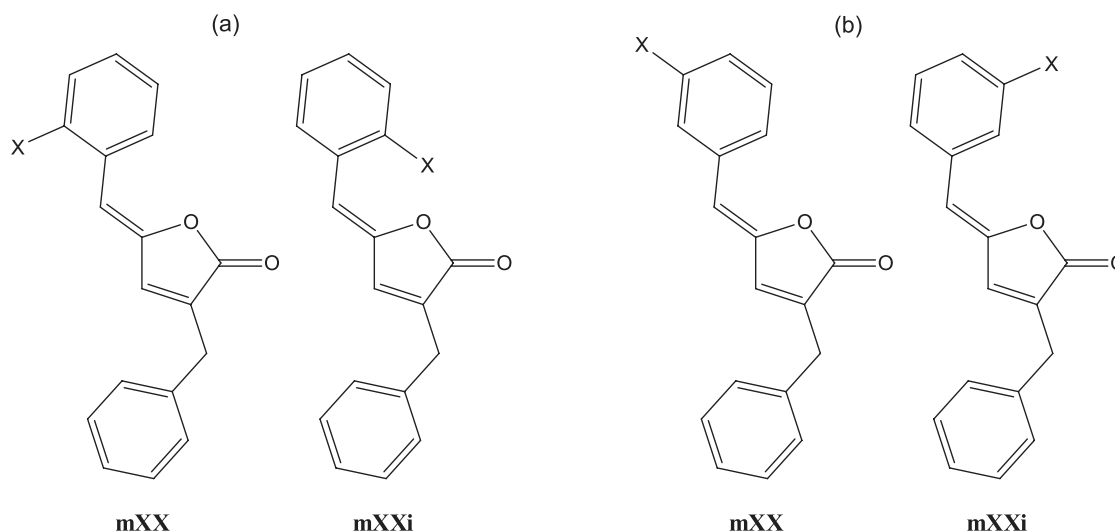


Figure 3. General structure of the two preferential conformations considered in this work for substituents in: (a) R₁ and R₃; (b) R₂ and R₃.

the combination of conformers corresponding to the models with the best statistics was selected.

Descriptor selection

The descriptors selection was carried out on the conformer combination selected in “Conformer selection” sub-section. The 2D descriptors were inspected to remove those degenerated. The following step was to eliminate descriptors that had the absolute value of Pearson’s correlation coefficient ($|r|$) with biological activity lower than 0.3. The ordered predictors selection algorithm (OPS) implemented in the public domain QSAR modeling software was used to reduce the number of descriptors to be kept in the regression model.^{29,30} Finally, they were manually selected in Pirouette-3.11³¹ software by building PLS regressions and comparing their statistics.

The 3D descriptors were selected by applying the digital filter for molecular interaction field descriptors (comparative distribution detection algorithm, CDDA).³² OPS algorithm and Pirouette-3.11 software were used in the next step of the selection. Descriptors’ probes and molecules were visualized using Chimera-1.11.³³

All the QSAR-models were built using PLS regressions method on autoscaled data (mean centered and scaled to unity variance) and internally validated by LOOCV.³⁴

Conformer reselection

To ensure that the best combination of conformers was selected, they were retested following the same method described in “Conformer selection” sub-section, but using only the descriptors selected in “Descriptor selection” sub-section. The same procedure was applied for the 2D and 3D molecular descriptors. PLS regressions were performed to choose the best combination of conformers.

The selected combination of conformers and their molecular descriptors were used to build 2D and 3D QSAR models. The 2D and 3D selected descriptors were also combined and reselected manually in Pirouette-3.11³¹ to build a hybrid QSAR model. All the QSAR-models were built using PLS regressions and internally validated by LOOCV (in MATLAB software).³⁵ Other validation tests such as leave- N -out (LNO) cross-validation and y -randomization were carried out in QSAR modelling software.^{36,37}

The molecular structure figures were created using ChemDraw Ultra 12.0,³⁸ Chimera-1.11,³³ GaussView 5.0³⁹ and MATLAB software³⁵ and the graphics were built using Pirouette-3.11³¹ and MATLAB software.³⁵

Results and Discussion

Conformers and molecular descriptors selection

Since conformation at the active site is unknown, two preferential conformations of nostoclidic analogues were initially considered (Figure 3). However, only one conformer from each pair was selected for model building. In a first step, the optimization energies of each pair of conformers were compared using equation 1. Only those conformers significantly less stable than its counterpart ($P < 1\%$) were eliminated. The pairs of conformers for which the optimization energies were not significantly different ($1 \leq P \leq 99\%$) were selected in the next step.

Conformers **m02i**, **m08i**, **m14i**, **m17i**, **m22i**, **m25i**, **m29i** and **m32i** presented $P < 1\%$, thus were excluded, whereas their counterparts, **m02**, **m08**, **m14**, **m17**, **m22**, **m25**, **m29** and **m32**, were kept. The eliminated conformers have an R_5 substituent, suggesting instability at this position. Similar observation was previously proposed in a structural and conformational study of nostoclidic analogues; it was assumed that the preferential conformation was due to a non-bonding steric repulsion between O13 and the substituent at R_5 position.²³

The preferred conformer from the remaining pairs were selected in order to choose those molecular descriptors which provided better predictions based on LOOCV of PLS regressions. Thus, the selected conformers are expected to have a conformation which is closer to the one they assume in its active site, or that better correlates with it.

Molecular descriptors selection was carried out using only the analogues that remained after the first conformer selection. When PLS models were built, 5 analogues appeared to be outliers. Two of them, molecules **m23** and **m28**, are the only structures having substituents in both positions R_1 and R_5 . The non-bonding steric repulsion between O13 and the substituent at R_5 position is unavoidable, causing the arylidene and furanone rings to be in different planes. Although it does not imply a problem in the experimental biological activity data, all the other analogues presented an optimized structure with those rings in the same plane (Figure 4). Therefore, the molecular descriptors calculated for molecules **m23** and **m28** are significantly different from the other compounds. The plots scores and sample’s residues *versus* Mahalanobis distance from principal component analysis, presented in Figure 5, confirm the atypical behavior of these two molecules. The remaining outliers were molecules **m05**, **m25** and **m34**, which presented atypical behavior compared to tendency of similar nostoclidic analogues, e.g., **m02-m04**, **m26-m27** and **m32-m33**, respectively.

