J Comput Aided Mol Des (2008) 22:747–759
DOI 10.1007/s10822-008-9211-x

The effects of characteristics of substituents on toxicity of the nitroaromatics: HiT QSAR study

Victor E. Kuz'min · Eugene N. Muratov · Anatoly G. Artemenko · Leonid Gorb · Mohammad Qasim · Jerzy Leszczynski

Received: 9 December 2007 / Accepted: 18 March 2008 / Published online: 2 April 2008

Abstract The present study applies the Hierarchical Technology for Quantitative Structure–Activity Relationships (HiT QSAR) for (i) evaluation of the influence of the characteristics of 28 nitroaromatic compounds (some of which belong to a widely known class of explosives) as to their toxicity; (ii) prediction of toxicity for new nitroaromatic derivatives; (iii) analysis of the effects of substituents in nitroaromatic compounds on their toxicity in vivo. The 50% lethal dose concentration for rats (LD_{50}) was used to develop the QSAR models based on simplex representation of molecular structure. The preliminary 1D QSAR results show that even the information on the composition of molecules reveals the main tendencies of changes in toxicity. The statistic characteristics for partial least squares 2D QSAR models are quite satisfactory ($R^2 = 0.96$–0.98; $Q^2 = 0.91$–0.93; $R^2_{test} = 0.89$–0.92), which allows us to carry out the prediction of activity for 41 novel compounds designed by the application of new combinations of substituents represented in the training set. The comprehensive analysis of toxicity changes as a function of substituent position and nature was carried out. Molecular fragments that promote and interfere with toxicity were defined on the basis of the obtained models. It was shown that the mutual influence of substituents in the benzene ring plays a crucial role regarding toxicity. The influence of different substituents on toxicity can be mediated via different C–H fragments of the aromatic ring.

Keywords Applicability domain · HiT QSAR · Nitroaromatic compounds · SiRMS · Targeted design · Toxicity in vivo · Virtual screening

Abbreviations

AVS	Automatic variables selection
DA	Applicability domain
GA	Genetic algorithm
HiT QSAR	Hierarchical QSAR technology
HQSAR	Hologram QSAR approach
LD_{50}	50% lethal dose concentration
PLS	Partial least squares or projection on latent structures statistical method
Q^2	Cross-validation determination coefficient
QSAR/QSPR	Quantitative structure–activity/property relationship
R^2	Determination coefficient for training set
R^2_{test}	Determination coefficient for test set
SD	Simplex descriptor
SiRMS	Simplex representation of molecular structure QSAR approach
TV	Trend-vector statistical method

Electronic supplementary material The online version of this article (doi:10.1007/s10822-008-9211-x) contains supplementary material, which is available to authorized users.

V. E. Kuz'min · E. N. Muratov · L. Gorb · J. Leszczynski (✉)
Department of Chemistry, Computational Center for Molecular Structure and Interactions, Jackson State University, 1400 J.R. Lynch St, Jackson, MS 39217, USA
e-mail: jerzy@ccmsi.us

V. E. Kuz'min · E. N. Muratov · A. G. Artemenko
A.V. Bogatsky Physical-Chemical Institute, National Academy of Sciences of Ukraine, Odessa, Ukraine

L. Gorb
US Army ERDC (SpecPro), Vicksburg, MS 39180, USA

M. Qasim
US Army ERDC, Vicksburg, MS 39180, USA

Introduction

Military sites throughout the United States often contain soils contaminated with such nitroaromatic compounds as trinitrotoluene, trinitrobenzene and their derivatives. These compounds are toxic to a wide range of living organisms [1]. For instance, toxic effects in humans after dermal, oral, or respiratory exposures include gastrointestinal, neurological, and reproductive disorders; cirrhosis of the liver; hepatitis; cataracts; respiratory and skin irritation; nephrotoxicity; and hematological defects. Moreover, nitroaromatic compounds are widely used in medicine, industry and agriculture. Nitroaromatic pesticides as well as the explosive residues are considered as being toxic environmental pollutants. Some of these compounds have mutagenic or carcinogenic activity and may bioaccumulate in the food chain. Therefore, the presence of aromatic and nitroaromatic xenobiotics in the environment may create serious public health and environmental problems, and both nature and degree of aromatic substitutions may have profound effects on the toxicity of nitroaromatic compounds [2].

Quantitative structure–activity relationships (QSAR) are widely used to predict toxicity from chemical structure and corresponding physicochemical properties. The development and application of QSAR started with the prediction of toxicity caused by baseline toxicants [2, 3]. A prerequisite for correct predictive assessment of chemical toxicity using QSAR is the accurate assignment of toxic action modes that could be caused by different factors related to the structure of molecules-toxicants in a complex manner. The two best known chemical mechanisms of nitrocompound toxic action are one and two electron reduction [4].

The one electron reduction is accompanied by formation of high-reactive particles (free radicals) e.g. O_2^- and OH. These species cause the oxidative stress of living cells and oxidation of lipids [4]. Two electron reduction of nitroaromatics is accompanied by creation of corresponding nitroso compounds and hydroxylamines. These compounds form adducts to proteins and DNA that block normal functioning of the latter [4]. In addition, nitroaromatics participate in S_NAr reactions with nucleophilic sites in peptides and DNA (for example –OH, –SH, and $–NH_2$ groups), forming complexes with electron-donating heterocycles of peptides and DNA, which interfere with their functions [5] and act as uncoupling agents in oxidative phosphorylation [2].

In other words, the nitroaromatics display complex mechanisms of toxicity, and numerous QSAR studies have been carried out to explain and predict toxicity of nitrocompounds on different living systems [6, 7]. In a recent paper [8] the QSAR analysis of oral toxicity on rats has been extended to 28 selected nitroaromatic molecules. In spite of acceptable QSAR models being developed on the basis of topological and quantum-chemical indexes, a lot of questions remain unanswered. One of them, addressed in this paper in great detail, is the relationship between chemical structure and toxicity. This information could provide for the formation of new hypotheses as to mechanisms of chemical toxic action. Therefore, the aim of the present study is to extend recent investigations [8] by applying the Hierarchical Technology for Quantitative Structure–Activity Relationships (HiT QSAR) for (i) evaluation the influence of structure, including analysis of the substitution characteristics of nitroaromatic compounds on their toxicity, (ii) in vivo prediction of toxicity for new nitroaromatic derivatives.

Materials and methods

Hierarchical Technology for Quantitative Structure–Activity Relationships (HiT QSAR) [9, 10] based on the Simplex Representation of Molecular Structure (SiRMS) method [11, 12] was applied for analysis and prediction of the influence of selected nitroaromatic compounds on oral rat toxicity. This method has proved efficient in previous studies for solving different "structure–activity" problems [13–16]. It allows solving the QSAR task, not ab ovo, but with the use of information received from a previous stage by means of the system of improved solutions.

Application of HiT QSAR results in a set of the different QSAR models that supplement each other (consensus strategy). Consensus QSAR modeling becomes more and more popular nowadays and can be briefly described by the slogan "More models, good and different". Advantage of this technique can be easily explained by the fact that near the same predictions obtained by different and independent methods (either statistical and descriptors generation) are more reliable than singular prediction made by even the best fitted and predictable model [10, 17].

In the frameworks of SiRMS any molecule can be represented as a system of different simplexes (tetratomic fragments of fixed composition, structure, chirality and symmetry) [11–14, 16, 18] (Fig. 1). Since only 1D and 2D models are used in the current studies, we present the protocols to obtain the information at the 1D and 2D levels only.

