LIGAND-BASED PHARMACOPHORE MODEL AND QSAR STUDIES ON HERBICIDES TARGETING PHOTOSYSTEM II FROM CHLAMYDOMONAS REINHARDTII[ǂ]

Ana Borota, Simona Funar-Timofei, Luminita Crisan*

Department of Computational Chemistry, Institute of Chemistry of Romanian Academy, Timisoara, Mihai Viteazul Avenue, 24, 300223 Timisoara, Romania

lumi_crisan@acad-icht.tm.edu.ro

ABSTRACT

The resistance of weeds is a problem which can be overcome by finding new herbicides. For this purpose, beyond the experimental methods, *in silico* approaches can be helpful, as a starting point. In this regard, pharmacophore mapping and 3D-QSAR studies were carried out on several series of herbicide, already known to act on the Photosystem II (PS II) D1 protein. Using PHASE software, three pharmacophore features, H-bond acceptor (A), hydrophobic (H) and aromatic ring (R) were taken into account to be the best hypothesis. For this hypothesis an atom-based 3D-QSAR model was generated with statistically significant parameters (the correlation coefficient of regression (R^2) of 0.839, the standard error of estimates (SD) of 0.370, the Fisher test (F) of 53.7 for the training set, the external explained variance Q^2 = 0.640, the Pearson-R = 0.916 and Root Mean Square Error (RMSE) = 0.572, for the test set). This hypothesis, validated by the 3D atom-based QSAR approach, assures the selection of novel scaffolds of herbicide derivatives and can be used for the design of new chemical entities active on the PS II D1 protein.

Keywords: Pharmacophore, 3D-QSAR, herbicide, PS II D1

INTRODUCTION

The chemical compounds utilized to control undesired plants (weeds) are named herbicides and they can act by inducing the various inhibition mechanisms in the plants. Photosystem II (PS II)

[ǂ] Dedicated to the 150th anniversary of the Romanian Academy

is a protein complex which is found in plants [1] and contains around 20 polypeptide chains. Among them, two subunits (D1 and D2), own the important redox-active cofactors. [2] The herbicides targeting PS II, in general, inhibit photosynthesis by binding to the D1 subunit. These compounds compete with the endogenous quinone (Q_A and Q_B) ligands, thus, the electron transport from Q_A to Q_B is blocked, CO_2 fixation is stoped and the growth of plants is damaged. [3, 4]

Computational methods such as ligand-based pharmacophores and quantitative structure-activity relationship (QSAR) are widely used in the discovery of new chemicals for pharmaceutical and agrochemical fields.

Our aim is to develop a valid pharmacophore model based on different scaffolds of PS II D1 herbicide derivatives (pyrimidine, pyridine, cinnoline, triazine and quinine) which further can be used for screening molecular databases in order to find potential new herbicides and for the prediction of their activity.

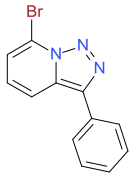
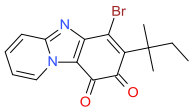
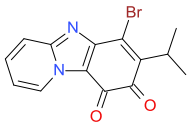
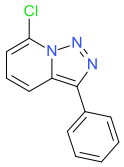
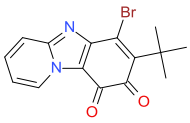
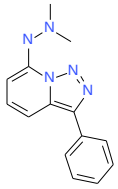
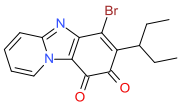
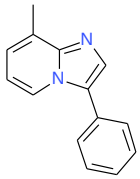
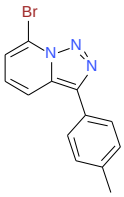
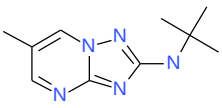
METHODS

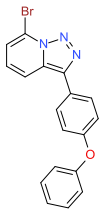
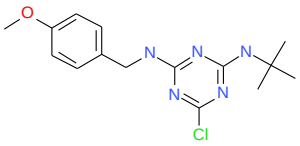
Data set selection and processing