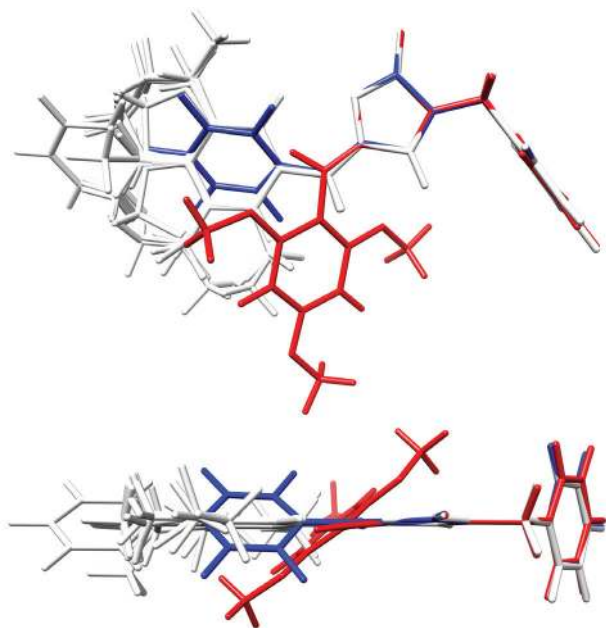


Figure 4. All optimized structures aligned. **m23** in red; **m28** in blue; others: in white.

The conformer selection did not consider the descriptors selection and the exclusion of the outliers. Therefore, the excluded conformers were re-included and a new conformers selection was carried out (conformer reselection). A similar methodology was applied, but excluding outliers and using only the molecular descriptors selected. The final set of molecules used to build the QSAR models are highlighted in grey in Table 1.

2D, 3D and hybrid QSAR models' statistics

A QSAR model predictivity can be evaluated by the values of Q^2 , R^2 , R_c^2 and k_c , which are, respectively, the cross-validated correlation coefficient, correlation coefficient of multiple determination, Pearson's correlation coefficient of

calibration and slope of measured *versus* predicted regression line through the origin. According to the literature, a QSAR model can be considered predictive if:^{36,37}

$$Q^2 > 0.5 \quad (2)$$

$$R_c^2 > 0.6 \quad (3)$$

$$\frac{(R_c^2 - R^2)}{R^2} < 0.1 \quad (4)$$

$$0.85 \leq k_c \leq 1.15 \quad (5)$$

$$R^2 - Q^2 < 0.3 \quad (6)$$

The statistics of the 2D, 3D and hybrid QSAR models are presented in Table 2. All the models satisfy the conditions of inequalities 2 to 6. The root mean square error of calibration (RMSEC) and root mean square error of cross-validation (RMSECV) values of the hybrid QSAR models were slightly better than the 3D QSAR's, while the 2D QSAR presented higher RMSEC and RMSECV.

The measured *versus* predicted percentages of inhibition plots are presented in Figures 6, 7 and 8 for the 2D, 3D and hybrid QSAR models, respectively. The 2D-QSAR model plot presented absolute deviations greater than 10% for 13 molecules, most of them with measured values of PI between 5 and 27%. The 3D and hybrid QSAR models showed better predictions in this range, although not for all molecules (6 of them with absolute deviation larger than 10%). However, the hybrid QSAR model presented lower absolute deviations overall, resulting in a more evident linearity.

Although statistical parameters previously calculated are necessary to indicate the quality of QSAR models, further validations are required to evaluate its robustness and the existence of a real structure-activity relationship. This can be done through the tests of leave-N-out (LNO) cross-validation and of y-randomization, which are presented in Figures 9, 10 and 11.^{36,37}

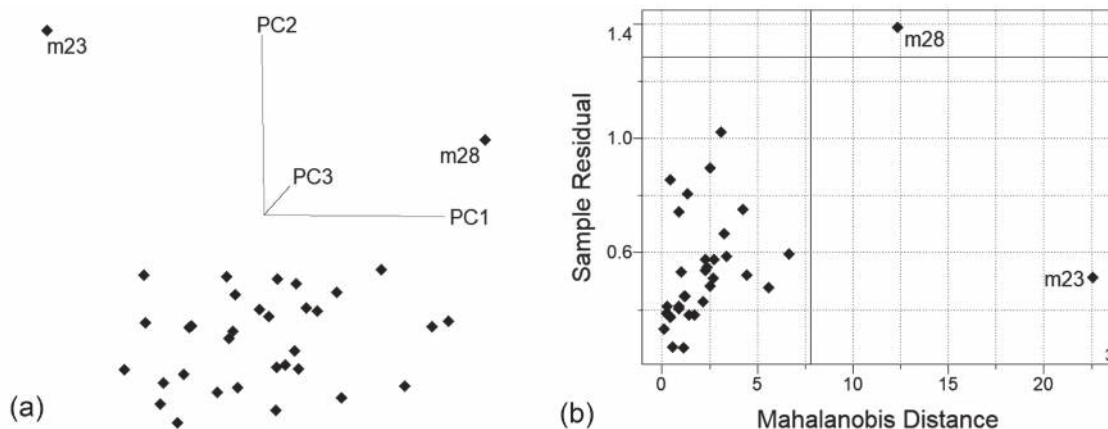
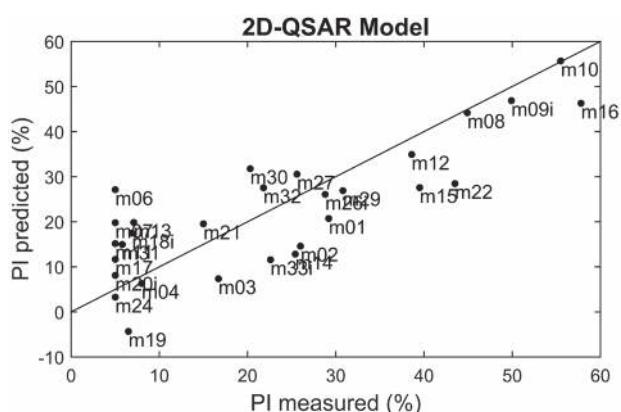


Figure 5. PCA for the 2D and 3D molecular descriptors: (a) PC1 \times PC2 \times PC3 scores plot; (b) applicability domain graph (Mahalanobis distance *versus* sample's residues).