1D models

At the 1D level, a simplex is a combination of four atoms contained in the molecule (Fig. 1). The simplex descriptor (SD) at this level is the number of quadruples of atoms of definite composition. For compound ($A_aB_bC_cD_dE_eF_f$...), the

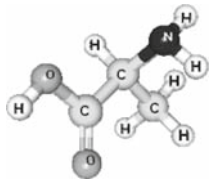
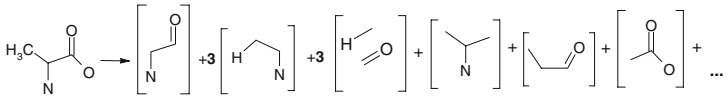
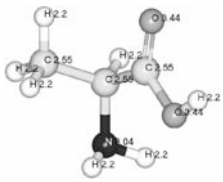


Fig. 1 Simplex representation of molecular structure on example of alanine

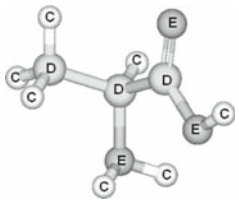
1D **$C_3H_7O_2N$ → 6** CCNO, **42** CNOH, **21** CCNH, **42** NOHH, **7** CCCH, **35** NHHH, •••

2D 

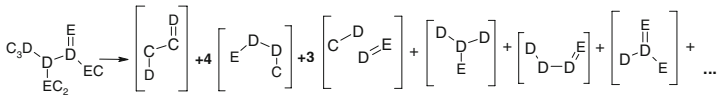
Differentiation of atoms by electronegativity



Division by groups (A ... F), depending on atom electronegativity:
A≤1.5<B≤2<C≤2.5<D≤3<E≤3.5<F≤4



1D **$C_7D_3E_3$ → 9** DDEE, **63** DEEC, **63** DDEC, **63** EECC, **7** DDDC, **105** ECCC, •••

2D 

value of SD ($A_iB_jC_lD_m$) is $K = f(i)\cdot f(j)\cdot f(l)\cdot f(m)$, where, for example,

$$f(i) = \frac{a!}{(a-i)!\cdot i!}.$$

The values of $f(j)$, $f(l)$, $f(m)$ have been calculated analogically. It is possible to define the number of smaller fragments ("pairs", "triples") by the same scheme. In this case, some of parameters of i, j, l, and m are equal to zero.

2D models

At the 2D level, the connectivity of atoms in simplex, atom type and bond nature (single, double, triple, aromatic) are considered. Atoms in simplex can be differentiated on the basis of their different characteristics, especially: a mark (symbol) that reflects atom individuality (nature or more detailed type of atom); partial atom charge [19]; and lipophilicity of atom [20]; atomic refraction; a mark, that characterizes the atom as possible donor or acceptor of H-bond (A—acceptor of hydrogen in H-bond, D—donor of hydrogen in H-bond, I—indifferent atom); electronegativity of atom; and Van der Waals attraction and repulsion potentials [21].

The usage of sundry variants of simplex vertexes (atoms) differentiation represents the important part of SiRMS. We consider that executed in many QSAR methods specification of atoms only by theirs nature (actually reflects atom identity, for example, C, N, O) limits the selection of pharmacophore fragments. For example, if the –NH– group has been selected as the determining activity fragment (pharmacophore) and the ability of H-bond formation is the factor determining its activity, then we shall miss such donors of H-bonds as, for example, OH-group, etc. The usage of atoms' differentiation by donor/acceptor of H-bond allows avoiding the situation illustrated above. Analogical examples can be made for other atomic properties (lipophilicity, partial charge, refraction, etc.).

The simplex descriptor (SD) at 2D level is the number of simplexes of the fixed composition and topology. For QSAR analysis, structural parameters corresponding to molecular fragments of different sizes can also be used together with simplex descriptors. Thus, in our work, molecular fragments containing from 1 to 7 atoms for 1D and 2–5 atomic fragments for 2D level have been used. Further extension of the fragment length could increase the probability of model over-fitting and decrease its predictivity and DA. All

fragmentary descriptors (except simplexes) have been differentiated by atom individuality only.

The large number of simplex descriptors was generated on the basis of the SiRMS approach. The PLS-method [22–24] proved efficient with a great number of variables. The removal of highly correlated and constant descriptors, genetic algorithm (GA) [25], trend-vector method [5, 26] and the automatic variable selection (AVS) strategy based on interactive [23] and evolutionary [22] selection of variables was used for selection of descriptors in PLS.

It is well-known [24] that the PLS-equation can be represented as

$$Y = b_0 + \sum_{i=1}^{N} b_i x_i,$$

where Y is an appropriate activity; b_i is PLS regression coefficients; x_i is the ith descriptor value (the number of simplexes of ith type in the SiRMS); and N is the total number of descriptors. Using this equation, it is not difficult to carry out reverse analysis (interpretation of QSAR models) by using the SiRMS approach. The contribution of each j-atom (C_j) in a molecule can be defined as the ratio of the sum of PLS regression coefficients (b_i) of all simplexes that this atom contains (M), to the number of atoms in the simplex [9]:

$$C_j = \frac{1}{4}\sum_{i=1}^{M} b_i$$

The statistical fit of QSAR can be assessed in several statistical terms, e.g., the determination coefficient R^2, cross-validation determination coefficient Q^2, and the standard error of prediction S, etc. We used Leave-Group-Out for cross-validation where the number of groups = 7 [24].

As mentioned above, the results of QSAR analysis should be used to make predictions for compounds with unknown activity values (sometimes called "virtual screening"). With the purpose of analyzing the competence regions [27] of PLS models, "domain applicability" (DA) procedures, developed by us, were used. It can be characterized as follows: the distribution of molecules from the training set in a space of latent variables T_1–T_A (axes of coordinates) can be obtained from PLS. For each coordinate axis (T_1 and T_2 in our case) the average values A_{T1} and A_{T2} and corresponding root-mean-square deviations S_{T1} and S_{T2} were determined. Domain applicability represents an ellipsoid built from the training set molecule distribution center with the 1/2 semi-axies length of $3S_{T1}$ and $3S_{T2}$, respectively. Thus, if a new molecule from the prediction set is located outside the DA (square inside ellipsoid), its prognosis by corresponding QSAR model is not very reliable.

The initial work set (compounds **1**–**28**, Fig. 2) was divided between training and test sets. Approximately 20% [28] (six compounds) from different groups of activity were selected to the test set; the 22 remaining compounds were assigned to the training set. In order to check the predictive power of the obtained QSAR models, three different test sets were used during this analysis. Most likely, the usage of several (three is the enough minimum) test sets constructed by different principles and subsequent comparison and averaging (consensus modeling [29]) of the obtained results for the model validation is more preferable then the usage of only one set. In that way, the first test set has been constructed to maximize its diversity with training set, i.e. the compounds with the maximal dissimilarity were chosen. The second test set is created in order to minimize it diversity with training set, i.e. less dissimilar compounds from each group were removed. And the last test set has been chosen in random manner taking into account activity variation only. Procedure of test sets formation has been described in details in Ref. [10].

The experimentally determined (observed) values of toxicity of investigated structures [8] are presented in Table 1. Approximately 3,100 simplex descriptors were calculated during the initial stage of work.

Depending on their physicochemical characteristics (see above) the atoms were divided into several groups[1] corresponding to:

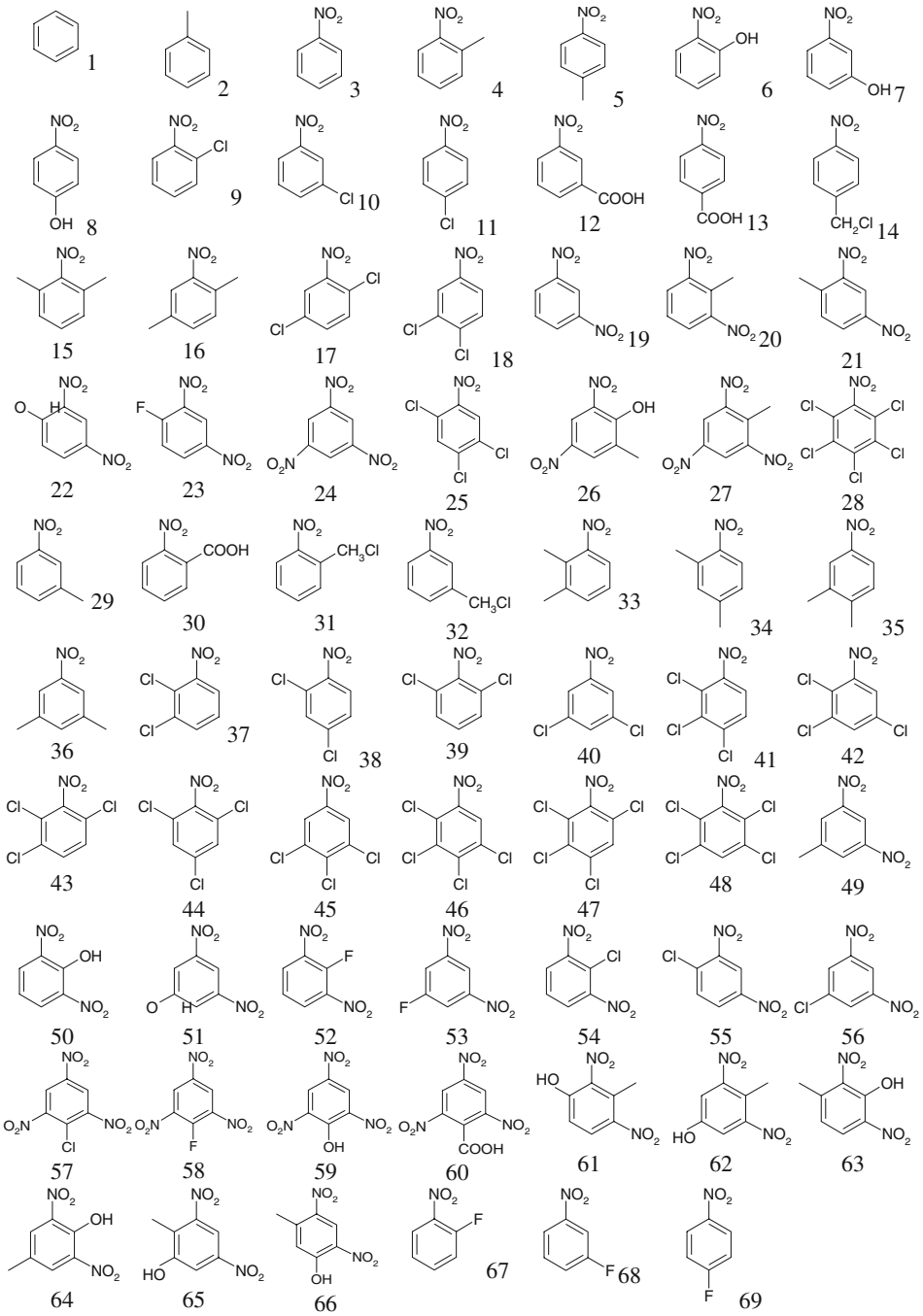
(i) partial charge ($A \le -0.1 < B \le -0.05 < C \le -0.01 < D \le 0.01 < E \le 0.05 < F \le 0.1 < G$);
(ii) lipophilicity ($A \le -1 < B \le -0.5 < C \le -0.1 < D \le 0.1 < E \le 0.5 < F \le 1 < G$);
(iii) refraction ($A \le 2 < B \le 3 < C \le 4 < D \le 6 < E \le 9 < F \le 12 < G$);
(iv) electronegativity ($A \le 1.5 < B \le 2 < C \le 2.5 < D \le 3 < E \le 3.5 < F$);
(v) Van der Waals attraction potential ($A \le 249 < B \le 447 < C \le 645 < D \le 844 < E \le 1042 < F \le 1240 < G$);
(vi) Van der Waals repulsion potential ($A \le 68{,}882 < B \le 130{,}603 < C \le 192{,}324 < D \le 254{,}045 < E \le 315{,}766 < F \le 377{,}487 < G$).

Results

As mentioned above, the 1D QSAR analysis was carried out as the preliminary stage of this study. The satisfactory two latent variables model ($R^2 = 0.82$; $Q^2 = 0.72$) was obtained using the PLS method with preliminary filtration of structural descriptors (vide supra) for all 28 molecules, indicating that the presence of certain combinations of atoms in the molecule substantially determines its toxicity.

[1] Number of groups is a tuning parameter of the model and can be varied.

Fig. 2 Investigated compounds



More detailed analysis of obtained information (Table 2) allowed observation of the trends described below.

The increase of nitrogen and/or oxygen atoms (N, O, OH_2 fragments) presence in molecule is important to the toxicity increase. Because the nitrogen atoms in the used training set belong only to the nitro group and oxygen could additionally be included in hydroxyl groups, we can conclude that these groups are essential factors in toxicity, i.e. increasing of the number of nitro and hydroxyl groups in the compound leads to the toxicity strengthening. On the other hand, the accumulation of nitrogen atoms (C_3HN_3 fragments) and, hence, nitro groups in nitroaromatic molecule decreases toxicity; i.e. the insertion of the third nitro group (N_3 means exclusively three nitro groups) into the aromatic ring interfere with toxicity of investigated compound. Summarizing the opposite effects of the mentioned fragments on toxicity one concludes that the toxicity of trinitroaromatics is not higher than toxicity of dinitroaromatics. Such results again indicate non-additive character of the simplex approach. In the molecules

Table 1 Observed and predicted values of investigated nitroaromatics toxicity

Structure	$-lgLD_{50}$ (rats), mmol/kg					
	Obs.	Pred. model 1	Pred. model 2	Pred. model 3	Pred. model 4	Consensus model 5
1	–1.86	−1.73	−1.76	−1.68	−1.65	−1.71
2[a,b,c]	–1.78	−1.95	−1.94	−1.68	−1.97	−1.89
3[a]	–0.69	−0.65	−0.69	−0.84	−0.66	−0.71
4	–0.81	−0.92	−0.94	−0.96	−0.86	−0.92
5[a]	–1.19	−1.02	−1.01	−1.08	−1.16	−1.07
6	–0.38	−0.30	−0.39	−0.45	−0.30	−0.36
7	–0.37	−0.34	−0.22	−0.15	−0.42	−0.28
8	–0.16	−0.15	−0.09	−0.32	−0.16	−0.18
9[a,c]	–0.23	−0.24	−0.56	−0.24	−0.52	−0.39
10[b]	–0.39	−0.36	−0.54	−0.52	−0.53	−0.49
11[a,c]	–0.43	−0.55	−0.59	−0.52	−0.35	−0.50
12	–0.61	−0.84	−0.89	−0.75	−0.92	−0.85
13[b,c]	–1.07	−0.98	−1.01	−0.78	−0.77	−0.89
14	–1.02	−0.93	−1.01	−1.04	−0.92	−0.98
15	–1.12	−1.11	−1.01	−1.08	−1.17	−1.09
16	–1.21	−1.10	−1.14	−1.16	−1.16	−1.14
17	–1.32	−1.29	−1.36	−1.24	−1.27	−1.29
18	–0.52	−0.49	−0.46	−0.64	−0.41	−0.50
19	0.31	0.19	0.23	0.23	0.15	0.20
20	–0.14	−0.09	−0.08	0.03	−0.22	−0.09
21[b]	–0.17	−0.13	−0.33	−0.12	−0.30	−0.22
22	0.41	0.50	0.56	0.57	0.62	0.56
23	0.57	0.50	0.46	0.57	0.56	0.52
24[b]	–0.11	−0.08	−0.08	−0.17	0.03	−0.08
25[b,c]	–0.67	−0.78	−0.61	−0.97	−0.66	−0.76
26[a,c]	0.52	0.51	0.13	0.47	0.38	0.37
27	–0.49	−0.53	−0.53	−0.57	−0.41	−0.51
28	–0.57	−0.63	−0.49	−0.43	−0.68	−0.56
29		−0.83	−0.89	−1.04	−1.13	−0.97
30		−0.92	−0.94	−0.50	−0.70	−0.77
31		−0.92	−0.94	−0.96	−0.84	−0.92
32		−0.83	−0.89	−1.04	−0.90	−0.92
33		−1.10	−1.14	−1.16	−1.14	−1.14
34		−1.06	−1.08	−1.18	−1.14	−1.12
35		−1.20	−1.21	−1.28	−1.63	−1.33
36		−0.84	−0.91	−1.24	−1.56	−1.14
37		−0.23	−0.37	−0.08	−0.61	−0.32
38		−0.53	−0.58	−2.20	−0.60	−0.98
39		0.05	−0.48	0.15	−0.60	−0.22
40		−0.84	−0.62	−2.63	−0.85	−1.24
41		−0.46	−0.40	−0.27	−0.49	−0.41
42		−1.47	−0.81	−2.52	−1.59	−1.60
43		−0.55	−0.34	−0.05	−0.74	−0.42
44		−0.64	−0.62	−4.08	−0.91	−1.56
45		−0.57	−0.75	−0.96	−0.47	−0.69
46		−0.85	−0.92	−0.75	−0.54	−0.77
47		−0.57	−0.27	−0.65	−0.79	−0.57

