 doi:10.20944/preprints201811.0246.v1

Article

Quantitative Structure-Activity Analysis of Triazines Immune Recognition Based on Immunoassay Data for polyclonal and Monoclonal Antibodies

Andrey A. Buglak[1,2*], Anatoly V. Zherdev[1], Hong-Tao Lei[3], Boris B. Dzantiev[1]

[1] A. N. Bach Institute of Biochemistry, Research Center of Biotechnology of the Russian Academy of Sciences, 33 Leninsky Prospect, Moscow 119071, Russia

[2] St. Petersburg State University, 7/9 Universitetskaya emb., St. Petersburg 199034 Russia

[3] Guangdong Provincial Key Laboratory of Food Quality and Safety, South China Agricultural University Guangzhou 510642, China

* Correspondence: andreybuglak@gmail.com; Tel.: +7 (495) 954-27-32

Abstract: A common task in the immunodetection of structurally close compounds is to analyze the selectivity of immune recognition: it is required to understand the regularities of immune recognition and to elucidate the basic structural elements which provide it. Triazines are compounds of particular interest for such a research due to their high variability and the necessity of their monitoring to provide safety of agricultural products and foodstuffs. We evaluated the binding of 20 triazines with polyclonal (pAb) and monoclonal (mAb) antibodies obtained using atrazine as the immunogenic hapten. A total of > 3000 descriptors was used in QSAR analysis of binding activities (pIC50). Comparison of the two enzyme immunoassay systems showed that the system with pAb is much easier to describe using 2D QSAR methodology, while the system with mAb can be described using the 3D QSAR COMFA. Thus, for the 3D QSAR model of the polyclonal antibodies, the main statistical parameter q^2 ('leave-many-out') is equal 0.498, and for monoclonal antibodies q^2 is equal 0.566. Obviously, in the case of pAb, we deal with several targets, while in the case of mAb the target is one, and therefore it is easier to describe it using specific fields of molecular interactions distributed in space.

Keywords: Monoclonal antibodies; Polyclonal antibodies; Triazines; Enzyme immunoassay; Quantitative structure-activity relationship analysis; 3D-QSAR; Atrazine.

1. Introduction

Triazines are herbicides which are widely used in agriculture and may accumulate in soil, as well as in food products [1]. Atrazines are bound in soil to solid as well as to dissolved fractions of humic and fulvic acids which leads not only to their accumulation in soil but also to contamination of surface and ground waters [2] since they are water soluble.

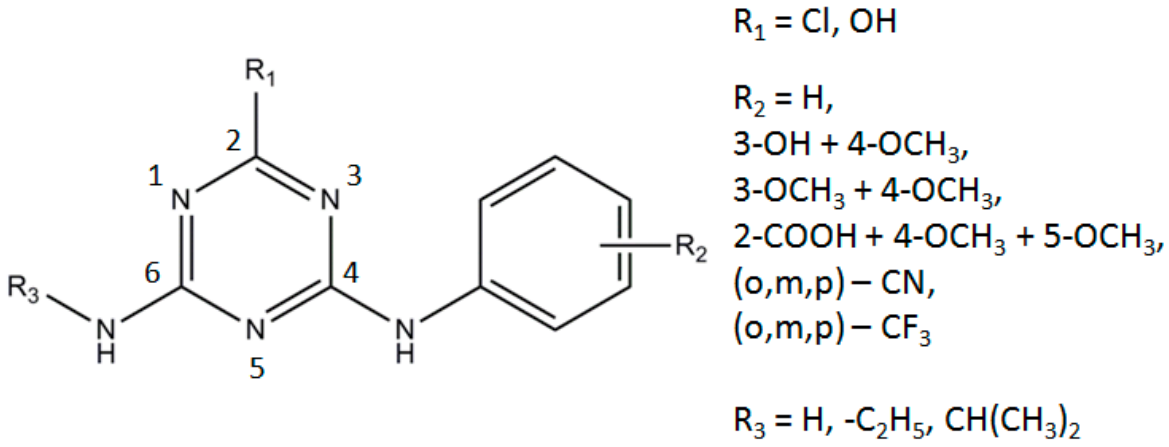
Triazines may undergo chemical transformations of both biotic and abiotic nature. The parent compound atrazine is subjected to oxidation of the alkyl substituents, oxidative dealkylation, hydroxylation, and, also to ring cleavage [3-5]. The derivatives of atrazine may be toxic to a greater or lesser extent. Triazines may induce different physiological disorders of humans and animals like immunity suppression [6] and birth defects [7]. Thus, contamination of the environment by

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 November 2018 doi:10.20944/preprints201811.0246.v1

triazine herbicides is a significant risk factor. For this reason it is necessary to monitor the triazine levels in soil, water, agricultural products and foods.

The methods of triazines detection like NMR, HPLC, mass-spectrometry and others are time and labour-consuming. The immunoassay determination of triazines is much easier, less expensive, highly sensitive and rapid. Due to its selectivity and simplicity the immunoassay has been used to detect a wide variety of environmental pollutants including triazines [8-10]. However, siginificant variety of triazines and their derivatives causes necessity in detailed analysis of regulations of their immune recognition.

Quantitative structure-activity relationship (QSAR) is widely used studying the immune recognition of diffferent classes of toxic food contaminants, veterinary drugs, including pesticides, etc [11-14]. The work by Yuan and co-authors reported about an immunoassay analysis of triazines, yet on a small set of 11 compounds [9]. QSAR is used for the analysis of immunoassays on quinolones and fluoroquinolones [15-19], organophosphorus pesticides [20-21], phenylurea herbicides [22], sulfonamides [23]. In this study we used molecular modeling to study triazine recognition by monoclonal and polyclonal antibodies. Twenty triazines along with 2D and 3D-QSAR methodology were used to study the relationship between antigen and antibody. Experimental data of microplate immunoenzyme assays with broad specificity for triazines was used from a classical work by A. Dankwardt and co-authors [24]. Comparative immunoassay of polyclonal and monoclonal antibodies used 4-arylamino-6-amino-1,3,5-triazines (Scheme 1). Antibodies were grown using atrazine as an immunizing hapten.



Scheme 1. Molecular structure of arylamino-s-triazines (based on Danckwardt et al., 1996).

The antibody-binding activity of triazines from Table 1, presented in the logarithmic form (pIC50), was used as the dependent variable y during QSAR calculations. In Table 1 S2 stands for polyclonal sheep antibodies, while K4E7 stands for monoclonal antibodies.

 doi:10.20944/preprints201811.0246.v1

Table 1. Cross-reactivity (CR) and concentration of 50% inhibition for the interaction of triazines with polyclonal (S2) and monoclonal antibodies (K4E7), according to Danckwardt, et al., 1996.

№	Compound	CR, %		IC50, nmol l^{-1}	
		S2	4E7	S2	K4E7
1	2-chloro-4-isopropylamino-6-ethylamino-1,3,5-triazine (atrazine)	100	100	0.93	0.37
2	2-chloro-4-isopropylamino-6-amino-1,3,5-triazine (deethylatrazine)	15	18	6.0	2.1
3	2-chloro-4-ethylamino-6-amino-1,3,5-triazine (deisopropylatrazine)	0.4	1.8	232	20.6
4	2-hydroxy-4-isopropylamino-6-ethylamino-1,3,5-triazine (hydroxyatrazine)	6	0.1	15.3	371
5	2-chloro-4-anilino-6-isopropylamino-1,3,5-triazine	106	140	0.88	0.27
6	2-chloro-4-(3′,4′-dimethoxyanilino)-6-isopropylamino-1,3,5-triazine	126	140	0.74	0.27
7	2-chloro-4-(3′,4′-dimethoxybenzylamino)-6-isopropylamino-1,3,5-triazine	109	148	0.84	0.25
8	2-chloro-4-(3′-hydroxy-4′-methoxyanilino)-6-isopropylamino-1,3,5-triazine	74	79	1.25	0.47
9	2-chloro-4-(2′-carboxy-4′,5′-dimethoxyanilino)-6-isopropylamino-1,3,5-triazine	9	<0.1	10.3	>500
10	2-chloro-4-(2′-nitrilanilino)-6-isopropylamino-1,3,5-triazine	134	70	0.70	0.52
11	2-chloro-4-(3′-nitrilanilino)-6-isopropylamino-1,3,5-triazine	134	108	0.70	0.34
12	2-chloro-4-(4′-nitrilanilino)-6-isopropylamino-1,3,5-triazine	101	107	0.93	0.35
13	2-chloro-4-(2′-triflourmethylanilino)-6-isopropylamino-1,3,5-triazine	91	81	1.02	0.46
14	2-chloro-4-(3′-triflourmethylanilino)-6-isopropylamino-1,3,5-triazine	133	75	0.70	0.46
15	2-chloro-4-(4′-triflourmethylanilino)-6-isopropylamino-1,3,5-triazine	92	129	1.02	0.49
16	2-chloro-4-(3′,4′-dimethoxyanilino)-6-ethylamino-1,3,5-triazine	5	10	18.6	3.68
17	2-hydroxy-4-anilino-6-isopropylamino-1,3,5-triazine	0.4	0.2	232	186
18	2-hydroxy-4-(3′,4′-dimethoxyanilino)-6-isopropylamino-1,3,5-triazine	0.4	0.2	232	186
19	2-chloro-4-anilino-6-amino-1,3,5-triazine	0.1	<0.1	>500	>500
20	2-chloro-4-(3′,4′-dimethoxyanilino)-6-amino-1,3,5-triazine	<0.1	<0.1	>500	>500