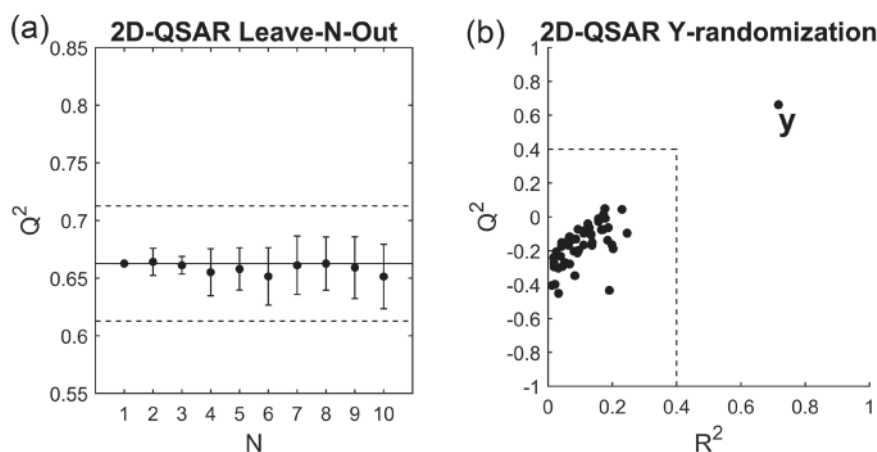
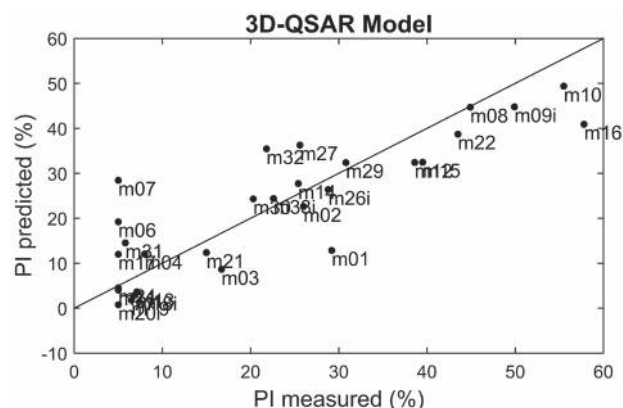
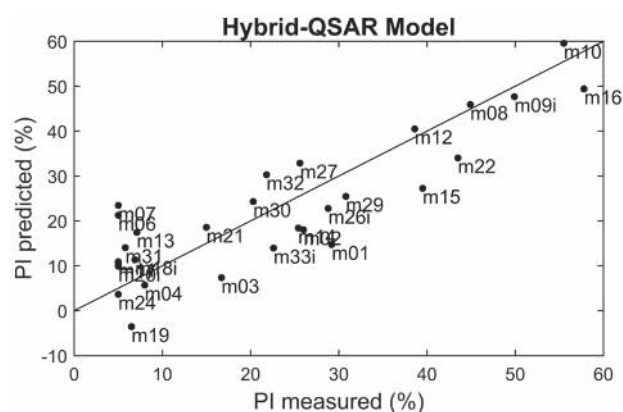
Table 2. 2D, 3D and hybrid quantitative structure-activity relationship (QSAR) model statistics

Model	2D-QSAR	3D-QSAR	Hybrid QSAR
Number of descriptors	5	7	7
Latent variables	1	2	2
Explained variance / %	57.4	63.4	61.2
k_c^a	0.901	0.935	0.939
RMSECV ^b / %	9.11	7.55	7.33
R_C^{2c}	0.717	0.813	0.824
R^{2d}	0.717	0.813	0.824
RMSECV ^e / %	9.60	8.52	8.38
Q^{2f}	0.663	0.734	0.743

^aSlope of measured *versus* predicted regression line through the origin; ^broot mean square error of calibration; ^cPearson correlation coefficient of calibration; ^dcorrelation coefficient of multiple determination; ^eroot mean square error of cross-validation; ^fcross-validated correlation coefficient.

**Figure 6.** Plot of measured *versus* predicted percentages of inhibition for the 2D-QSAR model.

The LNO cross-validation test was repeated 10 times for each value of N . The rows from the data matrix were randomized for each replicate, resulting in average values

**Figure 9.** 2D-QSAR model graphs of (a) leave-N-out cross-validation and (b) y-randomization test.**Figure 7.** Plot of measured *versus* predicted percentages of inhibition for the 3D-QSAR model.**Figure 8.** Plot of measured *versus* predicted percentages of inhibition for the hybrid QSAR model.

of Q_{LNO}^2 . For a robust QSAR model, the average Q_{LNO}^2 is expected to be close to Q_{LOO}^2 even for a large number of molecules (20-30% from original set) removed from the set. It is also desirable that two times the standard deviation for each N (including the critical one) presents values lower than 0.1.^{36,37} In this work, N varied from 2 to 10. Figures 9a, 10a and 11a show the mean value and standard deviation of Q^2 for each N and the limits $Q_{LOO}^2 \pm 0.05$. All the models

