Structure–Toxicity Relationships of Nitroaromatic Compounds

ARTICLE *in* MOLECULAR DIVERSITY · JUNE 2006

Impact Factor: 1.9 · DOI: 10.1007/s11030-005-9002-4 · Source: PubMed

4 AUTHORS, INCLUDING:

Olexandr Isayev
University of North Carolina at Chapel Hill
25 PUBLICATIONS **300** CITATIONS
SEE PROFILE

Bakhtiyor Rasulev
North Dakota State University
76 PUBLICATIONS **845** CITATIONS
SEE PROFILE

Molecular Diversity (2006) 10: 233-245
DOI: 10.1007/s11030-005-9002-4

Full-length paper

Structure-toxicity relationships of nitroaromatic compounds

Olexandr Isayev[1], Bakhtiyor Rasulev[1,2], Leonid Gorb[1] & Jerzy Leszczynski[1,*]
[1]*Computational Center for Molecular Structure and Interactions, Jackson State University, Jackson, Mississippi, USA;*
[2]*Institute of the Chemistry of Plant Substances AS RUz, Tashkent, Uzbekistan*
*(*Author for correspondence, E-mail: jerzy@ccmsi.us, Tel: 601-979-7824, Fax: +601-979-7823)*

Received 23 June 2005; Accepted 11 October 2005

Key words: nitroaromatic compounds, QSAR, toxicity, LD_{50}, genetic algorithm, topological descriptors, quantum-chemical descriptors, electronic properties, DFT

Summary

The toxicity data of 28 nitroaromatic compounds (nitrobenzenes and, for comparison, benzene and toluene) related to a 50% lethal dose concentration for rats (LD_{50}) were used to develop quantitative structure-activity relationships (QSARs).

A genetic algorithm and multiple regression analysis were applied to select the descriptors and to generate the correlation models. The obtained equations consist of one to three descriptors. A number of molecular descriptors was obtained from HF/6-31G(d) and DFT (B3LYP/6-311+G(d, p)) level calculations. The calculated molecular geometry and electronic properties were evaluated by comparison with the available experimental data (where applicable). All parameters obtained at the B3LYP/6-311+G(d, p) level and the topological descriptors derived from this geometry were found to be reliable, except for dipole moment, due to the large uncertainty of its estimation.

Satisfactory relationships were observed for the one-parameter structure-toxicity models between topological (X5Av, Ms) and quantum-chemical (E_{LUMO}) descriptors. For better predictability two- and three-parameter QSAR analyses were performed. These analyses resulted in much better equations with correlation coefficient values $r = 0.872 - 0.924$. These models have been obtained with a set of topological, fragment and quantum-chemical descriptors (Ms, PCR, PCD, BELe1, C-026 and E_{LUMO}).

The toxicity of nitroaromatic compounds appears to be governed by a number of factors, such as the number of nitrogroups, the electrotopological state, the presence of certain fragments and the electrophilicity/reactivity parameter (E_{LUMO}). Nitrobenzenes exhibited electrophilic reactivity (as was shown by correlation of the toxicity with the energy of the lowest unoccupied orbital, E_{LUMO}).

The toxicity LD_{50} parameter for rats has been utilized for the first time for QSAR analysis of nitrobenzenes. The predictive ability of the models is determined by a cross-validation "leave-one-out" method.

Abbreviations: HOMO, Highest Occupied Molecular Orbital; LUMO, Lowest Unoccupied Molecular Orbital; SOMO, Single-Occupied Molecular Orbital; DFT, density functional theory; B3LYP, Becke threeparameter hybrid functional combined with Lee–Yang–Parr correlation functional; AM1, Austin Model 1 semiempirical method; GA, Genetic Algorithm; MLRA, Multiple Linear Regression Analysis method; HF, Hartree-Fock, UHF, unrestricted Hartree-Fock; HLG, HOMO-LUMO gap; QSAR, quantitative structure-activity relationships; IP, ionization potential; LD_{50}, lethal dose which causes the death of 50% (one half) of a group of test animals

Introduction

N-substituted aromatic compounds (NACs), such as nitrobenzenes, nitrophenols, aminophenols, and aromatic amines, are released into the biosphere almost exclusively from anthropogenic sources. Some compounds are produced by incomplete combustion of fossil fuels; others are used as synthetic intermediates, dyes, pesticides, and explosives [1]. Nitrobenzene is produced annually on the order of 225,000 metric tons, and it has been estimated that as much as 9,000 metric tons of nitrobenzene are discharged annually into natural waters [2].