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 November 2018 doi:10.20944/preprints201811.0246.v1

2. Results

2.1. 2D QSAR

2.1.1. S2 system analysis (polyclonal antibodies)

For S2 system a series of multiple linear regression models was obtained:

Model 1

*pIC50 = 23.68 + 0.985 * ES_Count_sssCH − 64.66 * Mulliken_Charge_C4 + 0.056 * Solvation_E*

Model 2

*pIC50 = 12.43 + 4.393 * ES_Sum_sssCH − 107.3 * Jurs_FPSA_3*

Model 3

*pIC50 = 11.92 − 0.910 * CHI_2 + 1.932 * CHI_V_2 − 100.3 * Jurs_FPSA_3*

Model 4

*pIC50 = 8.807 + 5.366 * ES_Sum_sssCH + 0.828 * IC2 – 104.0 * Jurs_FPSA_3*

Model 5

*pIC50 = 6.302 + 6.182 * ES_Sum_sssCH + 0.946 * IC2 + 0.062 * Solvation_E*

It is believed that QSAR is predictive if the following conditions are satisfied: $R^2 > 0.6$, $q^2 > 0.5$ and $pred\_R^2 > 0.5$. Table 2 shows that all the presented models are statistically reliable and possess predictive ability. Model 5 has the highest statistical parameters among the five models.

Table 2. Statistical parameters of the models for S2 system.

№	Statistical parameter	Model 1	Model 2	Model 3	Model 4	Model 5
1	R^2	0.941	0.890	0.937	0.937	**0.945**
2	R^2 *adj*	0.926	0.873	0.921	0.921	**0.932**
3	q^2	0.893	0.843	0.898	0.873	**0.900**
4	$pred\_R^2$	0.913	0.924	0.870	0.920	**0.942**
5	*Max error*	0.567	0.904	0.746	0.549	**0.454**
6	*RMS error*	0.295	0.387	0.305	0.306	**0.284**
7	*LOF*	0.383	**0.329**	0.411	0.412	0.355

Let us consider the descriptors involved in Model 5. *Solvation_E* (relative contribution, α = 66.8%) is the solvation energy calculated with AM1 method. Compound **18** has the highest solvation energy value (-20.27 kJ mol^{-1}), while compound **13** has the lowest one (-71.22 kJ mol^{-1}) (Table S1). The lower the solvation energy the higher the solubility. However, the energy of solvation has a negative value, and therefore compounds that are less soluble in water (i.e., more hydrophobic) have higher pIC50 values.

ES_Sum_sssCH (26.1%) is the electrotopological index of methanthriyl groups [25]. The *ES_Sum_sssCH* value is directly proportional to the activity of the molecule (**1**, **2** and similar compounds). *ES_Sum_sssCH* is equal zero for the compounds without >CH-groups. Obviously, the presence of the *ES_Sum_sssCH* descriptor in Model 5 indicates the participation of >CH-groups of triazine compounds in van der Waals interactions with hydrophobic regions of antibodies.

IC2 (7.1%) is the information content index (neighborhood symmetry of 2-order). This

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 November 2018 doi:10.20944/preprints201811.0246.v1

descriptor makes possible to estimate the degree of heterogeneity of the molecular structure. Compound **1** has the lowest value of *IC2* (3.431), and compound **9** has, on the contrary, the highest value of this descriptor (4.426). *IC2* can be considered as a measure of the complexity of the triazine topology. *IC2* has a minor contribution to the model (α = 7.1%)) and, apparently, acts as a correction term in the linear regression equation.

Experimental and predicted pIC50 values for triazine interactions with polyclonal antibodies (S2) based on Model 5 are presented on Figure 1.

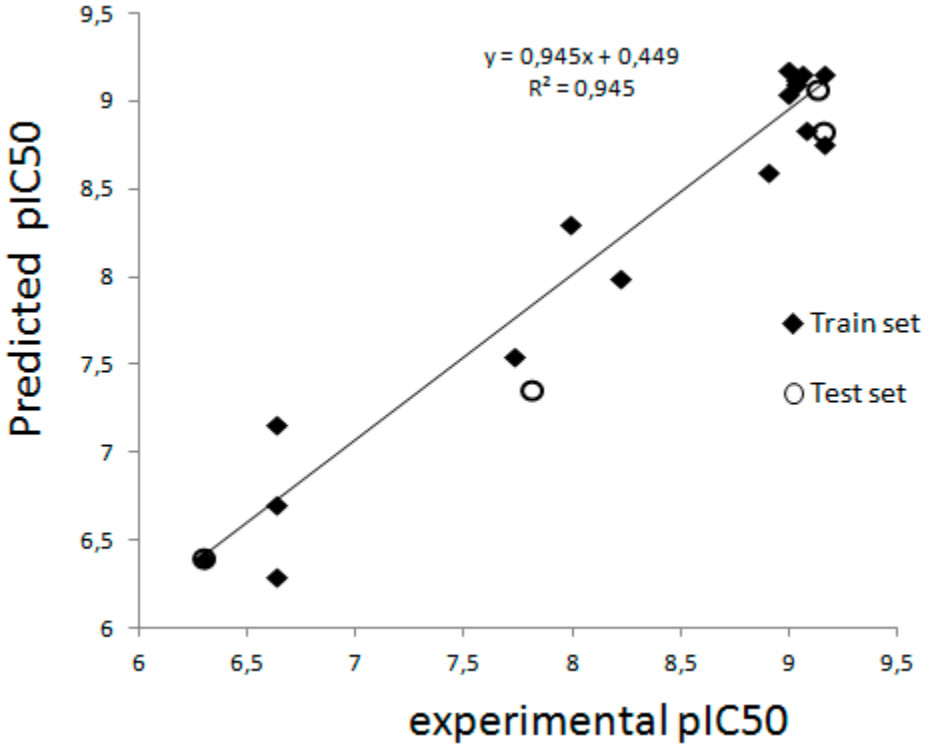


Figure 1. Comparison of experimental pIC50 values for triazine interactions with polyclonal antibodies (S2) and predicted activity based on Model 5; compounds **4**, **6**, **14** and **19** were used as a test set.

2.1.2 K4E7 system analysis (monoclonal antibodies)

For K4E7 system a series of multiple linear regression models was obtained:

Model 6

*pIC50 = 7.142 − 9.584 * Mor21v − 0.561 * NPlusO_Count + 33.83 * R6m+*

Model 7

*pIC50 = 5.859 − 0.018 * Jurs_TPSA − 7.118 * Mor21v + 27.20 * R6m+*

Model 8

*pIC50 = -3.941 + 1.291 * ES_Count_sssCH + 5.271 * IC2 − 0.216 * Jurs_DPSA_3*

Model 9

pIC50 = 7.173 − 8.061 * Mor21v − 0.0385 * TPSA_NO + 34.56 * R6m+

Model 10

*pIC50 = 7.369 + 1.637 * ES_Count_sssCH + 0.287 * ES_Sum_sCl − 0.791 * Num_H_Donors_Lipinski*

Model 10 has a high predictive ability, its statistical parameters are higher than for the models

 doi:10.20944/preprints201811.0246.v1

6-9 (Table 3).

Table 3. Statistical parameters of 2D QSAR models for K4E7 system.