The datasets consisting of 58 inhibitors of the D1 protein in photosystem II (PSII D1) from *Chlamydomonas reinhardtii* were collected from literature [5] and Pubchem database [6,7] AID1101260 and AID1101262. In case of ten compounds which show multiple experimental activities, we considered their average values. All structures were converted from smiles code into 3D structures, and ionization states and tautomers in the pH range of 6.2±0.3 were generated, using the LigPrep module [8] of Schrödinger suite [9]. The conformational space for each ligand was developed with the help of ConfGen module [10,11] using the default options. 217 compounds resulted after conformer generation and energy minimization based on the OPLS-2005 force field.

The pharmacophore hypotheses were generated using eight most active (with $pIC_{50}>7$) compounds, while the threshold for inactivity was set to 5 using the Phase module [12-14] of Schrödinger suite [9].

Table 1. The structure of the most active compounds (**1** to **8**), the unaligned ligands (**9** and **10**) and the less active compounds (**11** and **12**) and their herbicidal activity in logarithmic units (pIC50)

No	Structure	pIC50	No	Structure	pIC50
1	 7-bromo-3-phenyl-triazolo[1,5-a]pyridine	7.620	7	 6-bromo-7-(1,1-dimethylpropyl)pyrido[1,2-a]benzimidazole-8,9-dione	7.050
2	 6-bromo-7-isopropyl-pyrido[1,2-a]benzimidazole-8,9-dione	7.510	8	 7-chloro-3-phenyl-triazolo[1,5-a]pyridine	7.022
3	 bromo-7-tert-butyl-pyrido[1,2-a]benzimidazole-8,9-dione	7.470	9	 1,1-dimethyl-2-(3-phenyltriazolo[1,5-a]pyridin-7-yl)hydrazine	6.328
4	 6-bromo-7-(1-ethylpropyl)pyrido[1,2-a]benzimidazole-8,9-dione	7.190	10	 8-methyl-3-phenyl-imidazo[1,2-a]pyridine	6.215
5	 7-bromo-3-(p-tolyl)triazolo[1,5-a]pyridine	7.187	11	 N-tert-butyl-6-methyl-[1,2,4]triazolo[1,5-a]pyrimidin-2-amine	4.208

6	 7-bromo-3-(4-phenoxyphenyl)triazolo[1,5-a]pyridine	7.071	12	 N2-tert-butyl-6-chloro-N4-[(4-methoxyphenyl)methyl]-1,3,5-triazine-2,4-diamine	4.237

Pharmacophore modeling and validation

The "Develop Pharmacophore Model" module of Phase software [12-14] implemented in the Schrödinger suite was used in order to generate all possible pharmacophore hypothesis using four PLS factors. The number of PLS factors was increased, but the model statistics or predictive ability did not improve.

The pharmacophore validation was carried out by atom-based 3D-QSAR regression including both internal and external validation. The training set includes 80% randomly selected molecules, whereas the remaining 20% were denominated to validate the model (test set). The external predictive ability for the test set prediction using Pearson-R was considered and the models which have values greater than 0.6 were selected.

Taking into account this statistical parameter but also high value of Q^2 test (correlation coefficient of prediction for the test set) and R^2 training (correlation coefficient for the training set) we selected the best QSAR model.

The statistical parameters were calculated based on the following equations (see Phase user manual):

For the training set:

(i) Standard deviation of regression

$$SD = \sqrt{sse / df_2} \qquad (1)$$

where: sse is sum of squares errors and df_2 is the degree of freedom of data.

$$sse = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \ ;$$

$\hat{y}_i$ = predicted activity for the training set molecule *i*;

y_i = observed activity for the training set molecule *i;*

$df_2 = n - m - 2$ (n – the number of molecule in the training set, and m – the number of PLS factors in the models)

(ii) The coefficient of determination

$$R^2 = 1 - \frac{\sigma_{err}^2}{\sigma_y^2} \qquad (2)$$

where: $\sigma_{err}^2 = sse/n$, σ_{err} = variance in errors;