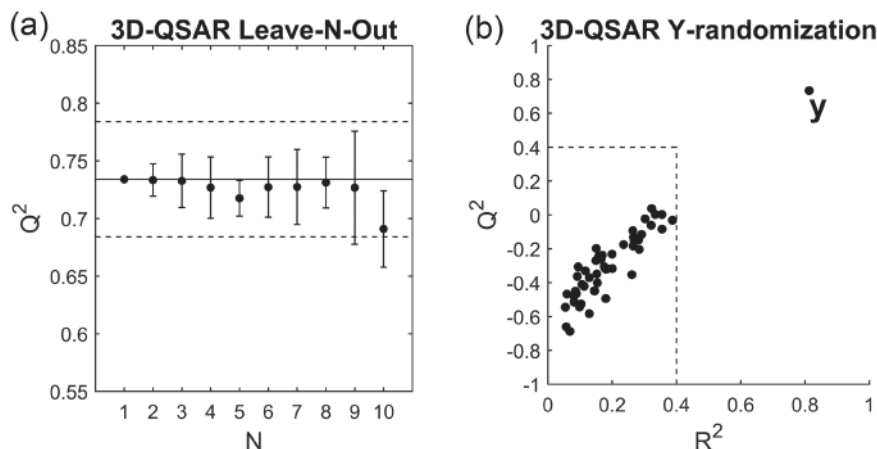


Figure 10. Plots of (a) leave-N-out cross-validation and (b) y-randomization test for the 3D-QSAR model.

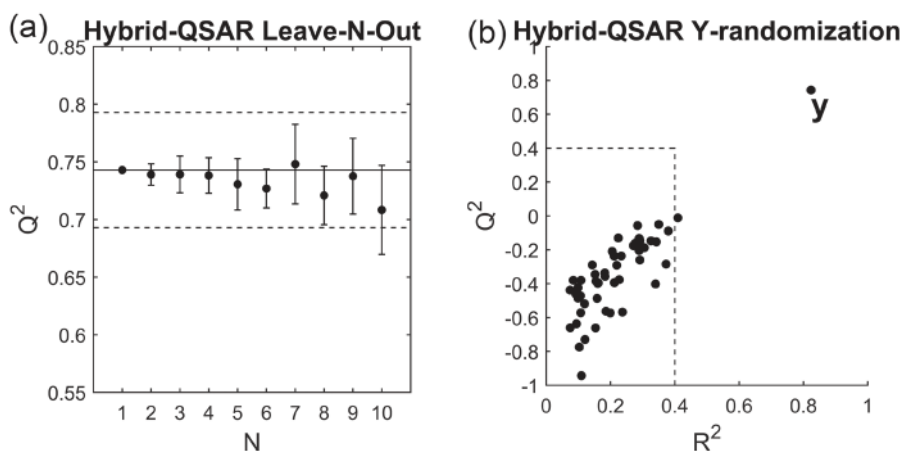


Figure 11. Hybrid QSAR model graphs of (a) leave-N-out cross-validation and (b) y-randomization test.

presented little fluctuation for N varying from 1 to 10. Therefore, all the QSAR models can be considered robust.

The y-randomization test consists in building several QSAR models (50 in this work) for which the vectors of biological activity (y vector) is randomly shuffled. For a good QSAR model, these randomized models are expected to be of poor quality and without real meaning. The basic LOO statistics of the randomized models should be worse than the original model: low Q_{yrand}^2 and R_{yrand}^2 values. A QSAR model that presents relatively high values of Q_{yrand}^2 and R_{yrand}^2 (> 0.2 and > 0.4 , respectively) indicates the existence of chance correlation, which means that the real model may contain descriptors which are statistically well correlated to y , but in reality there is no cause-effect relationship encoded in such correlations.^{36,37,40} Figures 9b, 10b and 11b show the plots of R^2 versus Q^2 for the y-randomization test for the three models: 2D, 3D and hybrid QSAR, respectively. The original model is labeled as y and the dashed lines indicate the cutoff = 0.4. Based on these results, none of the QSAR models built presented chance correlation.

Analysis of the selected molecular descriptors

The correlation matrix and regression coefficients for the 2D, 3D and hybrid QSAR models are shown in Tables 3, 4 and 5. The correlation matrices of these tables present the descriptor-PI and descriptor-descriptor correlation coefficients for each molecular descriptor in the model. Comparison between the regression and the descriptor-PI correlation coefficients for each molecular descriptor shows that they present the same sign. This implies that the contribution of each descriptor to its model is consistent with its correlation to the PI. Table 6 presents the details of the 2D descriptors and Figure 12 shows a tridimensional plot of the aligned nostoclide analogues and the position of the selected 3D descriptors, where “L” and “E” stands for the Lennard-Jones and the electrostatic potentials, respectively.

The most important molecular descriptor for the 2D model (see regression coefficients in Table 3) is QMbZZ, which is one of the components of the electrostatic quadrupole moment of the aromatic ring of the benzyl

Table 3. Descriptors correlation matrix, regression coefficients and model for the 2D QSAR (quantitative structure-activity relationship) model

	CO+CM	wHO02	QMafYY	QMbYY	QMbZZ
wHO02	0.310				
QMafYY	0.107	0.439			
QMbYY	-0.679	-0.212	-0.232		
QMbZZ	0.778	0.558	0.297	-0.794	
PI ^a	-0.683	-0.521	-0.439	0.705	-0.792
R. coef. ^b	-0.238	-0.182	-0.153	0.246	-0.276
R. model ^c	$PI = -498(CO + CM) - 283(wHO02) - 0.358(QMafYY) + 75.6(QMbYY) - 82.2(QMbZZ) - 335$				

^aPercentage of photosynthetic inhibitory activity; ^bregression coefficients for descriptors in autoscaled form; ^cregression mathematical model.

Table 4. Descriptors correlation matrix, regression coefficients and model for the 3D QSAR (quantitative structure-activity relationship) model

	E1	E2	E3	E4	E5	L1	L2
E2	-0.622						
E3	-0.340	0.391					
E4	-0.333	0.407	0.991				
E5	-0.0848	0.0630	0.165	0.190			
L1	-0.293	-0.0468	0.213	0.239	0.487		
L2	-0.335	0.526	0.381	0.382	0.403	-0.0209	
PI ^a	-0.429	0.567	0.571	0.550	-0.344	-0.400	0.407
R. coef. ^b	-0.180	0.244	0.248	0.228	-0.343	-0.373	0.157
R. model ^c	$PI = -0.109(E1) + 0.151(E2) + 0.142(E3) + 0.124(E4) - 0.143(E5) - 44.2(L1) + 4.75(L2) - 293$						

^aPercentage of photosynthetic inhibitory activity; ^bregression coefficients for descriptors in autoscaled form; ^cregression mathematical model.