№	Statistical parameters	Model 6	Model 7	Model 8	Model 9	Model 10
1	R^2	0.921	0.915	0.917	0.925	**0.966**
2	R^2 *adj*	0.902	0.894	0.896	0.906	**0.957**
3	q^2	0.819	0.778	0.862	0.832	**0.939**
4	*pred_*R^2	0.767	0.788	0.898	0.704	**0.969**
5	*Max error*	0.947	1.110	1.065	0.913	**0.469**
6	*RMS error*	0.431	0.449	0.444	0.422	**0.285**
7	*LOF*	0.822	0.890	0.869	0.787	**0.359**

Num_H_Donors_Lipinski (α = -37.9%) is the number of hydrogen bond donors (Table S2). The value of the descriptor is inversely correlated with the activity: the greater the number of hydrogen bond donors the lower the activity. Apparently, the presence of amine groups (-NH- and -NH2) negatively affects the activity.

ES_Sum_sCl (34.6%) is the electrotopological index of chlorine atoms.[35] The *ES_Sum_sCl* value is directly proportional to the activity of the molecule. Apparently, the chlorine atom, as an electron-acceptor substituent, participates in electrostatic interactions.

ES_Count_sssCH (27.4%) - the number of methantriyl groups. Obviously, the presence of the *ES_Count_sssCH* descriptor in Model 10 indicates the participation of hydrophobic substituents in the triazine compounds in van der Waals interactions with hydrophobic regions of antibodies.

Experimental and predicted pIC50 values for triazine interactions with monoclonal antibodies (K4E7) based on Model 10 are presented on Figure 2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 November 2018 doi:10.20944/preprints201811.0246.v1

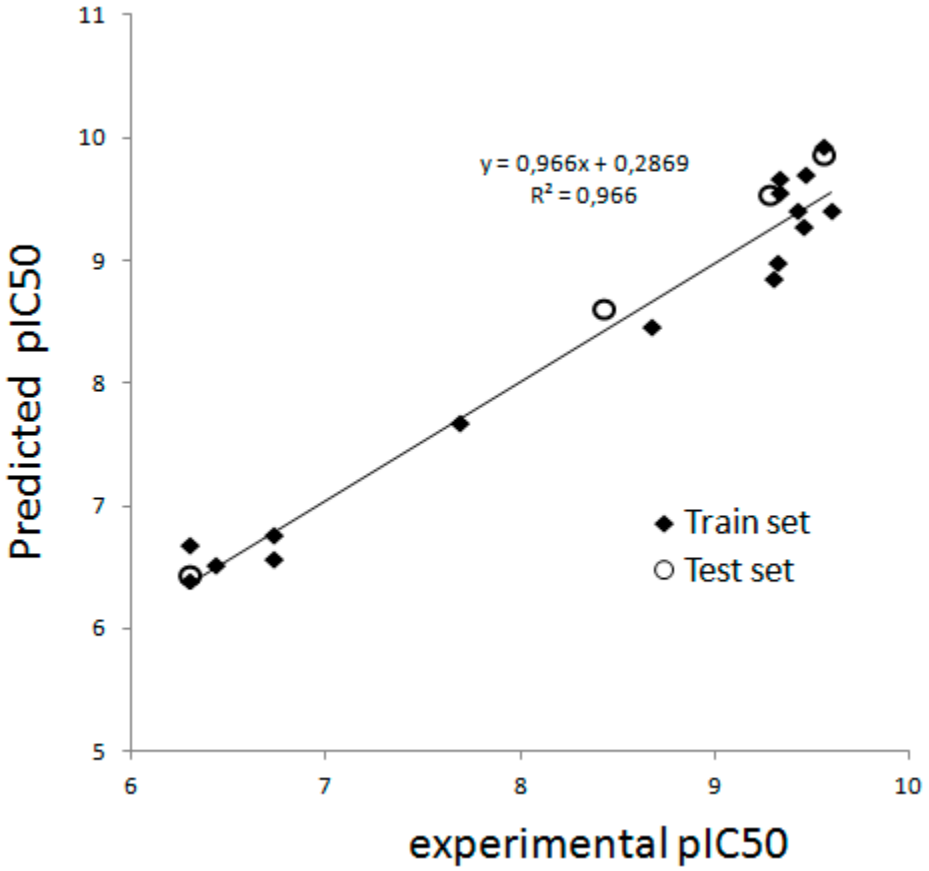


Figure 2. Comparison of experimental pIC50 values for triazine interactions with monoclonal antibodies (K4E7) and predicted activity based on Model 10; compounds **6**, **10**, **16** and **20** were used as a test set.

2.2 3D QSAR

2.2.1 S2 system study

3D QSAR analysis was carried out for all 20 compounds without splitting the sample into the training and test sets. The interaction energy of the "probe" (carbon atom with charge +1) with the target molecule was calculated at each point of the regular 3D lattice. The dimensions of the 3D cubic grid were 26 × 18 × 22 Å. The lattice spacing was 2 Å.

Three models were built consisting of 1, 2 and 3 principal components (PC). One can consider Table 4 to compare the accuracy of the activity prediction for individual molecules within each model including the optimal amount of PC for predicting the activity of each compound.

Table 4. Experimental and predicted values of triazine cross-reactivity logarithm (pIC50) in S2 system.

№	Experiment	PC 1	PC 2	PC 3	Optimal number of PC
1	9.032	8.741	8.725	9.056	3
2	8.222	8.369	8.210	8.405	2
3	6.635	7.384	6.709	6.846	2
4	7.815	7.314	7.238	7.297	1
5	9.056	8.305	8.593	8.641	3

 doi:10.20944/preprints201811.0246.v1

6	9.131	8.755	8.793	8.763	2
7	9.076	9.344	9.205	9.393	2
8	8.903	7.772	8.630	8.713	3
9	7.987	8.829	8.408	7.878	3
10	9.155	9.419	9.472	9.492	1
11	9.155	8.985	9.507	9.276	3
12	9.032	9.738	9.062	8.811	2
13	8.991	8.640	8.730	9.147	3
14	9.155	8.576	9.158	9.108	2
15	8.991	8.913	9.061	9.016	3
16	7.731	7.210	7.226	7.406	3
17	6.635	7.137	7.274	7.039	3
18	6.635	6.611	7.232	6.950	1
19	6.301	7.053	6.512	6.529	2
20	6.301	6.843	6.192	6.170	2

Van der Waals interactions make the main contribution to the interactions between triazine and antibody in the S2 system in all three models (Table S3).

To estimate the predictive ability of the 3D-QSAR models obtained the following statistical parameters were used: the coefficient of determination (R^2), the mean-square error of the model (SDEC), the F-statistic value, the correlation coefficient of the LOO method ($q^2_{(LOO)}$), the correlation coefficient of the LMO method ($q^2_{(LMO)}$), the mean-square error of prediction (SDEP). The model with 3 principal components showed satisfactory statistical results (Table 5).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 November 2018 doi:10.20944/preprints201811.0246.v1

Table 5. Statistical parameters of COMFA models for S2 system.

№	PC	R^2	SDEC	F-test	$q^2_{(LOO)}$	$SDEP_{(LOO)}$	$q^2_{(LMO)}$	$SDEP_{(LMO)}$
1	1	0.736	0.554	50.18	0.260	0.927	0.185	0.968
2	2	0.896	0.347	73.57	0.602	0.680	0.439	0.800
3	3	0.938	0.269	80.39	0.680	0.610	0.498	0.754

As can be seen from Table 5, the model with three principal components has the highest predictive ability. High coefficients of determination (R2 = 0.938) and LOO cross-validation (q^2 = 0.68) indicate the statistical significance of the model obtained. High correlation between predicted and experimental values of the model with three PC can be seen on Figure 3.