Table 1 continued

Structure	–lgLD$_{50}$ (rats), mmol/kg					
	Obs.	Pred. model 1	Pred. model 2	Pred. model 3	Pred. model 4	Consensus model 5
48		−0.72	−0.19	−0.65	−1.06	−0.66
49		0.01	0.02	−0.17	−0.19	−0.08
50		0.47	0.68	0.44	0.57	0.54
51		−0.29	−0.08	0.43	−0.25	−0.05
52		0.46	0.48	0.44	0.51	0.47
53		−0.27	−0.08	0.22	−0.25	−0.10
54		0.37	0.35	0.55	0.13	0.35
55		0.02	−0.05	0.27	0.26	0.13
56		−0.44	−0.35	−0.57	−0.42	−0.45
57		−0.42	−0.36	−0.36	0.56	−0.15
58		0.26	0.52	0.04	1.01	0.46
59		0.23	0.71	0.04	1.08	0.52
60		−0.46	−0.53	−0.57	−0.47	−0.51
61		0.22	0.33	0.37	0.36	0.32
62		−0.65	−0.39	0.23	−0.60	−0.35
63		0.17	0.21	0.09	0.23	0.18
64		0.29	0.41	0.04	0.34	0.27
65		−0.17	−0.46	0.13	−0.54	−0.26
66		0.49	0.08	0.27	0.22	0.27
67		−0.30	−0.49	−0.45	−0.28	−0.38
68		−0.34	−0.22	−0.25	−0.40	−0.30
69		−0.20	−0.09	−0.32	−0.19	−0.20

Notes: Training set molecules have been marked by bold font

[a] This molecule has been used for model 2 test set

[b] This molecule has been used for model 3 test set

[c] This molecule has been used for model 4 test set

investigated, an increase in number of carbon atoms (C_4 and C_6Cl fragments) was associated with a decrease in toxicity. Introduction of alkyl, specifically methyl and chlorine alkyl (for example, CH_2Cl) substituents leads to a decrease in toxicity, whereas introduction of fluorine atoms into the benzene ring (H_2F and CH_2F fragments) results in an increase of toxicity.

Table 2 1D QSAR model descriptors and their relative influence on toxicity change

Descriptor (differentiation)	Corresponding atom combination	Relative contribution in activity (%)
Toxicity increase		
E (electronegativity)	N or O	28
A_2B (Van der Waals attractive potential)	H_2O or H_2F	13
CFH_2 (element)	CFH_2	12
Toxicity decrease		
C_4 (element)	C_4	18
C_3N_3H (element)	C_3N_3H	14
AB_2D_2 (atomic refraction)	HC_2Cl_2, $HCNCl_2$, HN_2Cl_2, OC_2Cl_2, $OCNCl_2$, ON_2Cl_2	8
C_6Cl (element)	C_6Cl	7

As expected, the transition to the 2D level of molecular structure representation allows development of much more adequate QSAR models (Tables 1, and 3; Fig. 3). The obtained models satisfy all the requirements of QSAR Setubal Principles [30]. They have high statistical indexes: adequacy-of-fit, robustness, and predictivity ($R^2 = 0.96$–0.98; $Q^2 = 0.84$–0.93; $R^2_{test} = 0.89$–0.92). Interestingly, the training set decrease, due to the detachment of six molecules and their insertion into the test set, did not appreciably worsen the statistical characteristics of models 2–4 (Table 3). The consensus model 5 (averaged-by-models 1–4) prediction quality close to model 1 (all 28 molecules included). All the variables from every described model were sourced in the Supplementary information (Supplement.xls).

Thus, taking high statistical characteristics of obtained 2D QSAR models into account, the transition to the models of the next levels of HiT QSAR technology has been

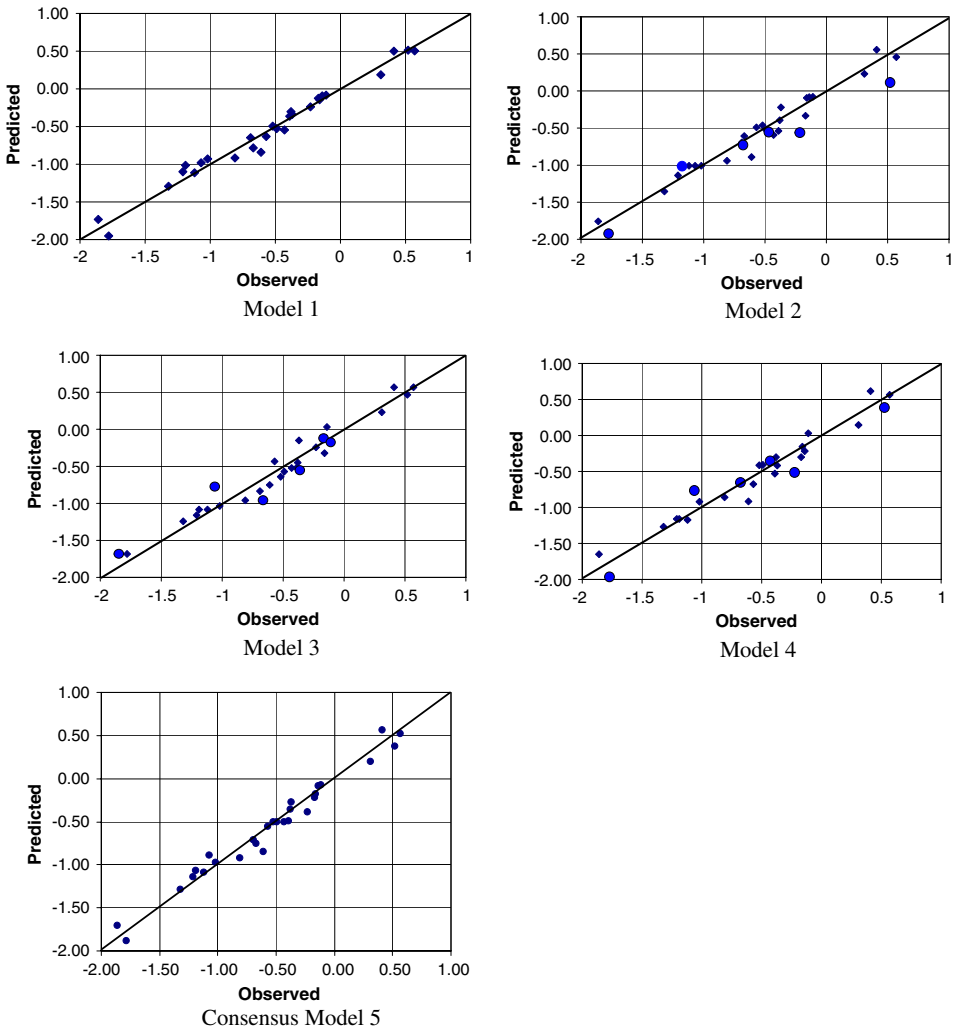


Fig. 3 Observed versus predicted diagrams for obtained 2D QSAR models. (training set molecules are rhomb-marked; test set molecules are sphere-marked)

considered inexpedient. The second part of the study involved application of this model to the other set of nitrobenzene derivatives having a similar chemical structure. Therefore, the set of compounds that includes different combinations of substituents in an aromatic ring was represented in Fig. 2 and Table 1. All of these compounds are nitroaromatics, i.e. belong to the same chemical class. The whole set of substituents used for their design is represented in the initial training set. Toxicity values for 41 structures from the new set of compounds (**29–69**) were predicted. Later, their link to domain applicability was estimated via model 1. As expected (Fig. 4), almost all investigated molecules, except compounds **58** and **59**, located inside the space of structural features fall into