$\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$, σ_y = variance in observed activities;

$\bar{y}$ = mean observed activity

(iii) Fischer test

$$F = \frac{ssy/df_1}{sse/df_2} \qquad (3)$$

where: $ssy = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$;

ssy - the variance in model;

$df_1 = m + 1$, df_1 - degrees of freedom in model

(iv) P - statistical significance; probability that correlation could occur by chance.

$$P = B(df_1, df_2, \frac{df_2}{df_2 + Fdf_1}) \qquad (4)$$

For the test set:

(v) Root-mean-square error

$$RMSE = \sqrt{\frac{1}{n_T}\sum_{j \in T}(\hat{y}_j - y_j)^2} \qquad (5)$$

where: y_j – observed activity for molecule $j \in T$;

$\hat{y}_j$ - predicted activity for molecule $j \in T$;

T - test set of molecules;

n_T - number of molecules in T

(vi) Coefficient for the predicted activities, analogous to R^2, but based on the test set predictions

$$Q^2 = R^2(T) \quad (6)$$

(vii) Pearson-R correlation coefficient

$$r = \frac{\sum_{j \in T} \left(y_j - \bar{y}_T\right)\left(y_j - \hat{\bar{y}}_T\right)}{\sqrt{\sum_{j \in T} \left(y_j - \bar{y}_T\right)^2 \left(y_j - \hat{\bar{y}}\right)^2}} \quad (7)$$

where: $\bar{y}_T$ represents the mean observed activity of the test compounds;

$\hat{\bar{y}}_T$ - mean calculated activity of the test compounds.

RESULTS AND DISCUSSIONS

Ten pharmacophore (Table 2) hypotheses based on different scaffolds of PSII D1 herbicide derivatives were generated using three minimum sites: H-bond acceptor (A), hydrophobic (H) and aromatic ring (R). The selected hypothesis AHR.7 (Figure 1) was used for the generation of the 3D QSAR model using four PLS factors. This model was built using the PHASE descriptors as independent variables and the herbicidal activity values (expressed as pIC50 values), as dependent variables. Two unaligned ligands (compound no **9** and no **10**) of AHR.7 hypothesis, were excluded as outlier, see Table 1.

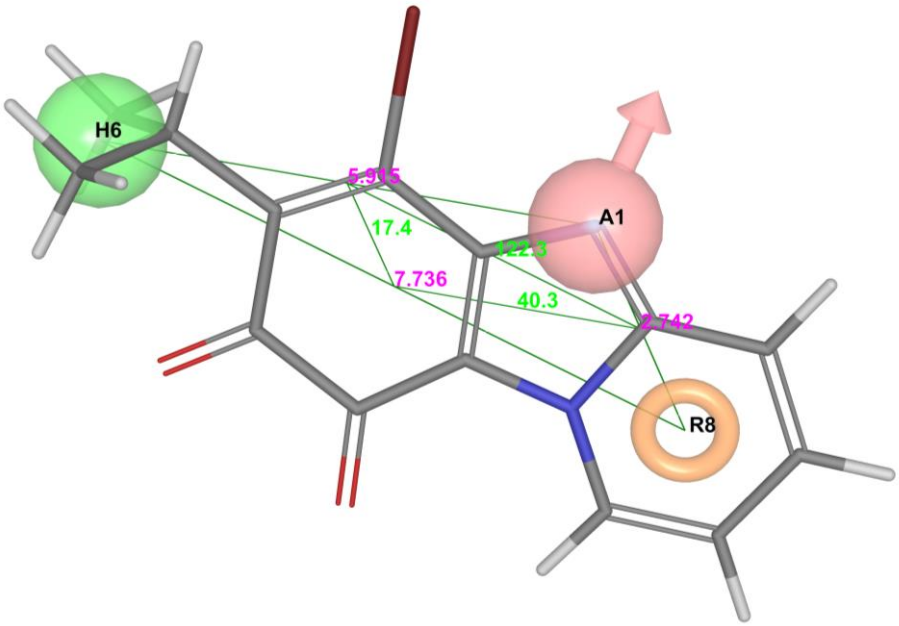


Figure 1. The pharmacophore hypothesis AHR.7 (acceptor (A1, pink), hydrophobic (H6, green), ring (R8, orange)) aligned to the compound **2** with best Fitness score = 3