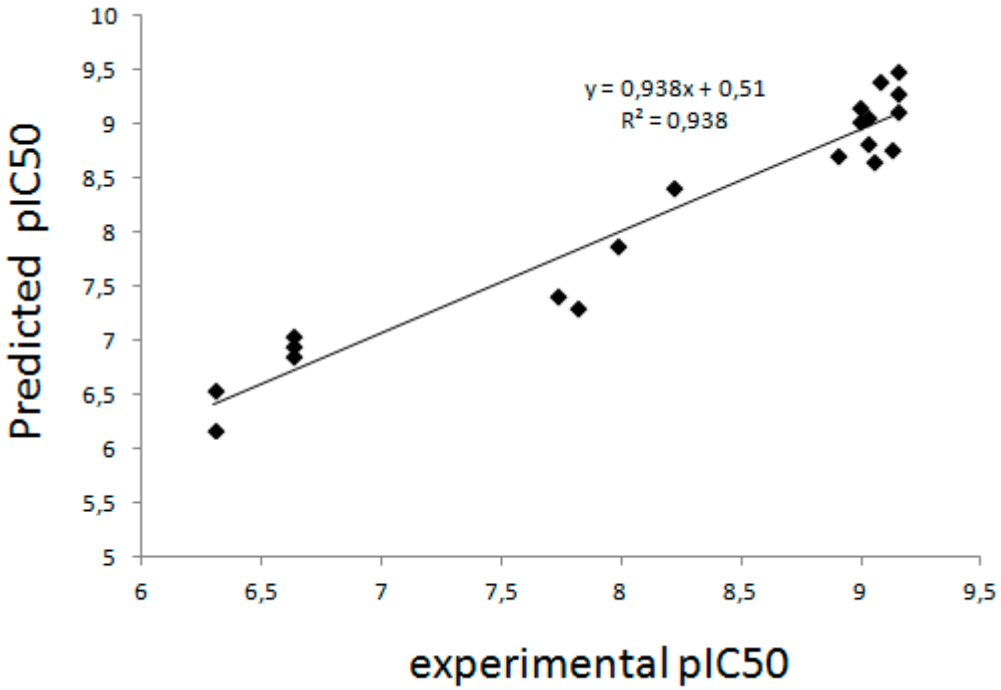


Figure 3. Experimental vs. predicted values of triazines cross-reactivity in S2 system according to the COMFA method (3D QSAR).

MIF contour maps were obtained to visualize the information about the 3D-QSAR models. Contour maps are presented in Figure 4. The steric fields presented by green contours are favorable for bulk substituents and have a positive effect on cross-reactivity (Figure 4, a), while the yellow contours represent the areas where the presence of bulk substituents is unfavorable for high activity (Figure 4, b). The electrostatic fields are represented by blue and red contours. The blue contour reflects the areas where negatively charged atoms have a positive effect on activity (Figure 4, c), while the red contours reflect the regions in which the presence of negatively charged atoms is unfavorable for high activity (Figure 4, d).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 November 2018 doi:10.20944/preprints201811.0246.v1

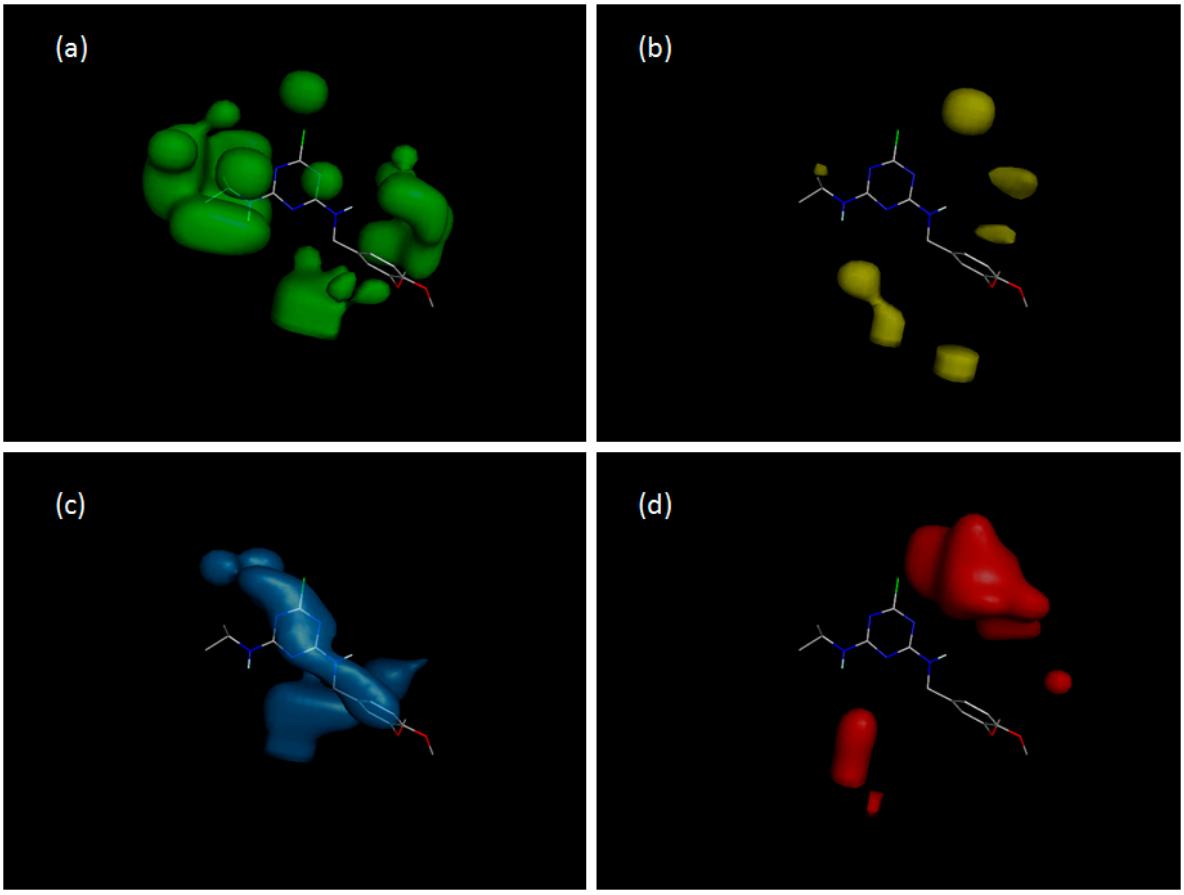


Figure 4. The MIF contour maps for the system S2 and compound **7**: (a) favorable steric interactions; (b) unfavorable steric interactions; (c) favorable electrostatic interactions; (d) unfavorable electrostatic interactions.

2.2.2 K4E7 system study

3D QSAR analysis was carried out for all 20 compounds without splitting the sample into the training and test parts. The same alignment was used as for the S2 system.

The predicted values of triazine activity are shown in Table 6: the calculated values of the model consisting of 3 principal components have the best match with the experimental data. The level of correlation between the activities predicted by the COMFA model and the experimental values is also shown on Figure 5.

Table 6. Experimental and predicted values of triazines cross reactivity in the K4E7 system.

№	Experiment	PC 1	PC 2	PC 3	Optimal number of PC
1	9.432	9.208	8.784	9.120	1
2	8.678	9.050	8.538	8.786	3
3	7.686	8.187	7.487	7.746	3
4	6.431	7.330	6.371	6.429	3
5	9.569	9.147	9.191	9.031	2
6	9.569	8.573	9.279	9.361	3
7	9.602	10.651	10.054	9.397	3
8	9.328	8.253	8.757	8.894	3
9	6.301	6.857	7.14	6.442	3

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 November 2018 doi:10.20944/preprints201811.0246.v1

10	9.284	9.399	9.533	9.455	1
11	9.469	9.134	9.502	9.462	3
12	9.456	9.400	9.595	9.871	1
13	9.337	8.744	8.763	9.212	3
14	9.337	9.287	9.809	9.776	1
15	9.310	9.212	9.704	9.527	1
16	8.434	7.852	8.203	8.497	3
17	6.731	7.493	7.023	6.721	3
18	6.731	6.256	6.404	6.362	2
19	6.301	7.279	6.947	6.931	3
20	6.301	5.971	6.199	6.266	3

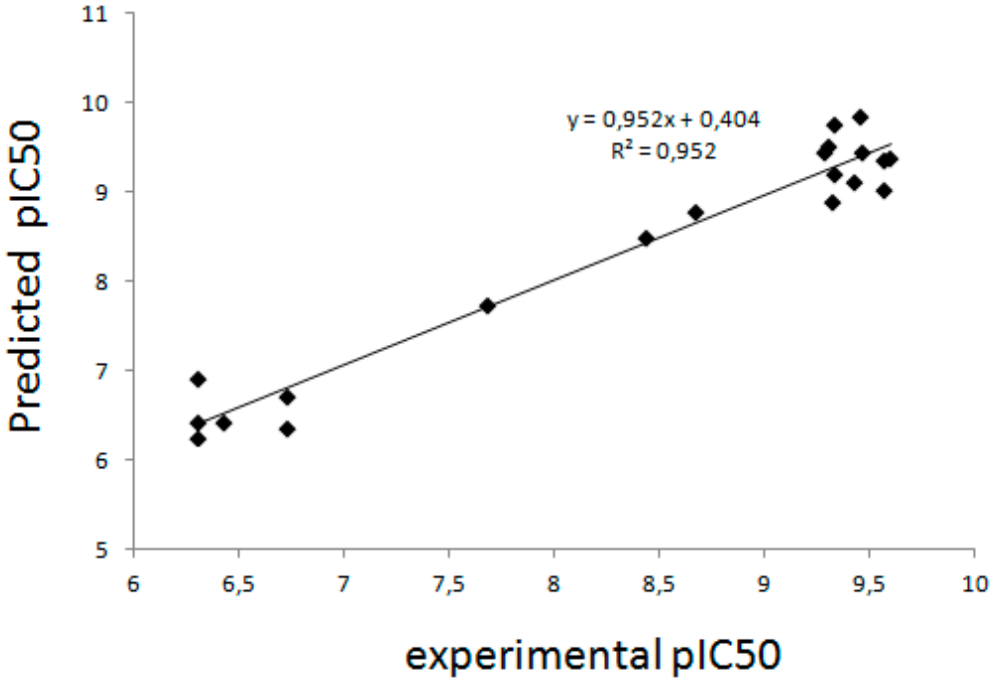


Figure 5. K4E7 system experimental and predicted pIC50 triazine values based on the COMFA method (3D QSAR).