Table 5. Descriptors correlation matrix, regression coefficients and model for the hybrid QSAR (quantitative structure-activity relationship) model

	CO+CM	wHO02	QMafYY	QMbYY	E1	E3	L1
wHO02	0.310						
QMafYY	0.107	0.439					
QMbYY	-0.679	-0.212	-0.232				
E1	0.184	0.430	0.322	-0.289			
E3	-0.285	-0.499	-0.440	0.441	-0.340		
L1	0.353	0.0225	-0.0269	-0.278	-0.293	0.213	
PI ^a	-0.683	-0.521	-0.439	0.705	-0.429	0.571	-0.400
R. coef. ^b	-0.284	-0.154	-0.137	0.275	-0.149	0.201	-0.292
R. model ^c	$PI = -593(CO+CM) - 240(wHO02) - 0.320(QMafYY) + 84.6(QMbYY) - 0.0907(E1) + 0.115(E3) - 34.6(L1) - 335$						

^aPercentage of photosynthetic inhibitory activity; ^bregression coefficients for descriptors in autoscaled form; ^cregression mathematical model.

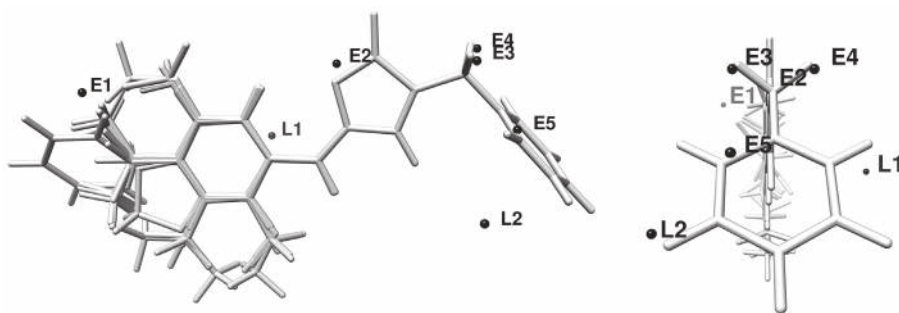
group. This descriptor is negatively correlated to the PI and its values are negative. It means that higher absolute values of QMbZZ correspond to higher inhibition of the photosystem II. The QMbYY molecular descriptor, another component of the same quadrupole moment, was also selected for the 2D QSAR model. It is positively correlated to the PI and its values are positive, resulting in the same behavior of the QMbZZ. The quadrupole moment of an aromatic ring can be associated with

aromatic stacking, sometimes called “pi-stacking”. This is an interaction that is important to many biological systems.^{41,42} At the photosystem II, these interactions were observed only between the plastoquinone Q_A and its site in the D2 protein.^{43,44} However, based on a QSAR study, Karacan *et al.*¹² proposed the existence of aromatic stacking interactions between a set of perfluoroisopropyl-dinitrobenzen derivatives and the plastoquinone Q_B site in the D1 protein. Likewise, the molecular descriptors QMbZZ

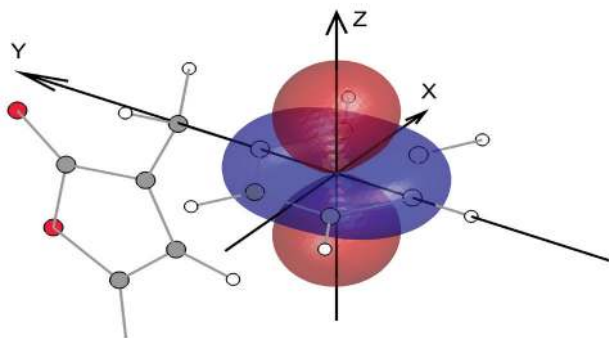
Table 6. 2D molecular descriptors names and details

	Detail
CO+CM	sum of CHELPG atomic charges for <i>ortho</i> and <i>meta</i> carbons of the benzyl's ring, i.e., C16, C17, C19, C20
HO+HM	sum of CHELPG atomic charges for <i>ortho</i> and <i>meta</i> hydrogens of the benzyl's ring, i.e., H23, H24, H26, H27
wHO2	sum of squares of the wave functions' molecular orbital coefficients for C02 carbon
QMafYY ^a	YY component of the electrostatic quadrupole moment from arylidene and furanone, i.e., atoms 1-13 and 28-34
QMbYY ^a	YY component of the electrostatic quadrupole moment from benzyl's ring, i.e., atoms 15-20 and 23-27
QMbZZ ^a	ZZ component of the electrostatic quadrupole moment from benzyl's ring, i.e., atoms 15-20 and 23-27

^aRegression coefficients for descriptors in autoscaled form. CHELPG: charges from electrostatic potentials using a grid based method.

**Figure 12.** Illustration of the aligned nostoclide analogues and the location of the 3D descriptors' probe from two points of view.

and QMbYY indicate the possibility of aromatic stacking interactions between the benzyl group of the nostoclide analogues and the residues of the plastoquinone Q_B site. Figure 13 shows the contribution of the benzyl quadrupole moment to the electrostatic potential for the **m01** analogue, which presents no substituents and medium PI. The blue and red isosurfaces correspond to the same absolute value of electric potential, but with positive and negative signs, respectively.

**Figure 13.** Electric potential isosurface of the quadrupole moment from the aromatic ring of the benzyl group; blue for positive and red for negative values.

The positive isosurface of this quadrupole moment is longer in the Y axis than in the X axis. It also means that there is a dipole moment in the aromatic ring of the benzyl group. Studies regarding the toluene aromatic stacking have shown that its interactions occur preferably

in a parallel offset manner as a consequence of its dipole moment.⁴² Therefore, the quadrupole moment presented in Figure 13 suggests the occurrence of an offset aromatic stacking, which is consistent with the interaction proposed by Karacan *et al.*¹² and with the one observed in the plastoquinone Q_A site.^{43,44} The molecular descriptor CO+CM presents reasonable correlation with the descriptors QMbZZ and QMbYY (ca. ± 0.7), because they were calculated based on the CHELPG partial atomic charges of the aromatic ring of the benzyl group. The CO+CM descriptor is relevant for the model and indicates that the proposed aromatic stacking is more sensitive to the electronic density of the *ortho* and *meta* carbon atoms of the aromatic ring.