Table 2. The statistical parameters obtained for the QSAR models

Model	#	SD*	R^{2*}	F*	P*	Stability$	RMSE#	$Q^{2\#}$	r#
AHR.3	1	0.596	0.534	37.900	6.14E-07	0.817	0.627	0.455	0.710
	2	0.452	0.740	45.500	4.37E-10	0.596	0.511	0.639	0.814
	3	0.361	0.840	54.200	1.97E-12	0.364	0.510	0.640	0.802
	4	0.241	0.931	100.500	6.30E-17	0.237	0.470	0.694	0.834
AHR.4	1	0.598	0.531	37.400	6.90E-07	0.809	0.578	0.538	0.778
	2	0.456	0.736	44.500	5.72E-10	0.602	0.559	0.568	0.757
	3	0.372	0.830	50.300	5.13E-12	0.394	0.645	0.424	0.658
	4	0.261	0.919	84.700	6.63E-16	0.225	0.629	0.453	0.677
AHR.2	1	0.644	0.442	26.900	9.91E-06	0.926	0.506	0.646	0.876
	2	0.386	0.806	68.500	1.78E-12	0.672	0.489	0.669	0.847
	3	0.325	0.866	69.200	4.46E-14	0.618	0.594	0.512	0.728
	4	0.258	0.919	87.500	1.98E-16	0.556	0.570	0.550	0.756
AHR.12	1	0.633	0.496	43.300	4.70E-08	0.851	0.652	0.532	0.767
	2	0.511	0.678	45.300	2.61E-11	0.669	0.590	0.617	0.817
	3	0.465	0.740	39.900	2.35E-12	0.571	0.644	0.544	0.779
	4	0.376	0.834	51.500	1.88E-15	0.414	0.636	0.555	0.808
AHR.14	1	0.545	0.610	51.700	3.05E-08	0.846	0.495	0.694	0.842
	2	0.431	0.764	51.800	9.16E-11	0.743	0.629	0.507	0.722
	3	0.341	0.857	61.700	3.56E-13	0.616	0.567	0.599	0.786
	4	0.297	0.895	64.000	2.95E-14	0.515	0.552	0.620	0.795

AHR.7	1	0.659	0.454	36.500	2.93E-07	0.760	0.657	0.525	0.742
	2	0.507	0.683	46.400	1.82E-11	0.606	0.484	0.743	0.889
	3	0.428	0.780	49.700	7.18E-14	0.572	0.574	0.637	0.860
	4	**0.370**	**0.840**	**53.700**	**9.03E-16**	**0.523**	**0.572**	**0.640**	**0.916**
AHR.6	1	0.681	0.417	31.500	1.27E-06	0.888	0.556	0.627	0.897
	2	0.505	0.686	47.000	1.52E-11	0.693	0.546	0.640	0.855
	3	0.433	0.775	48.200	1.16E-13	0.663	0.546	0.640	0.835
	4	0.337	0.867	66.800	2.07E-17	0.533	0.650	0.490	0.717
AHR.11	1	0.597	0.513	36.800	6.26E-07	0.872	0.642	0.430	0.661
	2	0.462	0.716	42.900	5.08E-10	0.765	0.581	0.532	0.734
	3	0.348	0.844	59.300	2.19E-13	0.577	0.700	0.322	0.605
	4	0.276	0.905	75.900	7.26E-16	0.419	0.723	0.276	0.552
AHR.9	1	0.634	0.449	28.500	5.71E-06	0.942	0.715	0.362	0.637
	2	0.480	0.694	38.500	1.86E-09	0.907	0.547	0.627	0.826
	3	0.389	0.805	45.400	8.29E-12	0.888	0.717	0.360	0.657
	4	0.264	0.913	84.000	1.67E-16	0.745	0.662	0.453	0.718
AHR.5	1	0.752	0.288	17.800	0.000122	0.976	0.731	0.355	0.622
	2	0.585	0.580	29.600	8.16E-09	0.873	0.512	0.684	0.838
	3	0.487	0.715	35.100	1.65E-11	0.786	0.642	0.503	0.711
	4	0.389	0.822	47.400	7.47E-15	0.612	0.643	0.501	0.712