The presence of these aromatic xenobiotics in the environment may create serious public health and environmental problems. Some of these compounds have mutagenic or carcinogenic activity and may bioaccumulate in the food chain [3]. Many nitroaromatics have also been shown to be toxic or mutagenic to microorganisms [4]. The toxicity has

been attributed to the fact that nitrophenols act as uncoupling agents in oxidative phosphorylation [5]. The nature and the degree of the aromatic substitution were concluded to have a profound effect on the toxicity of the NACs compounds [6]. In addition, the electron withdrawing nitro groups provide the compounds resistance to typical advanced oxidation destruction techniques as well as generally protect the compound from chemical and biological attack. Therefore, an understanding of the mechanisms of nitrocompounds chemical transformation is very important from both fundamental and practical points of view.

For the last few decades, nitrobenzenes have been the object of considerable interest of chemists working in the field of environmental science. For many years, QSARs have been efficiently used for the study of toxicity mechanisms of various reactive chemicals. This is a powerful technique, which quantitatively relates variations in biological activity to changes in molecular properties. In other words, it attempts to link activity data with descriptors chosen via identification of the "rules" that can be further used to monitor chemical transformations in the environment or in living organisms. The current status of the achievements of the QSAR technique have been discussed in a paper [7] which contains a review concerning QSAR analysis of experimental data on biotransformation and toxicity. The attempts have been made to define parameters for the QSAR studies on the basis of quantum mechanical computer calculations. In particular, there have been described the role in defining the toxicity properties of xenobiotics of such quantum-chemical parameters as HOMO and LUMO energies, heat of formations, ionization potentials, and octanol-water partition coefficients have been described.

Recently, Katritzky et al. [8] performed extensive toxicological investigation of the aquatic toxicity of ciliate *Tetrahymena pyriformis* for the series of nitrogen-containing aromatic compounds. They used the most extended set of descriptors that includes geometrical, topological, constitutional parameters, adopted from 2D and 3D structures of the target molecules, and a set of electronic properties, that has been obtained from AM1 quantum chemical calculations. Several models which include a different number of parameters (from two up to five) per model were obtained. One of the obtained regression models (the five-parameter model) that included topological (e.g., total molecular surface area) and quantum chemical (E^{SOMO}) descriptors was characterized by correlation coefficient R^2 equal to 0.815 in the case of a dataset for 97 nitrobenzenes (the model consists of 5 parameters). In other studies, such empirical parameters as LogP (octanol-water partition coefficient) were considered together with theoretical topological indexes, such as PI, MRI, Sz and others [9, 10]. In some cases only one-parametric models with quantum-chemical descriptors like E_{LUMO}, most positive or negative atomic charges where utilized for explanation of the toxicity of nitrobenzenes [9–13]. However, in the reported models for nitrobenzenes toxicity, these parameters were developed to explain growth inhibition for algae (i.e. *in vitro* toxicity to the algae) [8–13]. It is obvious that the toxicity specific to the algae does not take into account the number of factors occurring in mammals.

To study how the results presented in [8–12] may differ from the investigation that applies toxicity parameter such as LD_{50} for rats, we utilized the widest set (to date) of topological and quantum chemical descriptors, which totally cover more then 1500 characteristics. We expect that such an extension will allow more accurate predictions, which will better represent the experimental toxicity data. We also believe, that the obtained values of LD_{50} for rats will validate the proposed models and will allow the toxicity of nitroaromatic compounds to mammals to be described. For this purposes we have collected LD_{50} values from various sources [14–19]. The QSARs were developed for mechanism interpretation and prediction.