The molecular descriptor QMafYY is also a quadrupole moment, but calculated using only the arylidene and the furanone groups. This descriptor is illustrated in Figure 14 for the **m01** analogue, where blue stands for positive and red for negative values of the isosurface. It is negatively correlated to the PI and, consequently, a higher inhibitory activity corresponds to a more negative value of QMafYY. As a result, there is an approximation of the red portion of the electric potential isosurface to the Y axis. This descriptor is also reasonably well correlated to both charge of the atom at position R₃ and charge of the oxygen O13 (correlation coefficients of -0.8598 and -0.5139 , respectively). The substitution at position R₃ is the most relevant for the QMafYY descriptor, which influences the atomic charge of

the oxygen O13 and, consequently, the PI. For instance, an increase in the atomic charge of the atom in R_3 , such as in **m10** and **m16**, leads to a higher charge in oxygen O13 and an increase in PI. Several studies^{12,16,21} of herbicides targeting the D1 protein of photosystem II have shown the importance of the hydrogen bond interaction between these herbicides and the His215 residue for their photosynthetic inhibition activity. Moreover, a voltammetric study of rubrolide analogues showed the possibility of these compounds to act as hydrogen acceptor at photosystem I, being reduced in the process.⁴⁵ More recently, Teixeira *et al.*¹⁸ proposed a photoelectron transfer to the herbicide via His215, forming a radical anion. Based on these results and the QMafYY descriptor, it is reasonable to propose a similar photoelectron transfer to the nostoclides analogues via a hydrogen bond interaction with the oxygen O13, which is favored by an electron deficiency at the atom in position R_3 .

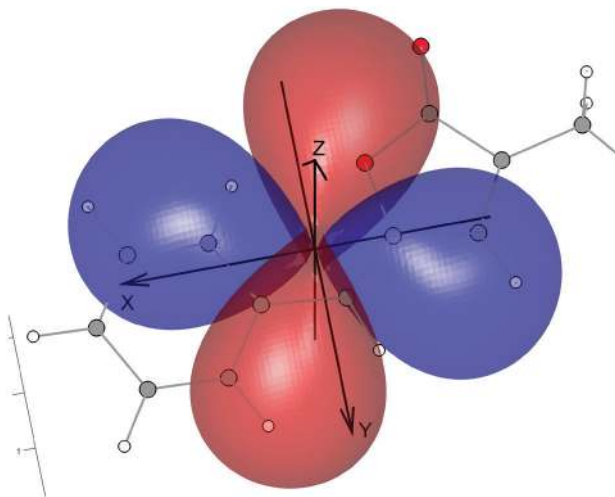


Figure 14. Electric potential isosurface of the quadrupole moment from the arylidene portion and furanone; blue for positive and red for negative values.

The WHO02 molecular descriptor indicates the electronic density of the HOMO located at carbon C2. However, the HOMO electron densities at the other atoms are also dependent on the WHO02 descriptor. Figure 15 shows the HOMO electronic density for molecules **m13** and **m16**, one of the least and one of the most active analogues, respectively; carbon C2 is highlighted. The electron density in this atom is greater for **m13** than for **m16**, as it is expected by the negative correlation between the WHO02 descriptor and PI. The change in electron density is observed in both arylidene and furanone rings, reaching the methylene of the benzyl. Although WHO02 have shown to be significant to the 2D QSAR model, it was not possible to attribute a conclusive interpretation for its concrete relation to the photosynthetic inhibition of the nostoclides.

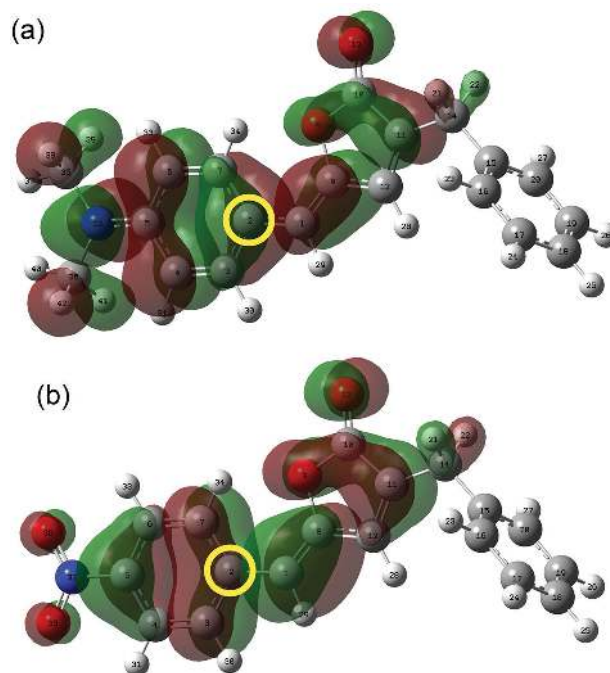


Figure 15. HOMO electronic density of molecules (a) **m13** and (b) **m16** with the carbon C2 highlighted.

Among the 3D descriptors, the most important to the 3D QSAR model was L1, a Lennard-Jones potential energy for the NH_3^+ probe located in the proximity of the aromatic ring of the arylidene portions (Figure 16). Its correlation with the PI is negative. On the other hand, the more negative the energy values of the van der Waals potential results in a more stable interaction. The L1 descriptor's probe is positioned above the aromatic ring of the arylidene portions, which is consistent with an aromatic stacking, known to have a contribution from the van der Waals potential.⁴⁶

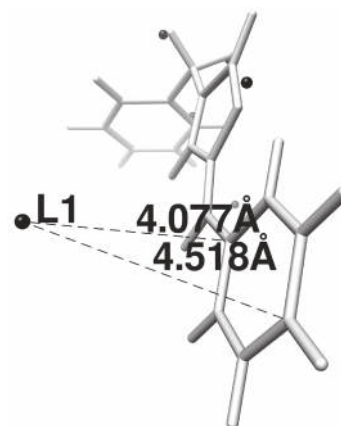


Figure 16. Illustration of the **m01** molecule and the position of L1 descriptor's probe.