Number of factors in the partial least squares regression model; SD - standard deviation of the regression; R^2 - the coefficient of determination; F - the ratio of the model variance to the observed activity variance; P - the significance level of variance ratio; Stability – the stability of the model predictions; RMSE – the root-mean-square error in the test set predictions; Q^2 - value for the predicted activities, analogous to R^2, but based on the test set predictions; r (Pearson-R) - value for the correlation between the predicted and observed activity for the test set (see equation 7); *for the training set; $^{\$}$ for the entire data set; $^{\#}$ for the test set.

The value of correlation coefficient for the training set (R^2) is higher than 0.8 and indicate a good correlation between the independent *versus* dependent variables. The predictive capacity of the QSAR model is satisfactory, from the values of the correlation coefficient for the test set (Q2 > 0.6) and of the Pearson's R (> 0.9). The statistical results for the test set proved that QSAR model is stable and predictive. The plot of observed versus predicted herbicidal activities for the training and test sets achieved for the 3D QSAR model of AHR.7 hypothesis is represented in Figure 2.

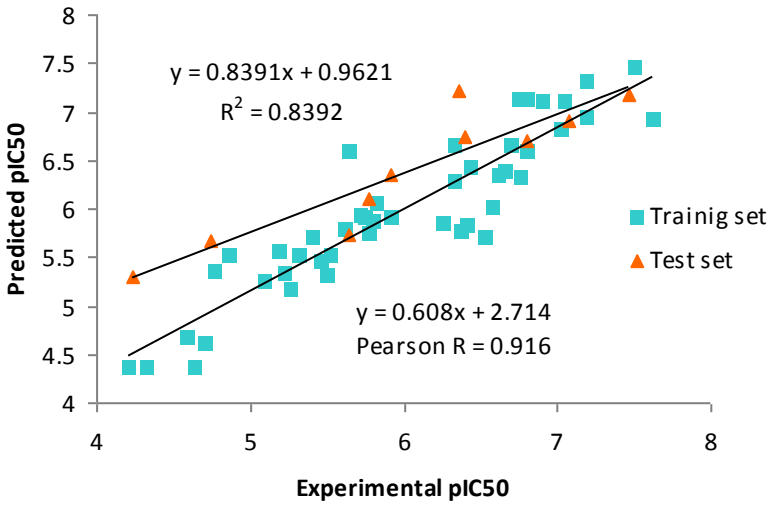


Figure 2. The plot of observed versus predicted herbicidal activities for the model with the pharmacophore AHR.7 hypothesis

Further understanding of the inhibitory activity of the PSII D1 protein can be achieved by mapping the 3D QSAR model over the ligands from the dataset series [5, 7]. A graphical representation of the significant favourable and unfavourable features for the herbicidal activity of the compounds that resulted when the QSAR model is applied is presented in Figures 3 to 6. In these pictures, the blue cubes show favorable regions, while red cubes indicate unfavorable regions for the herbicidal activity, regarding to following combined features: hydrogen bond donor, hydrophibic/ non-polar, electron-withdrawing and positive ionic.