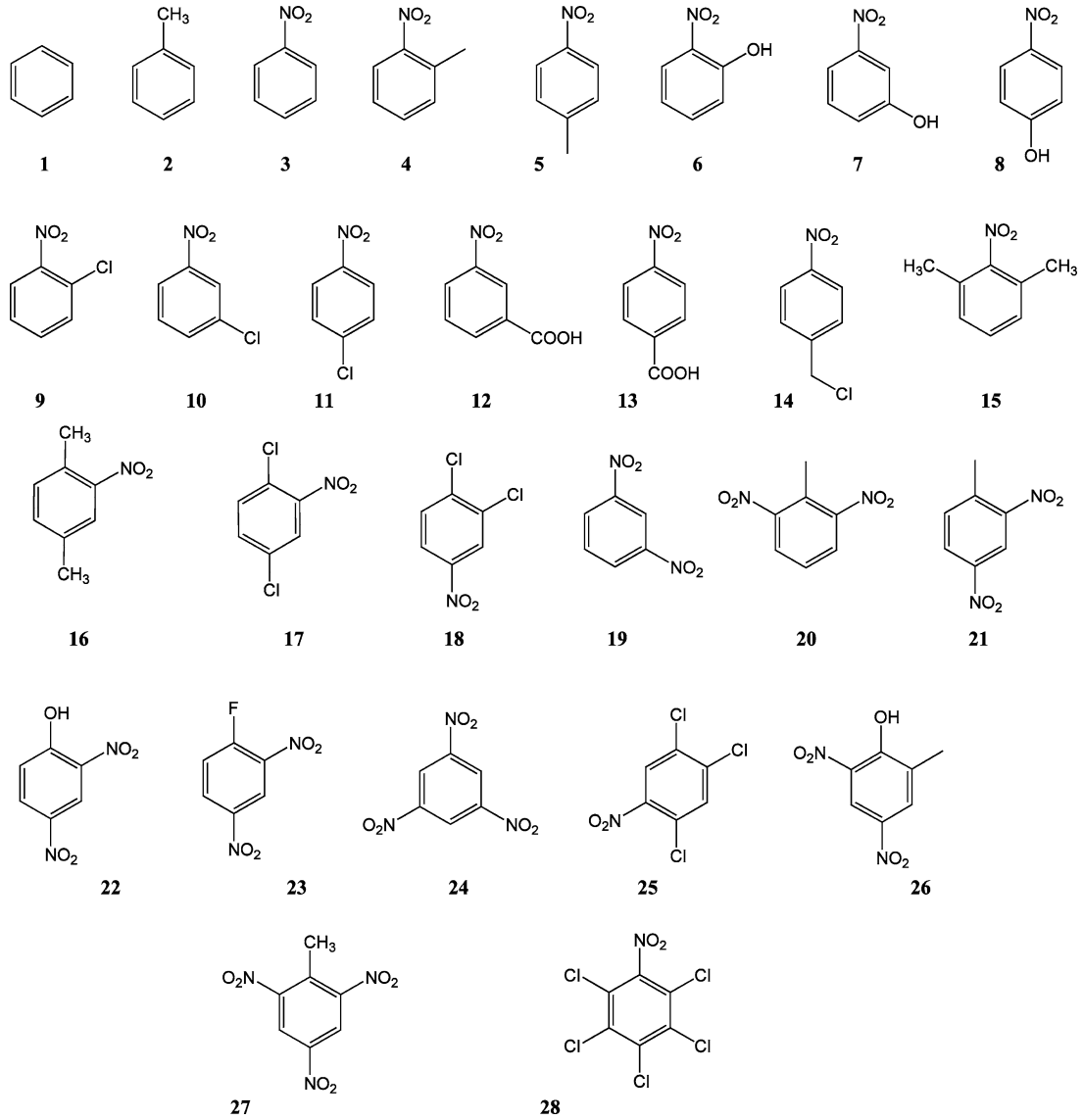
Although semi-empirical molecular orbital methods are currently not state-of-the-art techniques for studying electronic and geometrical properties, the AM1 calculations appear to be still the method of choice when calculating quantum-chemical descriptors for QSARs [8]. However, with the advent of modern computational capabilities and the development of fast algorithms, more reliable *ab initio* molecular orbital and first principles density functional methods could be applied to the studied species and these techniques were used in the present QSAR study.