Table 3 Statistical characteristics of the obtained 2D QSAR models

Model	R^2	Q^2	R^2_{test}	S(trs)	S(cv)	S(ts)	A	N	M	Outliers
1	0.98	0.91	–	0.10	0.19	–	2	18	28	No
2	0.96	0.93	0.89	0.11	0.16	0.26	2	8	22	No
3	0.96	0.84	0.91	0.12	0.26	0.20	2	8	22	No
4	0.96	0.85	0.92	0.13	0.24	0.22	2	12	22	No
5	0.97	–	–	0.10	–	–	–	–	–	No

where, R^2—correlation coefficient; Q^2—cross validation correlation coefficient; R^2_{test}—correlation coefficient for test set; S(trs)—standard error of a prediction for training set; S(cv)—standard error of prediction for training set in cross validation terms; S(ts)—standard error of a prediction for test set; A—number of PLS latent variables; N—number of descriptors; M—number of molecules in training set

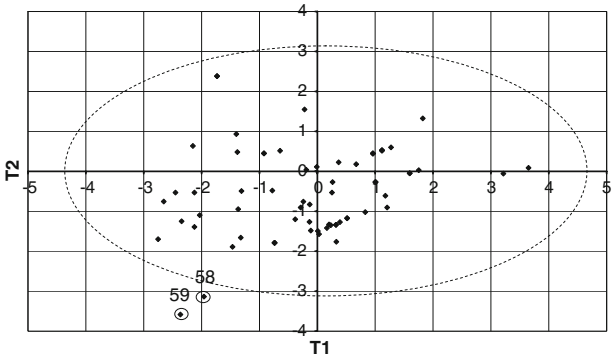


Fig. 4 Distribution of investigated molecules and QSAR model 1 domain applicability in the space of structural parameters (T1, T2)

domain applicability ellipsoid (see Materials and methods). Probably, it is caused by poor representation of trinitroaromatics in initial training set (only two compounds **24** and **27).** On the other hand there is only one Fluorine-containing compound (dinitroaromatic **23**) in a training set, and a lack of prediction reliability for monofluoro mononitroaromatic compounds **67–69** could be expected. In reality mentioned compounds are situated inside model DA, i.e. they have been predicted reliably. It can be easily explained by described above benefits of the SiRMS model—usage of various kinds of atom's differentiations in simplexes, including different physico-chemical properties of formers. It allows new compounds to be more similar to existing one than one can conclude just by comparing their structural formulas. Overall reliability of our predictions for new compounds is additionally confirmed by the following: We succeeded in finding additional information for some high-toxic molecules that were not included in the training set (United States National Library of Medicine http://sis.nlm.nih.gov/chimical.html). The data presented in Table 4 indirectly confirm the results of toxicity prediction.

Unfortunately, we cannot make a direct comparison of our results with these data, because they have been developed under different experimental conditions. The quality estimation of toxicity for all of these molecules was additionally carried out by the ToxTree computer program [35]. This software implements the Cramer and Verhaar classification schemes, and its development was commissioned by the European Chemicals Bureau. The highest level of toxicity (class III) was observed for all compounds.

Discussion

Based on the results obtained in the framework of Model 1, new information can be obtained as to nitroaromatic toxicity. Assuming reliable data on oral rat toxicity of 69 related nitro-aromatic compounds, some trends in the dependencies of toxicity on the structure of nitro-aromatic molecules can be analyzed. As follows from Fig. 5a, in practically all cases, an aromatic ring with nitro group(s) contributes positively to toxicity, even though this contribution varies widely-depending on the nature and number

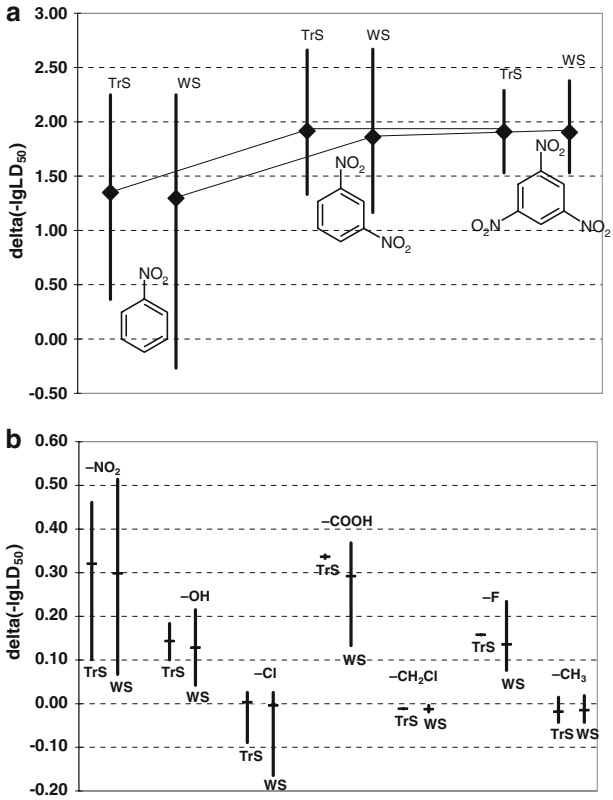


Fig. 5 Molecular fragment contributions $\Delta(-\lg LD_{50})$ to toxicity change: (**a**) nitroaromatic fragments; (**b**) substituents in benzene ring (TrS—training set, WS—work set)

of other substituents. Clearly, the contribution range is wider for the work set than for the training set. The increase in toxicity is noticeable during the transition from mononitrobenzene derivatives to dinitrobenzene ones. A subsequent transition to substituted trinitrobenzenes does not appreciably result in the increase of toxicity (Compounds **58** and **59** were not considered in this case). This corresponds quite well to the information obtained at the preliminary stage (1D QSAR).

More extended analysis of substituents on aromatic ring influence on oral rat toxicity is illustrated in a Fig. 5b. It can be concluded that nitro groups, hydroxyl, carboxyl, and fluorine promote an increase in toxicity, corresponding well to results of the 1D QSAR analysis. Both potential donor (substituents of the first species—OH, F) and

Table 4 Toxicity data from United States National Library of Medicine for some investigated nitroaromatics

Structure	CAS number	Information about toxicity
49	618-85-9	LD_{50} (rat, oral, mg/kg) = 216 or $-\lg LD_{50}$ (rat, oral, mmol/kg) = −0.07
50	573-56-8	LD_{50} (rat, intraperitoneal, mg/kg) = 38 [31] or $-\lg LD_{50}$ (rat, intraperitoneal, mmol/kg) = 0.69
55	97-00-7	LD_{50} (rat, intraperitoneal, mg/kg) = 280 [32] or $-\lg LD_{50}$ (rat, intraperitoneal, mmol/kg) = −0.14; LD_{50} (rat, oral, mg/kg) = 640 [32] or $-\lg LD_{50}$ (rat, oral, mmol/kg) = −0.50
59	88-89-1	LD_{50} (rat, oral, mg/kg) = 200 [33] or $-\lg LD_{50}$ (rat, oral, mmol/kg) = 0.06
64	609-93-8	LD_{50} (mouse, intraperitoneal, mg/kg) = 24.8 [34] or $-\lg LD_{50}$ (mouse, intraperitoneal, mmol/kg) = 0.90

potential acceptor (substituents of the second species—NO_2, COOH) groups contribute positively to toxicity. Methyl and chloromethyl groups slightly decrease activity. Chlorine, as a substituent to an aromatic ring, is ambiguous in its influence on toxicity.