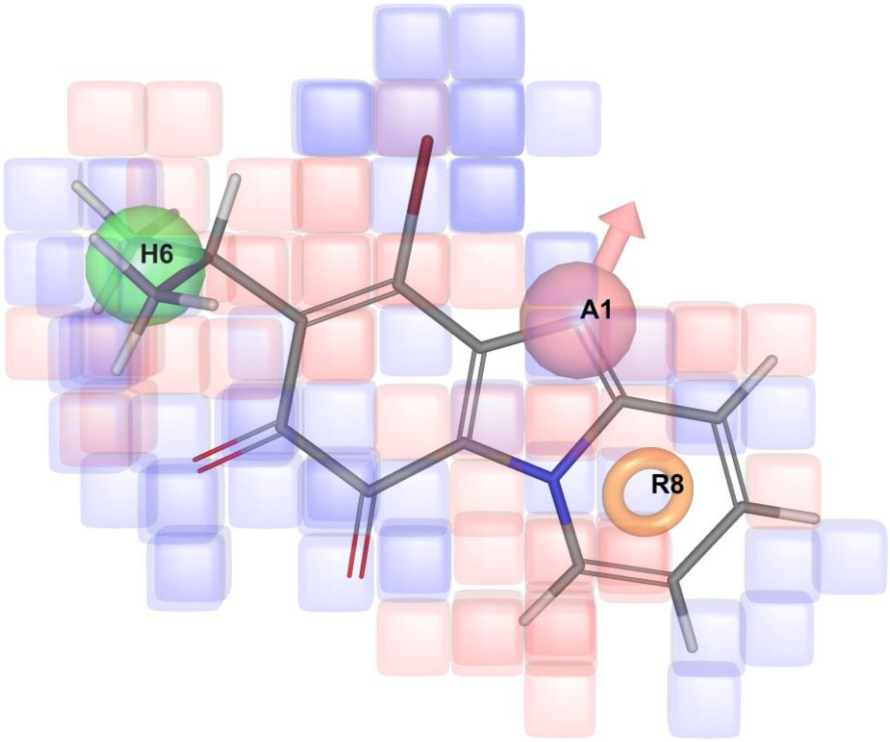


Figure 3. The QSAR model visualized in the context of the best aligned compound (no **2**) with AHR.7

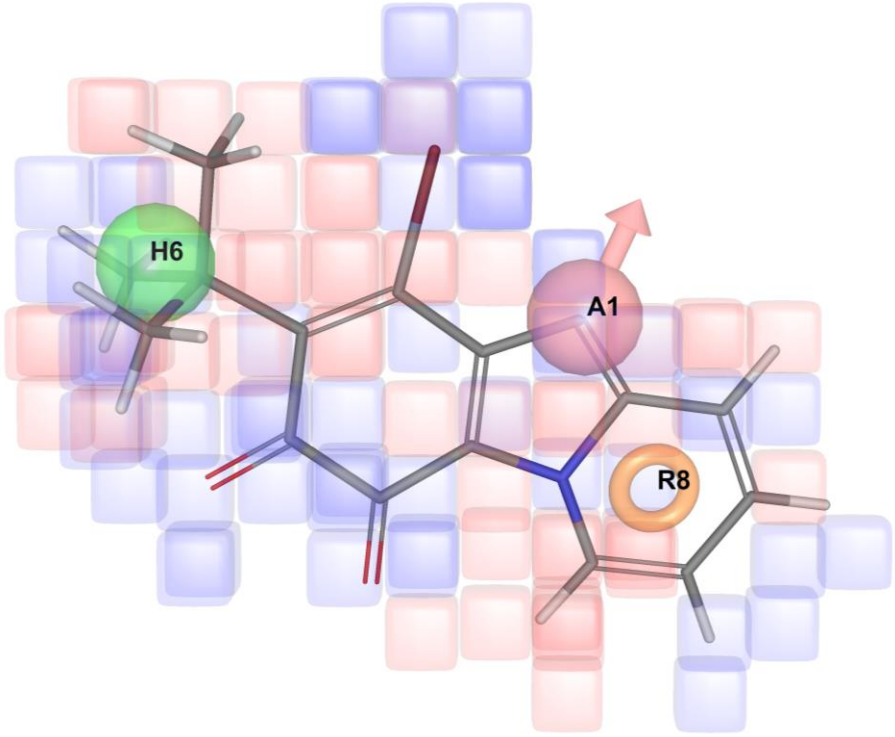


Figure 4. The QSAR model visualized in the context of the most active compound (no **3)** of the test set