The presented paper is organized in the following way. The section "Computational Methodology" describes a quantum chemical approximation and the QSAR strategy used. The section "Results and Discussion" is subdivided into two parts. The first part discusses the properties of NACs, which have been calculated at a modern quantum–chemical level and later will be used in QSAR analysis. The second part describes the QSAR analysis itself. The section "Conclusion" includes the most important results obtained during our study.

Computational methodology

Quantum-chemical calculations

Molecular geometries of all model structures (see Scheme 1) were fully optimized using the density functional theory (DFT) with Becke's three-parameter hybrid functional using the LYP correlation functional (B3LYP) [20,21] and the Hartee-Fock level of theory. The molecular symmetry of the model structures was maintained (where applicable) during the optimizations. The standard Pople's 6-311+G(d,p) and 6-31G(d) basis sets were used in the DFT and HF calculations, respectively. To verify that the obtained structures have been properly attributed to the local minima, the matrices of the energy second derivatives at the corresponding level of theory were checked to have zero imaginary values. Atomic charges were computed using the Mulliken population analysis. The values of the vertical (IP_v) and adiabatical (IP_{ad}) ionization potentials were determined by a standard approach through



Scheme 1. Molecular structures of the substituted and unsubstituted aromatic compounds studied in this work.

the calculation of energies of all involved species. IPs were not corrected for zero-point energy, assuming a negligible error. All calculations were carried out using the Gaussian 98 program package [22].

QSAR methodology

The training set for the present investigation was created from a series of 28 compounds (Scheme 1). The activities of the studied compounds are expressed in terms of LD_{50} dose for rats [14–22]; the corresponding values are represented in Table 1. All original LD_{50} toxicity data (mg/kg) have been converted to molar $-\log(LD_{50})$ response variables. The molecular geometries were obtained by the procedure explained in section "Quantum-Chemical Calculations". Constitutional, topological and molecular descriptors were calculated with the *DRAGON* software [23]. A set of 1497 molecular descriptors of different kinds was used to describe the chemical diversity of the compound. The descriptor typology is: (a) constitutional (atom and group fragments); (b) functional groups; (c) atom-centered fragments; (d) empirical; (e) topological; (f) walk counts; (g) various autocorrelations from the molecular graph; (h) Randic molecular profiles from the geometry matrix; (i) geometrical; (j) WHIMs; and (k) GETAWAYs descriptors and various indicator descriptors. The meaning of these molecular descriptors and the calculation procedure are summarized elsewhere [24].

Taking into account the importance to the QSAR/QSPR of some electronical molecular properties that cannot be derived by additive formulas (all descriptors, calculated by *DRAGON* software), we have used quantum-chemical *ab initio* calculations. Several parameters, evaluated by quantum-chemical calculations (as described above), were added as molecular descriptors. These descriptors represent the energies of the

Table 1. Toxicity (LD_{50}, mg/kg and molar $-LogLD_{50}$) of all considered compounds

Number	Compound	Toxicity, (oral LD_{50} in rats)	Toxicity, $Log[LD_{50}]^{-1}$
1	Benzene	5600	−3.748
2	Toluene	5500	−3.74
3	Nitrobenzene	600	−2.778
4	2-nitrotoluene	891	−2.950
5	4-nitrotoluene	2144	−3.331
6	2-nitrophenol	334	−2.524
7	3-nitrophenol	328	−2.516
8	4-nitrophenol	202	−2.305
9	1-chloro-2-nitrobenzene	268	−2.428
10	1-chloro-3-nitrobenzene	390	−2.591
11	4-chloronitrobenzene	420	−2.623
12	3-nitrobenzoic acid	680	−2.833
13	4-nitrobenzoic acid	1960	−3.292
14	4-nitrobenzylchloride	1809	−3.257
15	1,3-dimethyl-2-nitrobenzene	2000	−3.301
16	1,4-dimethyl-2-nitrobenzene	2440	−3.387
17	1,4-dichloronitrobenzene	4000	−3.602
18	3,4-dichloronitrobenzene	643	−2.808
19	1,3-dinitrobenzene	83	−1.919
20	2,6-dinitrotoluene	250	−2.255
21	2,4-dinitrotoluene	270	−2.428
22	2,4-dinitrophenol	71	−1.851
23	2,4-dinitrofluorobenzene	50	−1.699
24	1,3,5-threenitrobenzene	275	−2.439
25	2,4,5-trichloronitrobenzene	1070	−3.029
26	4,6-dinitrocresol	60	−1.778
27	2,4,6-trinitrotoluene	700	−2.845
28	Pentachloronitrobenzene	1100	−3.041