In considering toxicity changes within the separate groups of substituted mono-, di-, and trinitrobenzenes (Fig. 6a–c), it is evident from Fig. 6a, that methyl and chloromethyl derivatives of nitrobenzene are less toxic than unsubstituted nitrobenzene, whereas insertion of chlorine, hydroxyl or fluorine into nitrobenzene promotes toxicity. There are no simple trends for the influence of isomerism on the toxicity of nitrobenzene derivatives. However in most cases, *meta*-isomers are more and *para*-isomers are less toxic. In addition, for substituted 1,3-dinitrobenzenes (Fig. 6b), the influence of isomerism (A, B, C isomers) is more noticeable than in the previous case. In most cases, C-isomers are less and A-isomers are more toxic. The tendency toward toxicity increase in the group CH_3, Cl, OH, F is also observable. Similarly, insertion of chlorine, fluorine or hydroxyl in 1,3,5-trinitrobenzene increases toxicity (Fig. 6c). As noted before, information for compounds **58** and **59** is not sufficiently reliable. At the same time, 2,4,6-trinitrobenzoic acid is less toxic than the other species from the given group of compounds.

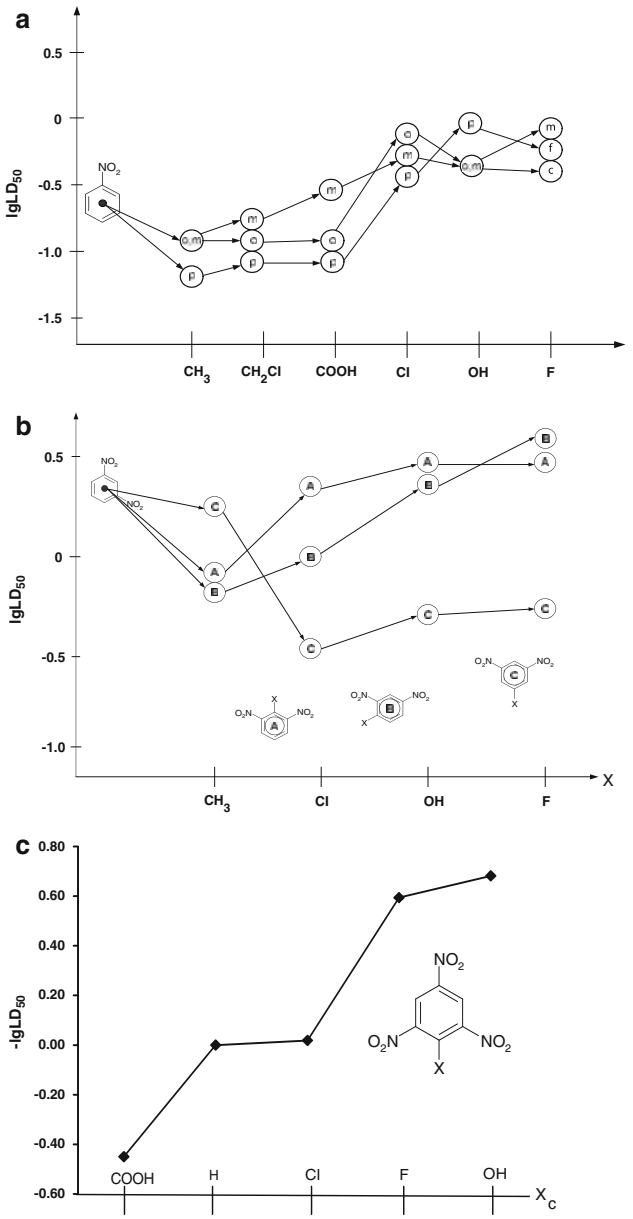
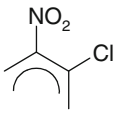
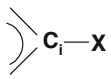


Fig. 6 Toxicity change for isomers (o–*ortho*, m–*meta*, p–*para*) of monosubstituted: (**a**) nitrobenzene; (**b**) dinitrobenzene; (**c**) trinitrobenzene

Since the chlorine derivatives are more comprehensively represented in the initial work set and since, as mentioned above, they deserve more careful study, we have made a more detailed investigation of their influence on toxicity. "Evolution" of toxicity change for this group of compounds is represented in Fig. 7. The largest toxicity changes are observed for nitrobenzene dichloro derivatives. An addition of new substituents (new chlorine atoms) decreases such changes. Presumably, the accumulation of chlorine atoms in the benzene ring results in cancellation of their influence on toxicity. Both the most toxic chlorine-substituted nitrobenzenes and the least toxic ones contain the chlorine atom in the position *ortho* to the nitro group. We expect that the  fragment is "toxicophore" [36, 37] and the other substituents (atoms of chlorine)—"modulators" that may increase or decrease toxicity.

Finalizing the analysis of chlorine substituents, we developed another way to describe the influence of chlorine atoms on toxicity values of chlorine derivatives of mono nitrobenzene (compounds **3**, **9**, **37**, **39**, **42**). In Fig. 8, the toxicity of each molecule is represented as six separate contributions,



where C_i — ith (i = 1–6) C atom on benzene ring and X—corresponds to substituent (H, Cl, NO_2). In this figure for compounds mentioned above theirs contributions to toxicity have been represented by irregular hexagons drawn by solid lines. Vertexes of these hexagons are marked by signs which are individual for each compound (◇–**3**; □–**9**; Δ–**37**; ▲–**39**; •–42). Positions of these vertexes on radial dashed lines outgoing from the center correspond to the contributions to toxicity values. Stepwise contributions to toxicity variation (0.1; 0.2;...; 0.6) has been represented on the Fig. 8 as the dashed regular hexagons.

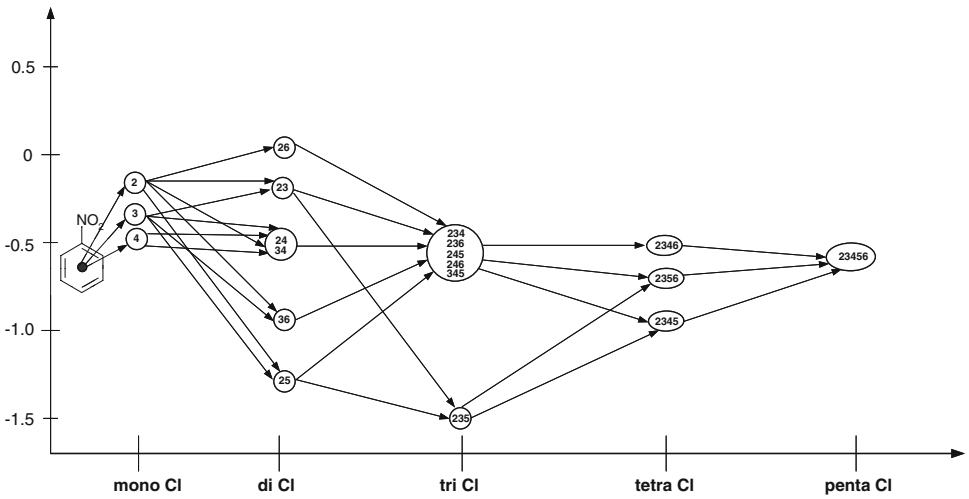


Fig. 7 Evolution of nitrobenzene chloroderivative toxicity change (digits in circles are the corresponding positions of substituents in benzene ring)