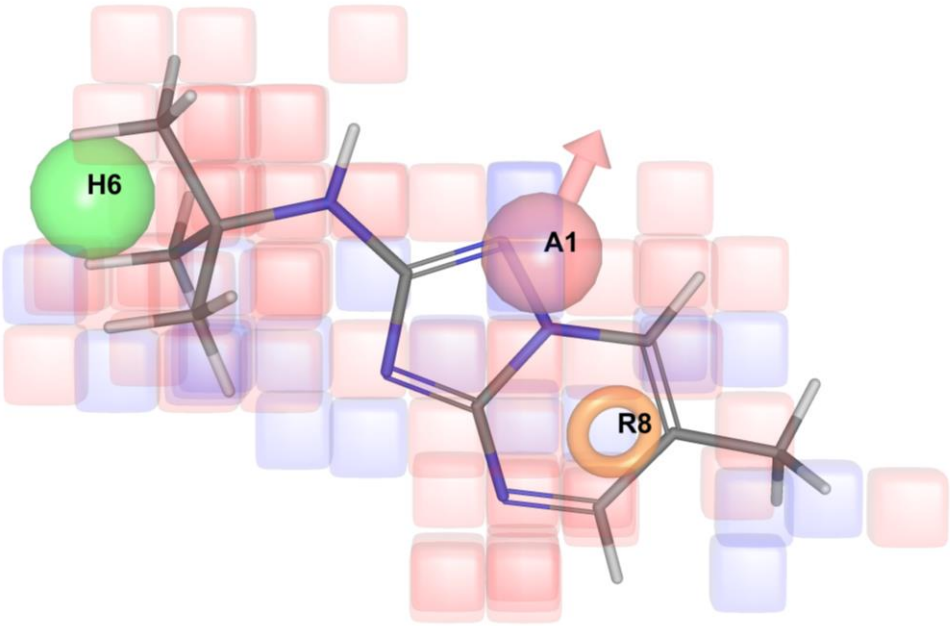


Figure 5. The QSAR model visualized in the context of the less active compound (no **11**) of the training set

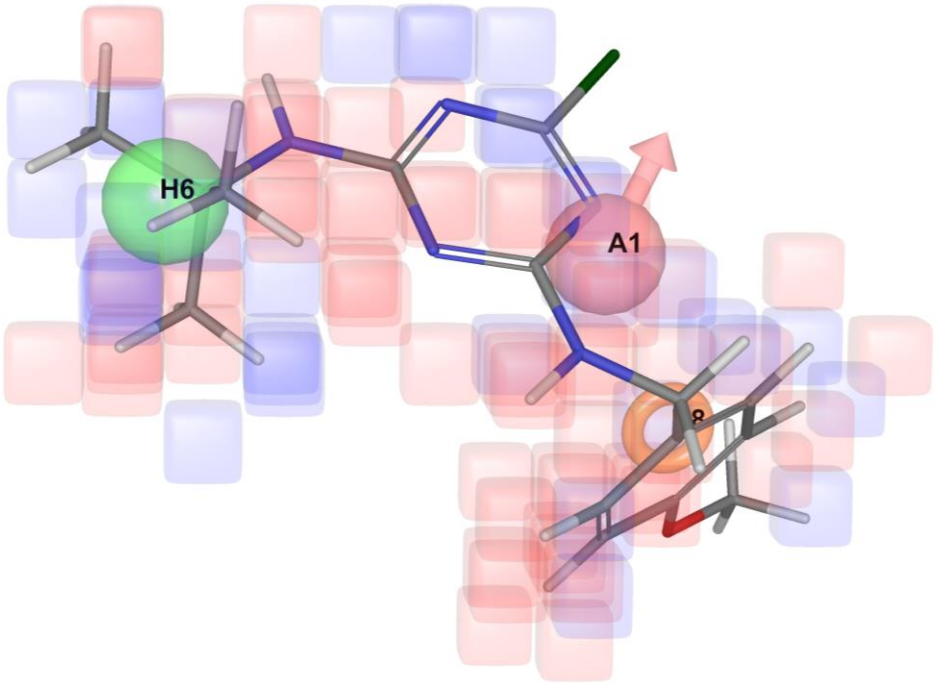


Figure 6. The QSAR model visualized in the context of the less active compound (no **12**) of the test set

In Figures 3 and 4 the active herbicide derivatives presents a superior number of blue cubes and a lower number of red cubes, in comparison with the less active ligands from Figures 5 and 6.