highest occupied molecular orbital (E_{HOMO}), the lowest unoccupied molecular orbital (E_{LUMO}), the HOMO-LUMO gap (HLG), vertical (IP_v) and adiabatical (IP_{ad}) ionization potentials, dipole moment (μ), atomic charges (q_i), electrostatic densities and charges, and total energies (E). The selected values of these parameters are collected in tables throughout the article. The full set of values is available from the authors upon request.

The correlation between biological activity and structural properties was obtained by using the variable selection the Genetic Algorithm (GA) and Multiple Linear Regression Analysis (MLRA) methods. Genetic Algorithms have been applied in recent years as a powerful tool to address many problems in drug design [25, 26]. We applied GA to select from all of the calculated descriptors, only the best combinations of those most relevant for obtaining models with the highest predictive power for toxicity. Finally, the combination GA-MLRA technique was utilized to select the appropriate descriptors and to generate different QSAR models. The GA technique started with a population of 100 random models and 1000 iterations to evolution. For GA analysis and the derivation of the QSAR models, the BuildQSAR program [27] has been used. A final set of QSARs was identified by applying the "leave-one-out" technique with its predictive ability being evaluated and confirmed by cross validation coefficient Q^2 based on a predictive error sum of squares (SPRESS).

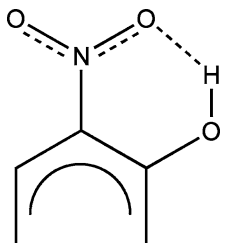
Results and discussion

Structural properties

The equilibrium structures of the 28 NACs were obtained by optimization at the B3LYP/6-311+G(d,p) level. Scheme 1 shows the molecular structures of the compounds calculated in the present study. The main geometrical parameters of all considered compounds are collected in Table 2. We expect that the nitro groups lie in plane (or slightly off plane) of the molecules with no substituents in the *ortho*-position to this group (**3, 5, 7, 8, 10–14, 18–27**) since this orientation facilitates the maximum resonance stabilization.

The geometry optimizations using the DFT method yield planar configurations for all mentioned molecules. The presence of a substituent in the *ortho*-position to the nitro group (**4, 9, 15–17, 21, 23, 25, 27, 28**) introduce significant changes to the geometry of this group since the substituents are close to each other and cause steric crowding. Thus, the presence of a methyl group ($-CH_3$) causes rotation by 26–30° (**4, 21**), and two groups will more than double this effect (56°), for 2.6-dimethyl-nitrobenzene (**15**). The highest effect was found for pentachloronitrobenzene (**28**), where the nitro group is perpendicular to the ring plane. To our knowledge, there are only two sets of experimental data for *ortho*-nitrotoluene (**4**) [28] and *ortho*-nitrochlorobenzene (**9**) [29] concerning the out-of-plane position of the nitro group. These values are 38 and 146 degrees. The corresponding calculated values of 26.5 and 139.6 are very reliable since the uncertainty of these experimental values is estimated to be ±15 degrees [28].

In contrast, in **6, 20** and **22**, the *ortho*-nitro groups are planar. This fact can be explained by the strong intramolecular interactions. In **6** and **22** the $-NO_2$ group is connected by intramolecular hydrogen bonding with the hydrogen of the OH group (see Scheme 2). The nature of this interaction and the properties of similar compounds were the subject of extensive experimental and theoretical studies [30, 31].