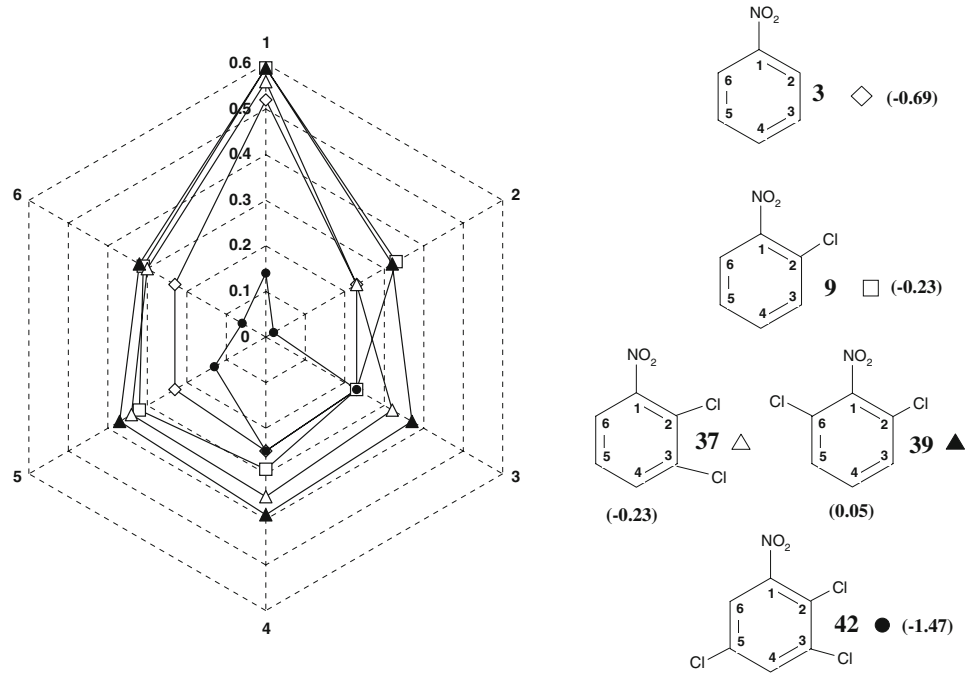


Fig. 8 The analysis of structural fragment influence on toxicity for some chlorosubstituted nitrobenzenes (compound number marked by bold; digits in brackets are the toxicity values)

We analyzed the influence of sequential insertion of substituents into the benzene ring as to toxicity. We found that insertion of a chlorine atom in *ortho* position to the nitro group (compound **9**) leads to a toxicity increase in comparison with nitrobenzene (compound **3**). This effect is not limited only to the chlorine atom. Actually, the contributions to toxicity of all other atoms are augmented (except C–H bond in *ortho* position to the C–Cl bond, Fig. 8). Insertion of one more chlorine in *ortho* position to the previous one (compound **37**) has only a small effect on toxicity. Although the new C–Cl bond (position 3) increases the toxicity of a molecule, the contributions of the nitro group and other C–Cl fragment (position 2) have been diminished. The influence of C–H bonds in positions 5 and 6 has virtually no effect on toxicity. The C–H bond in position 4 slightly increases toxicity. Thus, in spite of the redistribution of influence on toxicity between different fragments of molecule **37** (Fig. 8), the toxicity of this whole compound hardly changes compared to compound **9**.

The situation is completely different (Fig. 8) for the other isomer (compound **39**). Substitution of the hydrogen atom by chlorine in position 6 has practically no effect on this fragment's contribution to toxicity as well as to the toxicity of the nitro group in comparison with the compound **9**. However, the contributions of C–H bonds in positions 3–5 substantially increase toxicity. The last factor determines the increase in toxicity for compound **39**.

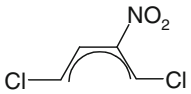
A dramatic change in toxicity was predicted for compound **42**. Substitution of hydrogen by chlorine in

position 5 of molecule **37** results in substantial lowering of toxicity. It is interesting that this result has the effect of diminishing contributions to the toxicity of all analyzed fragments (see Fig. 8). The influence on toxicity change of the C–Cl bond in position 2 is especially noticeable. It is interesting that this fragment is situated in *para* position to the new additional chlorine atom. Thus, it is possible to assume an important role for the chlorine atom as the substituent of the first species (electron-donating) and its polar influence on the aromatic system. Since we do not have any conclusive proofs about the mechanism of chlorinated nitro compounds action, one can conclude that the nitro groups making the halogen more susceptible to SNAr reactions or the halogens are facilitating the reduction of the nitro groups. Both of the mentioned mechanisms are probable.

Thus, the toxicity of all of considered compounds has been substantially affected by interference of the substituents.

Summarizing, the obtained results indicate that the influence on toxicity of substituents to the benzene ring is considerably non-additive, Fig. 8. In addition, the presented toxicity trends do not follow the simple donor acceptor rules known in organic chemistry (Substituent polar influence is discussed below).

As follows from our study, the 1,3-dinitrobenzene derivatives containing a fluorine or hydroxyl group (compounds **50**, **52**, and **66**) are the most toxic. Except for benzene and toluene, which do not possess nitro groups, the least toxic compounds are methyl derivatives of nitrobenzene (compounds **5**, **33**–**35**) and chlorine derivatives of nitrobenzene (compounds **17** and **42**) that contain the following fragment



As shown earlier [11, 12, 38], it is possible within the framework of the SiRMS approach to estimate the relative influence of different physical and chemical factors on toxicity change and to divide the total toxicity into contributions based on electrostatic, hydrophobic, and Van der Waals interactions of toxicants with the proper biological target. Results of the application of this technique suggest that all three factors contribute to toxicity in approximately equal amounts: electrostatic 34%, hydrophobic 33%, and Van der Waals 33%. It can also be concluded that the hydrophobic characteristics of nitroaromatics regulate their transport function (delivery to the biological target). Charge distribution in toxicants (electrostatic characteristics), likely determines the degree of its interaction with a biological target as well as the stability of the reaction species (ion and/or radical). Finally, Van der Waals forces can be an important factor of toxicant–target interaction.

Also, in spite of substituent polar effects undoubtedly being important, their influence does not follow well known rules of classical organic chemistry since we did not find meaningful Hammett σ and σ^+ constants-toxicity value dependences ($R = 0.06$–0.28) for the training set and its parts.

Conclusion

The HiT QSAR SiRMS approach proves to be a new and quite efficient tool for analyzing nitroaromatic toxicity. Among contributions analyzed in this work are electrostatic, hydrophobic and Van der Waals interactions of toxicants to biological targets. In particular, it was found that in most cases, insertion of fluorine and hydroxyl groups into nitroaromatics increases toxicity, whereas insertion of a methyl group has the opposite effect. The influence of chlorine on toxicity is ambiguous. Insertion of chlorine in *ortho* position to the nitro group leads to substantial increase in toxicity, whereas the second chlorine atom (in *para* position to the first one) results in considerable decrease in toxicity. It is also concluded that the influence of substituents is substantially non-additive. Finally, one needs to consider other QSAR descriptors for a more detailed analysis of chemical and biological mechanisms of nitroaromatic toxicity. These could include parameters obtained from quantum-chemical calculations, macro-parameters such as solubility, partition coefficients, etc. The obtained results should be verified by experimental study in order to provide feed-back on the accuracy of the proposed new technique.

Acknowledgments We appreciate and thank Dr. John Cullinane, ERDC Technical Director for Military Environmental Engineering and Science and Dr. Richard E. Price, Environmental Division Chief, ERDC, for their support and encouragement. Permission was granted by the Chief of Engineers to publish this information. This content does not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred. The authors are also thankful to Drs. B. Rasulev, A. Toropov and O. Isaev for fruitful discussion and useful comments.

Funding US Army Environmental Quality Technology Program (grant W912Z-04-P-139) administrated by the US Army Engineer Research and Development Center; US Army Engineer Research and Development Center CMCM program (grant 2T346GM007672-25A1). The authors confirm independence from the sponsors; the content of the article has not been influenced by the sponsors.