Table 7. Statistical parameters of 3D QSAR models for K4E7 system.

№	PC	R^2	SDEC	F-test	$q^2_{(LOO)}$	$SDEP_{(LOO)}$	$q^2_{(LMO)}$	$SDEP_{(LMO)}$
1	1	0.780	0.620	63.90	0.382	1.039	0.320	1.081
2	2	0.903	0.413	78.70	0.591	0.846	0.517	0.908
3	3	0.952	0.291	105.0	0.637	0.796	0.566	0.861

Table 7 shows that the model with three PC has the highest values of the determination coefficient ($R^2 = 0.952$) and LOO cross-validation parameter ($q^2 = 0.637$), which indicates the statistical significance of the model obtained. The contour maps of the molecular interaction for this model are presented in Figure 6.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 November 2018 doi:10.20944/preprints201811.0246.v1

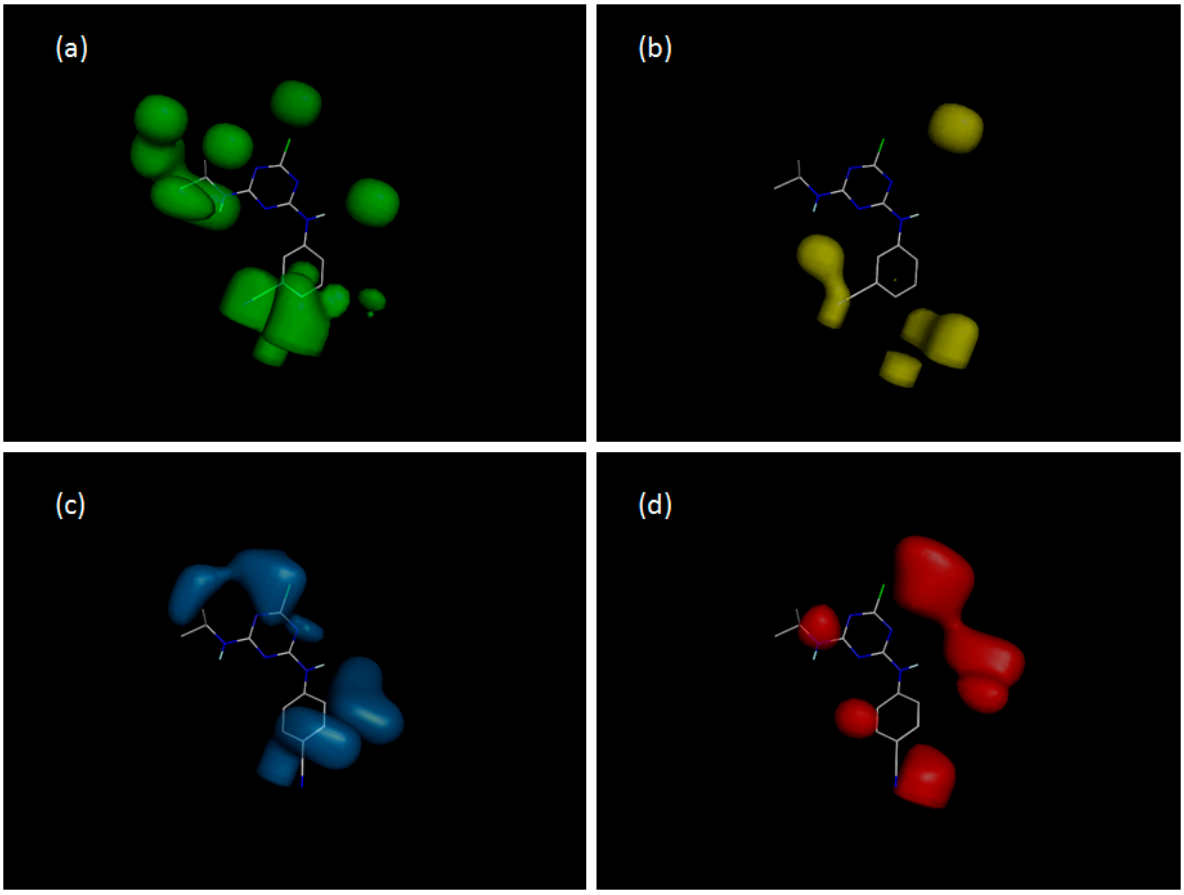


Figure 6. MIF contour maps for the K4E7 system: (a) favorable steric interactions and (b) unfavorable steric interactions for compound **11**; (c) favorable electrostatic interactions and (d) unfavorable electrostatic interactions for compound **12**.

The contributions of van der Waals and electrostatic interactions to the model with three principal components are approximately equal: 51% and 49%, respectively (Table S4).

3. Discussion

It seems reasonable to compare the most similar compounds with each other and find out which descriptors play a key role in their recognition.

Atrazine in the S2 system (polyclonal antibodies) exhibits significantly higher activity (CR 100%) than its metabolites - compounds 2, 3 and 4 (CR 15%, 0.4% and 6%, respectively). Atrazine has the highest value of the Solvation_E descriptor (see Table S1) and the lowest value of FPSA_3 among these four compounds. FPSA_3 is the partial polar surface area of the molecule. The value of this descriptor is inversely proportional to pIC50. FPSA_3 reflects the ability of molecules to form hydrogen bonds. The lower the proportion of the polar surface area of the molecule (N, O atoms and associated hydrogen) the lower the activity. For compounds 1-4 FPSA_3 is equal to 0.037, 0.048, 0.054 and 0.054, respectively. In general, this is consistent with our observation that the polarity of atrazines is unfavorable for the interaction with the antibodies.

For the K4E7 system (monoclonal antibodies) pIC50 values of 1-4 compounds are highly correlated with the Jurs_TPSA descriptor (the total area of the polar surface): 29.44, 41.45, 43.08 and 90.88, respectively. Jurs_TPSA is the analogue of FPSA_3 descriptor, and just like FPSA_3 it indicates

 doi:10.20944/preprints201811.0246.v1

that hydrophobic interactions play an important role in the recognition of triazine by the antibodies. Polar atoms (primarily nitrogen and oxygen) on the surface of the molecule, on the contrary, are unfavorable for the binding with the antibody.

For system S2, in the series of compounds 6-9 (CR 126%, 109%, 74% and 9%, respectively), the decrease in pIC50 is inversely correlated with the growth of FPSA_3: 0.043, 0.043, 0.048 and 0.049, respectively.

It is more difficult to interpret the COMFA contour maps than 2D descriptors, however, in the case of the K4E7 system and 6-9 compounds we should regard them. The radical at the C6 position does not differ in these compounds, and the variation of the substituent in the C4 position obviously plays the main role. Geometry of compound 7 is the most advantageous for the interaction with the antibody, and geometry 9 is the least advantageous (Figure 2). The contour map of electrostatic charges (Figure 8d) indicates that the presence of a negatively charged carboxyl group at position 2' (compound 9) is unfavorable for the binding with the antibody. The contour map of electrostatic charges on Figure 8c, on the contrary, shows how the presence of negatively charged oxygen atoms of the methoxy groups of compound 7 has a positive effect on the activity. The contour map on Figure 8a shows that the aniline residue at the C4 position (namely, the benzene ring) generally has a positive effect on the activity of atrazines; The contour map of van der Waals interactions on Figure 8b shows that the bulky substituents localized at meta position of the aniline have a negative effect on the activity: the methoxy substituents of compound 9 are located in these areas.