Scheme 2. Representation of intramolecular hydrogen bonding in **6** and **22**.

Table 2. Calculated (B3LYP/6-311+G(d)) and experimental (in parenthesis) geometrical parameters of NACs

System	Bond length, A									*R*	Dihedral Angle, °	Ref, expt
	C_1-C_2[a]	C_2-C_3	C_3-C_4	C_4-C_5	C_5-C_6	C_6-C_1	C-N[b]	N-O[b]	C-R		∠ (O-N-C-C)	
1	1.395[b] (1.399)											32
2	1.399 (1.398)	1.394	1.394						1.511 (1.509)	CH_3		28
3	1.392 (1.396)	1.391	1.395				1.479 (1.486)	1.224 (1.223)			0	32
4	1.405 (1.405)	1.402 (1.399)[b]	1.391	1.394	1.388	1.395	1.479 (1.490)	1.226 (1.231)	1.508 (1.508)	CH_3	26.5 (38±15)	28
5	1.393 (1.401)	1.388 (1.383)	1.403 (1.392)	1.399 (1.392)	1.391 (1.382)	1.390 (1.408)	1.475 (1.482)	1.226 (1.242)	1.508	CH_3	0	33
6	1.416 (1.396)	1.404 (1.384)	1.382 (1.354)	1.404 (1.373)	1.381 (1.365)	1.402 (1.393)	1.453 (1.457)	1.234 (1.219)	1.338	OH	0	34
7	1.388 (1.377)	1.394 (1.390)	1.398 (1.375)	1.394 (1.381)	1.390 (1.382)	1.392 (1.366)	1.483 (1.480)	1.225 (1.216)	1.364	OH	0	35
8	1.392 (1.383)	1.388 (1.380)	1.400 (1.393)	1.400 (1.387)	1.385 (1.377)	1.395 (1.388)	1.468 (1.442)	1.226 (1.234)	1.358	OH	0	36
9	1.399	1.395	1.391	1.394	1.388	1.393	1.478	1.223	1.743	Cl	139.6 (146)	29
10	1.391	1.389	1.394	1.394	1.392	1.390	1.483	1.224	1.752	Cl	0	
11	1.392 (1.39)[b]	1.387	1.394				1.477 (1.47)	1.225 (1.20)	1.750	Cl	0	37
12	1.390	1.396	1.401	1.391	1.393	1.390	1.482	1.224	1.504	COOH	0.5	
13	1.391	1.390	1.400	1.399	1.389	1.392	1.483	1.224	1.491	COOH	−0.1	
14	1.389	1.392	1.395	1.403	1.387	1.393	1.478	1.225	1.512	CH_2Cl	0	
15	1.401	1.307	1.391				1.481	1.224	1.509	CH_3	56.6	
16	1.403	1.396	1.391	1.400	1.389	1.402	1.478	1.226	1.508	CH_3	25.3	
17	1.398	1.392	1.387	1.393	1.390	1.395	1.480	1.222	1.745	Cl	137.1	
18	1.402	1.391	1.389	1.390	1.387	1.396	1.479	1.224	1.741	Cl	0	
19	1.388 (1.384)	1.388 (1.384)	1.392 (1.384)	1.392 (1.392)	1.393 (1.392)	1.392 (1.384)	1.484 (1.486)	1.223 (1.220)			0	38
20	1.401	1.397	1.391				1.504	1.223	1.511	CH_3	0	
21	1.406	1.390	1.385	1.390	1.388	1.403	1.481	1.224	1.506	CH_3	−29.6 (*o*-), 0.5 (*p*-)	
22	1.407 (1.403)	1.386 (1.391)	1.387 (1.385)	1.393 (1.389)	1.384 (1.382)	1.403 (1.400)	1.467 (1.466)	1.227 (1.223)	1.330	OH	0 (*o*-), 0 (*p*-)	39
23	1.399 (1.397)	1.389 (1.399)	1.386 (1.366)	1.393 (1.373)	1.386 (1.356)	1.390 (1.372)	1.479 (1.485)	1.222 (1.224)	1.330 (1.336)	F	150 (*o*-), 0.3 (*p*-)	40
24	1.389						1.487	1.220			0	
25	1.401	1.392	1.392	1.399	1.390	1.388	1.478	1.222	1.738	Cl	142.5	
26	1.410	1.393	1.385	1.386	1.388	1.403	1.478	1.224	1.509, 1.342	CH_3, OH	0	
27	1.407	1.387	1.385	1.385	1.388	1.405	1.485	1.221	1.506	CH_3	140.2 (*o*-), 128.1 (*o′*-), −1.0 (*p*-)	
28	1.390	1.401	1.404				1.488	1.217	1.732	Cl	90	