References

1. Johnson LR, Devenport R, Balbach HDJS (1994) Phototoxicology. Comparative toxicity of trinitrotoluene and aminodinitotoluenes

to Daphnia magna, Dugesia dorotocephala, and sheep erythrocytes. Ecotoxicol Environ Saf 27:44–49
2. Donlon BA, Razo-Flores E, Field JA, Lettinga G (1995) Toxicity of N-substituted aromatics to acetoclastic methanogenic activity in granular sludge. Appl Environ Microbiol 61:3889–3893
3. Slater EC (1962) Mechanism of uncoupling of oxidative phosphorylation by nitrophenols. Comp Biochem Physiol 4:281–301
4. Nemeikaitė-Čėnienė A, Miliukienė V, Šarlauskas V, Maldutis E, Čėnas N (2006) Chemical aspects of cytotoxicity of nitroaromatic explosives: a review. Chemija 17:34–41
5. Kuz'min VE, Artemenko AG, Kovdienko NA, Tetko IV, Livingstone DJ (2000) Lattice model for QSAR studies. J Mol Model 6:517–526
6. Agrawal WK, Khadikar PV (2001) QSAR prediction of toxicity of nitrobenzenes. Bioorg Med Chem 9:3035–3040
7. Cronin MTD, Schultz TW (2001) Development of quantitative structure-activity relationships for the toxicity of aromatic compounds to Tetrahymena pyriformis: comparative assessment of the methodologies. Chem Res Toxicol 14:1284–1295
8. Isayev O, Rasulev B, Gorb L, Leszczynski J (2006) Structure-toxicity relationships of nitroaromatic compounds. Mol Divers 10:233–245
9. Kuz'min VE, Artemenko AG, Polischuk PG, Muratov EN, Hromov AI, Liahovskiy AV, Andronati SA, Makan SY (2005) Hierarchic system of QSAR models (1D–4D) on the base of simplex representation of molecular structure. J Mol Model 11(6):457–467
10. Kuz'min VE, Artemenko AG, Muratov EN (2007) Hierarchical QSAR technology on the base of Simplex representation of molecular structure. J Comput Aided Mol Des 22 (doi: 10.1007/s10822-008-9179-6)
11. Kuz'min VE, Artemenko AG, Lozitska RN, Fedtchouk AS, Lozitsky VP, Muratov EN, Mescheriakov AK (2005) Investigation of anticancer activity of macrocyclic Schiff bases by means of 4D-QSAR based on simplex representation of molecular structure. SAR QSAR Environ Res 16(3):219–230
12. Kuz'min VE, Artemenko AG, Lozitsky VP, Muratov EN, Fedtchouk AS, Dyachenko NS, Nosach LN, Gridina TL, Shitikova LI, Mudrik LM, Mescheriakov AK, Chelombitko VA, Zheltvay AI, Vanden Eynde J-J (2002) The analysis of structure- anticancer and antiviral activity relationships for macrocyclic pyridinophanes and their analogues on the basis of 4D QSAR models (simplex representation of molecular structure). Acta Biochim Pol 49:157–168
13. Artemenko AG, Muratov EN, Kuz'min VE, Kovdienko NA, Hromov AI, Makarov VA, Riabova OB, Wutzler P, Schmidtke M (2007) Identification of individual structural fragments of N,N′-(bis-5-nitropyrimidyl)dispirotripiperazine derivatives for cytotoxicity and antiherpetic activity allows the prediction of new highly active compounds. J Antimicrob Chemother 60(1):68–77
14. Kuz'min VE, Artemenko AG, Muratov EN, Volineckaya IL, Makarov VA, Riabova OB, Wutzler P, Schmidtke M (2007) Quantitative structure-activity relationship studies of [(biphenyloxy)propyl]isoxazole derivatives—human rhinovirus 2 replication inhibitors. J Med Chem 50:4205–4213
15. Muratov E, Artemenko A, Kuz'min V, Konup I, Konup L, Kotlyar S, Kamalov G, Fedtchuk ANM (2006) Computational design of the new antimicrobials based on the substituted crown ethers. Clin Microbiol Infec 12(4):1558
16. Muratov EN, Artemenko AG, Kuz'min VE, Lozitsky VP, Fedchuk AS, Lozitska RN, Boschenko YA, Gridina TL (2005) Investigation of anti-influenza activity using hierarchic QSAR technology on the base of simplex representation of molecular structure. Antiviral Res 65(3):A62–A63
17. Zhang S, Golbraikh A, Oloff S, Kohn H, Tropsha A (2006) A novel automated lazy learning QSAR (ALL-QSAR) approach: method development, applications, and virtual screening of chemical databases using validated ALL-QSAR models. J Chem Inf Model 46:1984–1995
18. Kuz'min VE (1995) About homo- and heterochirality of dissymetrical tetrahedrons (chiral simplexes). Stereochemical tunneling. Zh Strucur Khim 36:873–878
19. Jolly WL, Perry WB (1973) Estimation of atomic charges by an electronegativity equalization procedure calibration with core binding energies. J Am Chem Soc 95:5442–5450
20. Wang R, Fu Y, Lai L (1997) A new atom–additive method for calculating partition coefficients. J Chem Inf Comp Sci 37: 615–621
21. Rappe AK, Casewit CJ, Colwell KS, Goddart WA, Skiff WM (1992) UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. J Am Chem Soc 114:10024–10035
22. Kubinyi H (1996) Evolutionary variable selection in regression and PLS analyses. J Chemometrics 10:119–133
23. Lindgren F, Geladi P, Rannar S, Wold S (1994) Interactive variable selection (IVS) for PLS. Part 1: theory and algorithms. J Chemometrics 8:349–363
24. Rannar S, Lindgren F, Geladi P, Wold S (1994) A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: theory and algorithm. J Chemometrics 8:111–125
25. Hasegawa K, Miyashita Y, Funatsu K (1997) GA strategy for variable selection in QSAR studies: GA-based PLS analysis of calcium channel antagonists. J Chem Inf Comput Sci 37:306–310
26. Carhart RE, Smith DH, Venkataraghavan R (1995) Atom pairs as molecular features in structure—activity studies. Definition and application. J Chem Inf Comput Sci 25:64–73
27. Jaworska J, Nikolova-Jeliazkova N, Aldenberg T (2005) QSAR applicability domain estimation by projection of the training set in descriptor space: a review. Altern Lab Anim 33:445–459
28. Gramatica P (2004) Evaluation of different statistical approaches for the validation of quantitative structure–activity relationships. ECVAM, Ispra, p 177
29. Tropsha A, Gramatica P, Gombar VK (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb Sci 22:69–77
30. QSAR; Expert; Group (2004) The report from the expert group on (Quantitative) Structure–Activity Relationships [(Q)SARs] on the principles for the validation of (Q)SARs.; 49; Organisation for Economic Co-operation and Development, Paris, p 206
31. Harvey DG (1959) On the metabolism of some aromatic nitro compounds by different species of animal. III. The toxicity of the dinitrophenols, with a note on the effects of high environmental temperatures. J Pharm Pharmacol 11:462–474
32. Sziza M, Magos L (1959) Toxicologic investigation of some aromatic nitrogen compounds approved for use in Hungarian industry. Arch Gewerbepathol Gewerbehyg 17:217–226
33. Wyman JF, Serve MP, Hobson DW, Lee LH, Uddin DE (1992) Acute toxicity, distribution, and metabolism of 2,4,6-trinitrophenol (picric acid) in Fischer 344 rats. J Toxicol Environ Health 37(2):313–327
34. Harvey DG (1953) The toxicity of the dinitrocresols. II. The formation and toxic properties of some nitrocompounds derived from meta- and para- cresols. J Pharm Pharmacol 8:497–510
35. Cramer GM, Ford RA, Hall RL (1978) Estimation of toxic hazard—a decision tree approach. J Cosmet Toxicol 16:255–276
36. Hakimelahi GH, Khodarahmi GA (2005) The identification of toxicophores for the prediction of mutagenicity, hepatotoxicity and cardiotoxicity. J Iran Chem Soc 2:244–267
37. Kazius J, McGuire R, Bursi R (2005) Derivation and validation of toxicophores for mutagenicity prediction. J Med Chem 48(1): 312–320
38. Kuz'min VE, Artemenko AG, Muratov EN, Lozitsky VP, Fedchuk AS, Lozitska RN, Boschenko YA, Gridina TL (2005) The hierarchical QSAR technology for effective virtual screening and molecular design of the promising antiviral compounds. Antiviral Res 65(3):A70–A71