Compounds 10, 11 and 12 differ have ortho-, meta- and para-positions of the nitrile of the aniline substituent, respectively (Figure 2). The activity of these compounds in the system S2 is determined by the contour maps of electrostatic interactions (Figures 6c, 6d).

Let us regard compounds 13, 14 and 15 which possess trifluoromethyl radical (-CF3) in the ortho-, para- and meta- positions, respectively. In the S2 system, the molecule 14 is more active than 13 and 15 (Table 1). In order to answer the question why these compounds exhibit different activities, it is required to consider the FPSA_3 descriptor involved in 2D QSAR models 2, 3 and 4. This descriptor reflects the fractional polar surface area of the molecule, and is inversely correlated with pIC50. For compounds 13, 14 and 15, FPSA_3 is equal 0.033, 0.032 and 0.033, respectively. Obviously, the smaller the fraction of the positively charged surface of the molecule, the lower the activity. In general, this is consistent with our observations that hydrophobicity of atrazines facilitates their recognition by the antibodies.

Obviously, the low activity of compounds 16, 19 and 20 in both systems (S2 and K4E7) is due to the absence of the isopropylamino radical at the C6 position. This is confirmed both by the 2D descriptors ES_Count_sssCH, ES_Sum_sssCH, and by the contour maps of van der Waals interactions (Figures 6a and 8a). Also, it is quite obvious that compounds 18, 19 have low pIC50 values due to the presence of the -OH group in the C2 position instead of chlorine.

If we compare the 2D QSAR models for monoclonal antibodies (mAb) and polyclonal antibodies (pAb) one can see that both models (Model 4 for mAb and Model 7 for pAb) contain polar surface area parameter: Jurs_FPSA_3 and Jurs_TPSA, respectively. The polar surface area is inversely correlated with the activity, it means that the polarity of atrazines is unfavorable for the interaction with both monoclonal and polyclonal antibodies. Both mAb and pAb are affected by the amount of methantryil groups (ES_Sum_sssCH and ES_Count_sssCH parameters, respectively), which means that in both cases bulk substituents are favorable for higher activity. The latter statement is supported

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 November 2018 doi:10.20944/preprints201811.0246.v1

by 3D molecular interaction field contour maps (Figures 6a and 8a). We may see that both mAb and pAb systems have very much in common.

4. Materials and Methods

4.1 Conformational analysis and geometry optimization

Preparation of molecular geometries for 2D and 3D QSAR analysis was carried out in Spartan v.16 program. A series of conformers for each of the 20 compounds was obtained by the systematic search method using the MMFF force field [26]. Optimization of the geometry for each conformer was carried out using the semi-empirical AM1 method [27]. In order to confirm the correspondence between the obtained geometry and the minimum surface of the potential energy for each conformer, the Hessian was calculated. The conformers with the lowest total energy were used for 2D QSAR analysis.

4.2 2D QSAR

Using a random number generation triazines were divided into 2 samples: training set (80%) and test set (20%). The requirements for the maximum and minimum values in the test sample were as follows: 1) the maximum value of pIC50 should be less than or equal to the maximum value of pIC50 in the training sample; 2) the minimum value of pIC50 must be greater than or equal to the minimum value of pIC50 in the training sample.

The required linear regression equation should have the following form:

$$y = a_1 * x_1 + \ldots a_n * x_n + c$$

where y is the dependent variable (pIC50); a_1 and a_n are regression coefficients; x_1 and x_n are independent variables (descriptors); c is a regression constant.

We estimated the contribution of a descriptor to the model using following equation:

$$\alpha(x_1) = \frac{R^2(x_1, x_2, x_3) - R^2(x_2, x_3)}{3 \times R^2(x_1, x_2, x_3) - R^2(x_1, x_2) - R^2(x_1, x_3) - R^2(x_2, x_3)} \times 100\%$$

where $\alpha(x_1)$ - is the relative contribution of the descriptor x_1 to the model with three descriptors; $R^2(x_1, x_2, x_3)$ is the determination coefficient of the model with all three descriptors; $R^2(x_2, x_3)$ is the determination coefficient of the model with two descriptors: x_2 and x_3.

During the 2D QSAR analysis, different types of descriptors were used: constitutional (number of hydrogen bond donors, number of hydrogen bond acceptors, number of rings, number of chains, number of CH3 groups, number of OH groups, etc.), electrostatic descriptors (maximum positive charge, maximum negative charge, the number of positively charged atoms, the number of negatively charged atoms, etc.), topological descriptors (kappa-indices describing the shape of the molecule, the indices of molecular bonds of Kier and Hall, etc.), 3D descriptors (volume, surface area, length and area of the projection on the coordinate axis, etc.), physico-chemical (lipophilicity (LogP), molecular refraction, polarizability, solubility in water, etc.), quantum-chemical descriptors (energy of frontier molecular orbitals, electrostatic charges of atoms, the population of atoms according to Mulliken, etc.) – a total of more than 3000 descriptors. The values of the descriptors were calculated using program packages Spartan v. 16 and E-Dragon 1.0.

For QSAR model obtained the cross-validation was performed. Also, the models were tested for their predictive power using an external test set of compounds. The following statistical indicators were used:

 doi:10.20944/preprints201811.0246.v1

- r^2 - coefficient of determination for the training sample;
- the determination coefficient R^2adj, corrected for the number of descriptors involved in the model;
- q^2 - r^2 from the results of internal cross-validation of the training sample using the leave-one-out (LOO) method;
- LOF - error of Friedman approximation (Friedman Lack of Fit);
- RMSE - root-mean-square error;
- Max Error - the maximum prediction error for all the compounds in the test and training set;
- pred_r^2 - r^2, which estimates the predictive ability of the model relative to the test set.

4.3 3D QSAR

4.3.1 Spatial alignment of molecules

Alignment of molecules plays a key role in 3D-QSAR. As a rule, the most energetically favorable conformation of the most active compound or the geometry corresponding directly to the interaction of the ligand and target (antibody), obtained on the basis of X-ray diffraction analysis or docking data, is chosen as the template for the alignment [28]. In our case, the most active conformation is unknown, and a number of the most active compounds (**10**, **14**) have a large number of conformers, which makes it difficult to analyze them correctly using quantum-chemical methods of required accuracy. Compounds **5**, **11** and **12** are also highly active, but with a small number of conformers compared to other triazines in the sample. All conformers of compounds **5** (12 conformers), **11** (23 conformers) and **12** (12 conformers) were optimized using the Hartree-Fock method and 6-31G(d) basis set (Figure 7). It was found that the most energetically favorable conformers of these compounds are identical, and therefore it can be assumed that this particular conformation is the most active and corresponds to the interaction with the antibody.

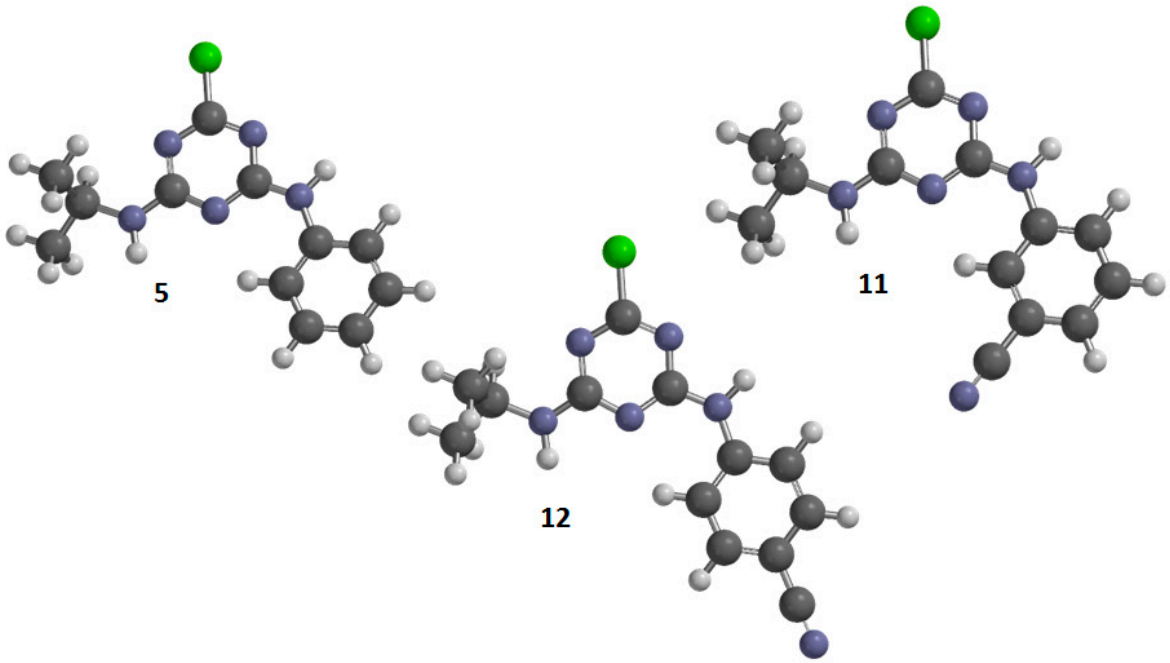


Figure 7. The most energetically favorable conformations of compounds **5**, **11** and **12** according to the calculation by HF/6-31G(d) method.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 November 2018 doi:10.20944/preprints201811.0246.v1

Triazine molecules in various conformations, optimized with AM1 method, were superimposed on the template (compounds **5** and/or **11**) using the algorithm presented in the open3DALIGN program [29]. Conformations corresponding to the largest overlap with the atoms of compounds **5** and **11** were selected (Figure 8).