CONCLUSIONS

Pharmacophore-based 3D-QSAR study of PSII D1 inhibitors is carried out in order to explain the structural features of some herbicide derivatives (pyrimidine, pyridine, cinnoline, triazine and quinine) required for their inhibitory activity. The selected 3D-QSAR model indicates a significant correlation and a good predictive capacity. One hydrogen bond acceptors (A), one lipophilic/hydrophobic group (H) and one aromatic ring (R), as pharmacophore features, are important for the PSII D1 herbicidal activity. The best hypothesis AHR.7, in this study, is characterized by the best values of the R^2 regression coefficient (0.839) and the highest values for the Pearson-R coefficient (0.916). In future studies this pharmacophore model will be used for screening molecular databases in order to find potential new herbicides.

Acknowledgements

This project was financially supported by Project 1.1 of the Institute of Chemistry of the Romanian Academy. The authors thank Dr. Ramona Curpăn (Institute of Chemistry Timisoara of Roumanian Academy), for providing access to Schrödinger software acquired through the PN-II-RU-TE-2014-4-422 projects funded by CNCS-UEFISCDI Romania.

REFERENCES

1. J. Barber, Russian in Biokhimiya, **79** (2014) 248-262.

2. M.D. Lambreva, D. Russo, F. Polticelli, V. Scognamiglio, A. Antonacci, V. Zobnina, G. Campi, G. Rea, Curr. Protein Pept. Sci., **15 (**2014) 285-295.

3. http://herbicidesymptoms.ipm.ucanr.edu/MOA/Photosystem_II_Inhibitors (accessed August 2016).

4. M. Broser, C. Glöckner, A. Gabdulkhakov, A. Guskov, J. Buchta, J. Kern, F. Müh, H. Dau, W. Saenger, A. Zouni, J. Biol. Chem. **286** (2011) 15964–15972.

5. U. Egner, K.P. Gerbling, G.A. Hoyer, G. Kriiger, P. Wegnerb, Pestic. Sci., **47** (1996) 145–158.

6. Y.Wang, J. Xiao, T.O. Suzek, J. Zhang, J. Wang, Z. Zhou, L. Han, K. Karapetyan, S. Dracheva, B.A. Shoemaker, E. Bolton, A. Gindulyte, S.H. Bryant, Nucleic Acids Res., **40** (2012) D400-412.

7. PubChem BioAssay. https://pubchem.ncbi.nlm.nih.gov/ (accessed June 2016).

8. Schrödinger Release 2016-3**:** LigPrep, Schrödinger, LLC, New York, NY, 2016.

9. Small-Molecule Drug Discovery Suite 2016-3**,** Schrödinger, LLC, New York, NY, 2016.

10. K.S. Watts, P. Dalal, R.B. Murphy, W. Sherman, R.A. Friesner, J.C. Shelley, J.Chem. Inf. Model., **50** (2010) 534-546.

11. Schrödinger Release 2016-3**:** ConfGen, Schrödinger, LLC, New York, NY, 2016.

12. S.L. Dixon, A.M. Smondyrev, E.H. Knoll, S.N. Rao, D.E. Shaw, R.A. Friesner, J. Comput. Aided. Mol. Des. **20** (2006) 647-671.

13. S.L. Dixon, A.M. Smondyrev, S.N. Rao, Chem. Biol. Drug. Des. **67**(2006) 370-372.

14. Schrödinger Release 2016-3**:** Phase, Schrödinger, LLC, New York, NY, 2016.