The E5 descriptor has the second highest absolute regression coefficient, it is negatively correlated to PI and its values are negative. The NH_3^+ probe is located very close to

carbon C16 of the benzyl group, which is one of the *ortho* ones. This descriptor is possibly related to the QMbYY 2D descriptor, located in the same ring, because their correlation coefficient is -0.6164 and both have electrostatic nature. Their interpretation is also similar: a more negative value of E5 leads to more negative electrostatic density close to the probe, which results in a more positive value of the QMbYY. E5 might be related to aromatic stacking interactions between the nostoclode analogues and the plastoquinone Q_B site, as well as the QMbYY was theorized to be. Similarly, the descriptors E3 and E4 show some correlation to QMbZZ (correlation coefficients of -0.5964 for E3, and -0.5792 for E4), thus, could be related to the aromatic stacking of the benzyl's ring.

The E2 descriptor's probe is in the proximity of oxygen O9 and the carbonyl $C10=O13$ and its values are negative, which corresponds to a negative electrostatic density. It is positively correlated to the PI, a similar behavior compared to QMaYY molecular descriptor regarding positions R_3 and the oxygen O13. The E2 descriptor suggests a similar chemical interpretation as the QMaYY, which could be related to the photoelectron transfer to the nostoclode analogues via His215.

The probe of E1 descriptor is located between the R_3 and R_4 positions of the aromatic ring of the arylidene portions. This descriptor's values vary from negative to positive depending on the substituent, but it is negatively correlated with PI. Consequently, substituents with more negative electrostatic density at positions R_3 and R_4 (i.e., CF_3 , NO_2 and F) favor the photosynthetic inhibition property. Barbosa *et al.*¹⁵ observed that these substituents are also important for photosynthetic inhibition of rubrolide analogues. L2 is a Lennard-Jones potential descriptor located close to the benzyl group, whose atoms do not present significant variance of their positions. Therefore, it was not possible to make a consistent interpretation for this molecular descriptor.

Descriptors included in the hybrid QSAR model were selected among those used to build both the 2D and the 3D QSAR models. The selection was performed using PLS regressions and analyzing which combination would give rise to a QSAR model of good quality. The molecular descriptors CO+CM, QMbYY, and E3 were selected over the QMbZZ, E4 and E5, regarding the electrostatic interactions of the benzyl group. Also, the molecular descriptor QMaYY was selected over the E2 to take into account the role of the electrostatic density of the oxygens from the furanone ring. The descriptors E1, L1 and wHO02 presented complementary importance for the hybrid QSAR model, and thus, they were all kept in the model. The descriptor L2 was excluded. As can be seen

from Table 2, Figures 8 and 11, the statistics of this final model is somewhat better than the previous.

Conclusions

Experimental data of the percentage of inhibition of a set of 34 nostoclode analogues, published by Teixeira *et al.* in 2008, was used to build QSAR models based on 2D and 3D molecular descriptors. Although the 2D QSAR model presented higher errors in the leave-one-out cross-validation, the selected descriptors showed superior correlation coefficients with PI. Furthermore, its chemical interpretation gave some insight regarding the interactions between the analogues and the plastoquinone Q_B active site from photosystem II. On the other hand, the 3D QSAR model showed lower errors in the internal validations and some complementary information about the ligand-receptor interactions regarding the photosynthetic inhibitory activity of the studied compounds. The hybrid QSAR model, presenting the best statistics, highlighted the most important molecular descriptors for the PI. Analysis of the molecular descriptors used to build the QSAR models showed the major importance of electrostatic interactions for the activity of the nostoclode analogues. These interactions were observed, primarily, in the aromatic ring of the benzyl group, and associated with an aromatic stacking interaction due to the electrostatic quadrupole moment of the ring. For the studied nostocliodes, strong electron withdrawing substituents, such as CF_3 and NO_2 (present in the several of the most actives analogues, i.e., **m08-10** and **m15-16**), tend to increase the quadrupole moment of the aromatic ring of the benzyl group. Another important interaction proposed in this work was the photoelectron transfer to the nostoclode analogues via hydrogen bond between their carbonyl and the His215 residue of the D1 protein. Van der Waals interaction observed near the arylidene group presented correlation with the PI. The analyses of the chemical properties of the nostoclode analogues and their relation to the inhibition of the photosystem II presented important insights about the ligand-receptor interactions of these herbicides. The results obtained in this work are expected to be useful in future researches of herbicides targeting the photosystem II, especially for the further experimental development of novel nostoclode analogues.

Supplementary Information

Supplementary information regarding the selected molecular descriptors, the QSAR models and their validations is available free of charge at <http://jbc.sbq.org.br> as PDF file.

Acknowledgments

The authors acknowledge the Brazilian Governmental Agencies CNPq and FAPESP for financial support.