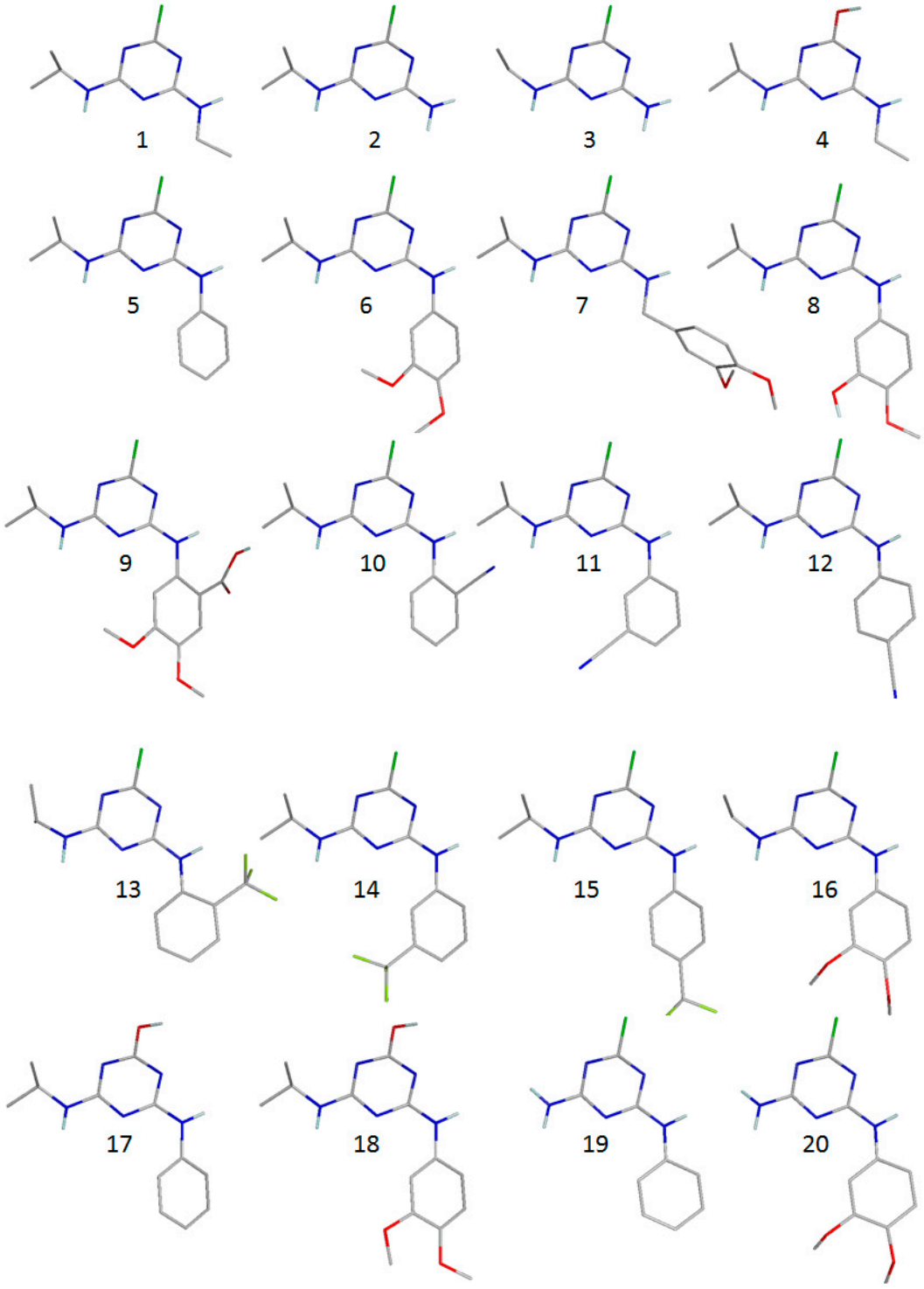


Figure 8. Optimized with AM1 method triazine geometries (hydrogen atoms are represented only in -OH groups).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 November 2018 doi:10.20944/preprints201811.0246.v1

The aligned geometries (Figure S1) were used to create and analyze the fields of molecular interactions using the open3DQSAR program [30]. Visualization of 3D structures and contour maps of the force fields was carried out in the PyMol program.

4.3.2 Building of molecular interaction fields (MIF)

Molecular interaction fields (MIF) were calculated to analyze the pIC50 values of triazines and to create COMFA models for S2 and K4E7 systems.

Two potentials were used to create the fields of molecular interaction: the steric potential in the form of the Lennard-Jones function 6-12 between the atoms of the molecule and the sp3 carbon atom:

$$E_{VdW} = \sum_{i=1}^{n} \left[\frac{A_i}{r_i^{12}} - \frac{B_i}{r_i^{6}} \right];$$

The electrostatic field was calculated by summing the Coulomb interactions between the test atom with charge +1 and triazine molecules:

$$E_{ele} = k \sum_{i=1}^{n} \left[\frac{q_i}{r_i^{m}} \right]$$

4.3.3 MIF preprocessing

The force field data was preliminarily processed to remove non-informative *x* variables:

1) equated to zero the values of the variables less than 0.05 kcal mol^{-1};
2) cut off values above 30 kcal mol-1 and below -30 kcal mol^{-1};
3) deleted *x* variables with a standard deviation value below <0.1;
4) removed variables that have 4 or less nonzero values within the sample.

4.3.4 Building of regression models

After all the preliminary operations were completed, we proceeded to construct the models using Partial Least Squares (PLS) method [31], based on the NIPALS algorithm (Non-linear Iterative PArtial Least Squares) [32]. The predictive power of the models was evaluated using "leave-one-out" (LOO) cross-validation method:

$$q^2 = 1 - \frac{\sum (y_{obs} - y_{pred})^2}{\sum (y_{obs} - y_{mean})^2};$$

Also, we used "leave-many-out" cross-validation method: a sample of 20 compounds was randomly broken into the training (75%) and test (25%) parts. The q^2 parameter was calculated for the test part. The partition procedure was repeated 50 times, and the average of the 50 partitioning variants was given the value q^2 (LMO).

To estimate the internal stability and predictive power of the 3D models, the standard prediction error (SDEP) parameter was calculated:

$$SDEP = \sqrt{\sum \frac{(y_{obs} - y_{pred})^2}{N}};$$

 doi:10.20944/preprints201811.0246.v1

where y_{obs} is the experimental value; y_{pred} is the predicted value; y_{mean} – mean value; N is the number of molecules in the sample.

Procedures for selecting variables and methods for creating clusters of variables with common characteristics were used in order to increase the predictive ability of models:

– The SRD (Smart Region Definition) method [33] groups variables based on their localization in three-dimensional space. This procedure reduces the number of descriptors bearing the same information;

– Fractional Factorial Design (FFD) [34,35] allows the selection of variables that have the greatest impact on the predictive ability of models. The FFD selection was conducted based on the leave-many-out cross analysis, using the SRD of the variable group.

5. Conclusions

In this study we evaluated the efficiency of the interaction of 20 triazines with polyclonal and monoclonal antibodies grown using atrazine as the immunizing hapten.

Comparison of the two immunoassay systems showed that the system with the polyclonal antibodies (S2) is much easier to describe using 2D QSAR methodology, and the system with monoclonal antibodies can be described using the 3D QSAR COMFA method. Thus, for the 3D QSAR model of the S2 system, the main statistical parameter q^2 (LMO) is equal 0.498, and for K4E7 q^2 (LMO) = 0.566. Apparently, this is explained by the fact that in the case of polyclonal antibodies, we deal with several targets, while in interaction with monoclonal antibodies the target is one, and therefore it is easier to describe it using specific fields of molecular interactions distributed in space.