References

1. Teixeira, R. R.; Barbosa, L. C. A.; Forlani, G.; Piló-Veloso, D.; Carneiro, J. W. M.; *J. Agric. Food Chem.* **2008**, *56*, 2321.
2. Oerke, E. C.; *J. Agric. Sci.* **2006**, *144*, 31.
3. Timmons, F. L.; *Weed Sci.* **2005**, *53*, 748.
4. <https://edis.ifas.ufl.edu/pi219>, accessed in February 2018.
5. Neve, P.; Vila-Aiub, M.; Roux, F.; *New Phytol.* **2009**, *184*, 783.
6. <http://www.weedscience.org/Graphs/ChronologicalIncrease.aspx>, accessed in July 2018.
7. Duke, S. O.; *Pest Manage. Sci.* **2012**, *68*, 505.
8. Martins, P. A.; Barbosa, G.; Pasqualoto, K. F. M.; Ferreira, M. M. C.; *J. Chem. Inf. Model.* **2009**, *49*, 1428.
9. Todeschini, R.; Consonni, V.; Gramatica, P. In *Comprehensive Chemometrics*, vol. 4; Brown, S. D.; Tauler, R.; Walczak, B., eds.; Elsevier BV: Amsterdam, Netherlands, 2009, ch. 5, p. 129.
10. Huppatz, J. L.; Mcfadden, H. G.; *Z. Naturforsch., C: J. Biosci.* **1993**, *48*, 140.
11. Huppatz, J. L.; *Weed Sci.* **1996**, *44*, 743.
12. Karacan, M. S.; Yakan, Ç.; Yakan, M.; Karacan, N.; Zharmukhamedov, S. K.; Shitov, A.; Los, D. A.; Klimov, V. V.; Allakhverdiev, S. I.; *Biochim. Biophys. Acta* **2012**, *1817*, 1229.
13. Banjare, P.; Singh, J.; Roy, P. P.; *Beni-Suef Univ. J. Basic Appl. Sci.* **2017**, *6*, 219.
14. Teixeira, R. R.; Pinheiro, P. F.; Barbosa, L. C. A.; Carneiro, J. W. M.; Forlani, G.; *Pest Manage. Sci.* **2010**, *66*, 196.
15. Barbosa, L. C. A.; Maltha, C. R. A.; Lage, M. R.; Barcelos, R. C.; Donà, A.; Carneiro, J. W. M.; Forlani, G.; *J. Agric. Food Chem.* **2012**, *60*, 10555.
16. Funar-Timofei, S.; Borota, A.; Crisan, L.; *Mol. Divers.* **2017**, *21*, 437.
17. Xie, Y.; Peng, W.; Ding, F.; Liu, S. J.; Ma, H. J.; Liu, C. L.; *Pest Manage. Sci.* **2018**, *74*, 189.
18. Teixeira, R. R.; Barros, M. V. A.; Bressan, G. C.; Siqueira, R. P.; dos Santos, F. S.; Bertazzini, M.; Kiralj, R.; Ferreira, M. M. C.; Forlani, G.; *Pest Manage. Sci.* **2017**, *73*, 2360.
19. Takahashi, R.; Hasegawa, K.; Takano, A.; Noguchi, T.; *Biochemistry* **2010**, *49*, 5445.
20. Trebst, A.; *Z. Naturforsch., C: J. Biosci.* **1987**, *42*, 742.
21. Gleason, F. K.; *FEMS Microbiol. Lett.* **1990**, *68*, 77.
22. Mallipudi, L. R.; Gleason, F. K.; *Plant Sci.* **1989**, *60*, 149.
23. Teixeira, R. R.; Barbosa, L. C. A.; Carneiro, J. W. M.; Corrêa, R. S.; Ellena, J.; Doriguetto, A. C.; *J. Mol. Struct.* **2009**, *917*, 1.
24. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; *Gaussian 09, Revision D.01*; Gaussian Inc., Wallingford, CT, USA, 2009.
25. Becke, A. D.; *J. Chem. Phys.* **1993**, *98*, 5648.
26. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J.; *J. Phys. Chem.* **1994**, *98*, 11623.
27. Weigend, F.; Ahlrichs, R.; *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
28. Atkins, P.; Paula, J.; Friedman, R.; *Quanta, Matter, and Charge: A Molecular Approach to Physical Chemistry*; OUP: Oxford, UK, 2009.
29. Teófilo, R. F.; Martins, J. P. A.; Ferreira, M. M. C.; *J. Chemom.* **2008**, *23*, 32.
30. Martins, J. P. A.; Ferreira, M. M. C.; *Quim. Nova* **2013**, *36*, 554.
31. *Pirouette 3.11*; Infometrix Inc., Woodinville, WA, USA, 2003.
32. Barbosa, E. G.; Ferreira, M. M. C.; *Mol. Inf.* **2012**, *31*, 75.
33. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E.; *J. Comput. Chem.* **2004**, *25*, 1605.
34. Ferreira, M. M. C.; *J. Braz. Chem. Soc.* **2002**, *13*, 742.
35. *MATLAB, 8.5*; The MathWorks, Inc., Natick, MA, USA, 2015.
36. Tropsha, A.; Gramatica, P.; Gombar, V. K.; *QSAR Comb. Sci.* **2003**, *22*, 69.
37. Kiralj, R.; Ferreira, M. M. C.; *J. Braz. Chem. Soc.* **2009**, *20*, 770.
38. *ChemDraw Ultra 12.0*; PerkinElmer Informatics, Massachusetts, USA, 2016.
39. Dennington, R.; Keith, T. A.; Millam, J. M.; *GaussView*, version 5; Semichem Inc., Shawnee Mission, KS, USA, 2016.
40. Wold, S.; Eriksson, L.; Clementi, S. In *Chemometric Methods in Molecular Design*; Waterbeemd, H., ed.; Wiley-VCH: Weinheim, Germany, 1995, p. 309.
41. Hunter, C. A.; Sanders, J. K.; *J. Am. Chem. Soc.* **1990**, *112*, 5525.
42. Martinez, C. R.; Iverson, B. L.; *Chem. Sci.* **2012**, *3*, 2191.
43. Kern, J.; Renger, G.; *Photosynth. Res.* **2007**, *94*, 183.
44. Lambreva, M. D.; Russo, D.; Polticelli, F.; Scognamiglio, V.; Antonacci, A.; Zobnina, V.; Campi, G.; Rea, G.; *Curr. Protein Pept. Sci.* **2014**, *15*, 285.

45. Varejão, J. O. S.; Barbosa, L. C. A.; Maltha, C. R. A.; Lage, M. R.; Lanznaster, M.; Carneiro, J. W. M.; Forlani, G.; *Electrochim. Acta* **2014**, *120*, 334.

46. Mcgaughey, G. B.; Gagne, M.; Rappe, A. K.; *J. Biol. Chem.* **1998**, *273*, 15458.

Submitted: May 3, 2018

Published online: September 11, 2018