There is an interesting fact that while for the S2 system the main contribution (72%) is made by hydrophobic interactions (Table S3), for K4E7 the fraction of van der Waals and electrostatic interactions is equal (Table S4).

Based on our analysis, we conclude that the effectiveness of the interaction with polyclonal antibodies significantly depends on the presence of the isopropylamino group in the C6 position of triazine, the presence of chlorine at the C2 position, and also the aniline radical in the C4 position, which substituents, in turn, participate in electrostatic interactions and hydrogen bonding.

The obtained data as well as earlier published results of QSAR techniques application for immune recognition demonstrate the efficiency of this technique in identification of key structures responsible for distinguishing of structurally close antigens by antibodies. This information allows theoretical formulation of criteria to molecules that could be recognized by the same antibodies (thus extending selectivity of one immunoassay) or needs in different antibodies for detection (thus realizing assays with mixtures or arrays of immunoreactants).

Supplementary Materials: The supplementary materials are available online.

Author Contributions: writing—original draft preparation, Andre A. Buglak; writing—review and editing, Anatoly V. Zherdev and Hong-Tao Lei; supervision, Boris B. Dzantiev.

Funding: The work was financially supported by the Russian Science Foundation (Project No. 14-16-00149).

Conflicts of Interest: The authors declare no conflict of interest

References

1. Li, Y.; Sun, Y.; Beier, R.C.; Lei, H.; Gee, S.; Hammock, B.D.; Wang, H.; Wang, Z.; Sun, X.; Shen, Y.; Yang, J.; Xu, Z. *Trends Anal. Chem.*, **2017**, *88*: 25-40.
2. Bottoni, P.; Grenni, P.; Lucentini, L.; Barra Caracciolo, A. *Microchem. J.*, **2013**, *107*: 136–142.
3. Cheney, M.A.; Y.Shin, J.; Crowley, D.E.; Alvey, S.; Malengreau, N.; Sposito, G. *Colloids Surf. A*, **1998**, *137*, 267-273.
4. Shin, J.Y.; Cheney, M.A. *Environ Toxicol Chem*. **2005**, *24*, 1353-1360.
5. Tian, Y.; Shen, W.; Jia, F.; Ai, Z.; Zhang, L. *Chem. Eng. J.*, **2017**, *330*, 1075-1081.
6. Rowe, A.M.; Brundage, K.M.; Barnett, J.B. *Basic Clin. Pharmacol. Toxicol.*, **2008**, *102*, 139-145.
7. Agopian, A.J.; Langlois, P.H.; Cai, Y.; Canfield, M.A.; Lupo, P.J. *Matern. Child. Health J.*, **2013**, *17*, 1768-1775.
8. Xu, Z.L.; Shen, Y.D.; Zheng, W.X.; Beier, R.C.; Xie, G.M.; Dong, J.X.; Yang, J.Y.; Wang, H.; Lei, H.T.; She, Z.G.; Sun, Y.M. *Anal. Chem.*, **2010**, *82*, 9314-9321.
9. Yuan, M.; Na, Y.; Li, L.; Liu, B.; Sheng, W.; Lu, X.; Kennedy, I.; Crossan, A.; Wang, S. *J. Agric. Food. Chem.*, **2012**, *60*, 10486-10493.
10. Liu, C.; Dou, X.; Zhang, L.; Kong, W.; Wu, L.; Duan, Y.; Yang, M. *Anal. Chim. Acta.*, **2018**, *1012*, 90-99.
11. Bitencourt, M.; Freitas, M.P. *Pest Manag. Sci.*, **2008**, *64*, 800-807.
12. Zuo, Y.; Wu, Q.; Su, S.W.; Niu, C.W.; Xi, Z.; Yang, G.F. *J. Agric. Food Chem.*, **2016**, *64*, 552-562.
13. Funar-Timofei, S.; Borota, A.; Crisan, L. *Mol. Divers.*, **2017**, *21*, 437-454.
14. Xie, Y.; Peng, W.; Ding, F.; Liu, S.J.; Ma, H.J.; Liu, C.L. *Pest Manag. Sci.*, **2018**, *74*, 189-199.
15. Wang, Z.; Zhu, Y.; Ding, S.; He, F.; Beier, R.C.; Li, J.; Jiang, H.; Feng, C.; Wan, Y.; Zhang, S.; Kai, Z.; Yang, X.; Shen, J. *Anal. Chem.*, **2007**, *79*, 4471-4483.
16. Mu, H.; Wang, B.; Xu, Z.; Sun, Y.; Huang, X.; Shen, Y.; Eremin, S.A.; Zherdev, A.V.; Dzantiev, B.B.; Lei, H. *Analyst*, **2015**, *140*, 1037-1045.
17. Chen, J.; Lu, N.; Shen, X.; Tang, Q.; Zhang, C.; Xu, J.; Sun, Y.; Huang, X.A.; Xu, Z.; Lei, H. *J. Agric. Food Chem.*, **2016**, *64*, 2772-2779.
18. Chen, J.; Wang, L.; Lu, L.; Shen, X.; Huang, X.A.; Liu, Y.; Sun, X.; Wang, Z.; Eremin, S.A.; Sun, Y.; Xu, Z.; Lei, H. *Anal. Chem.*, **2017**, *89*, 6740-6748.
19. Mu, H.; Xu, Z.; Liu, Y.; Sun, Y.; Wang, B.; Sun, X.; Wang, Z.; Eremin, S.; Zherdev, A.V.; Dzantiev, B.B.; Lei, H. *Spectrochim. Acta A Mol. Biomol. Spectrosc.*, **2018**, *194*, 83-91.
20. Xu ZL, Xie GM, Li YX, Wang BF, Beier RC, Lei HT, Wang H, Shen YD, Sun YM. *Anal. Chim. Acta*, **2009**, *647*, 90-96.
21. Xu, Z.L.; Shen, Y.D.; Zheng, W.X.; Beier, R.C.; Xie, G.M.; Dong, J.X.; Yang, J.Y.; Wang, H.; Lei, H.T.; She, Z.G.; Sun, Y.M. *Anal. Chem.*, **2010**, *82*, 9314-9321.
22. Yuan, M.; Liu, B.; Liu, E.; Sheng, W.; Zhang, Y.; Crossan, A.; Kennedy, I.; Wang, S. *Anal. Chem.*, **2011**, *83*, 4767-4774.
23. Wang, Z.; Kai, Z.; Beier, R.C.; Shen, J.; Yang, X. *Int. J. Mol. Sci.*, **2012**, *13*, 6334-6351.
24. Dankwardt, A.; Hock, B.; Simon, R.; Freitag, D.; Kettrup, A. *Environ. Sci. Technol.*, **1996**, *30*, 3493–3500.
25. Hall, L.H.; Mohney, B.; Kier, L.B. *J. Chem. Inf. Comput. Sci.*, **1991**, *31*, 76-82.
26. Halgren, T.A. *J. Comp. Chem.*, **1996**, *17*, 490–519.
27. Dewar, M.J.S.; Zoebisch, E.G.; Healy, E.F.; Stewart, J.J.P. *J. Am. Chem. Soc.*, **1985**, *107*, 3902-3909.
28. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E.W. Jr. *Pharmacol. Rev.*, **2013**, *66*, 334-395.
29. Tosco, P.; Balle, T.; Shiri, F.J. *Comput. Aided Mol. Des.*, **2011**, *25*, 777-783.
30. Tosco, P.; Balle, T. *J. Mol. Model.*, **2011**, *17*, 201-208.

31. Stahle, L.; Wold, S. *J. Chemom.*, **1987**, *1*, 185-196.
32. Wold, S.; Sjöström, M.; Eriksson, L. *Chemometrics Intell. Lab. Syst.*, **2001**, *58*, 109-130.
33. Pastor, M.; Cruciani, G.; Clementi, S. *J. Med. Chem.*, **1997**, *40*, 1455-1464.
34. Baroni, M.; Clementi, S.; Cruciani, G.; Costantino, G.; Riganelli, D.J. *Chemometr.*, **1992**, *6*, 347-356
35. Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. *Quant. Struct-Act. Relat.*, **1993**, *12*, 9-20.

Sample Availability: Samples of the compounds listed in the main text of the manuscript are available from the authors.