[a]Numbering here and after, clockwise from the upper carbon atom of the cycle, as depicted on Scheme 1.
[b]Average bond length values where applicable.

The experimental values of bond lengths have also been included in Table 2. The geometrical data of compounds **1–5** and **22** were obtained from gas phase electron diffraction [28, 32, 33]. Excellent correspondence was found between the calculated and experimental geometrical parameters, within less than 0.1% (mainly 0.001–0.005 Å) of difference. For other systems (**6–8**, **11**, **19** and **23**) only X-ray crystallographic data are available [29, 34–40]. In this case, the deviations of the calculated data from experiments are significantly higher (0.01–0.02 Å or 2%). The value of this difference depends on the crystal packing and the type of experiment. The experimental N═O, C─O and O─H bond lengths are typically longer than the calculated values; these difference are due to the interatomic interactions in the crystal lattice of these compounds. The substituent effects on the geometry of the benzene ring were subject of several recent extensive theoretical investigations [e.g., 40–43]; thus, we will not discuss this subject in details.

In summary, the structural changes of the carbon skeleton effect all bond distances and angles. They are the most significant at the location of substitution and depend on the electronegativity as well as on the donor/acceptor character of the substituent. The DFT approach can satisfactorily reproduce the molecular geometries of the considered species. All geometrical and topological descriptors evaluated from this geometry can be considered as highly reliable.

Electronic properties

The values of the calculated vertical (IP_v) and adiabatical (IP_{ad}) ionization potentials are listed in the Table 3 along with the available experimental values. The IPs obtained from the DFT calculations are in excellent agreement with the experimental IP values. The mean absolute deviation between IP values is 0.07 eV for the 16 compounds for which experimental data are available [44]. Few compounds highly contribute to the absolute deviation of the whole series: the computed values for 1-chloro-3-nitrobenzene (**10**) and 1-chloro-4-nitrobenzene (**11**) deviate from experiment by 0.16 and 0.18 eV, respectively; that is, the experimental IP is much higher than the calculated IP. All these experimental values were determined using an electron impact technique that has been shown [45] to predict ionization potentials that are typically too high. Removing the ionization potentials for these compounds and re-evaluating the deviation results in an improved mean absolute deviation value of 0.05 eV. The computed values for both trinitrobenzene (**24**) and trinitrotoluene (**27**) are higher than the experimental values by 0.21 and 0.09 eV, respectively. For these two cases, the experimental values were obtained from photoionization mass spectrometry and lie in between the calculated vertical and adiabatical ionization potentials. In general, B3LYP performs very good to reproduce relative IPs between different isomers (except isomeric chloronitrobenzenes where the difference in IPs is within the error of the experiment) and can be treated as trustworthy molecular descriptors.

The HF results significantly underestimate the energy needed to ionize the molecules. The difference between the calculated and experimental IPs amounts to 0.5–1.5 eV. In addition, the ordering of IPs does not follow the trend of the experimental data. The values of IP calculated by HF can be found in supplementary materials. The deficiency of the UHF method originates from the fact that it does not result in wave functions that are eigenfunctions of $\hat{S}^2$; that is, it does not produce a pure spin state [46].

To have a qualitative understanding of the trends in IP on the nature (either electron-withdrawing or electron-donating) and number of substituents, we analyzed some tendencies in the frontier orbital energies of the considered compounds. From Table 3 it can be seen that increasing the number of nitro groups in the molecule from 0 to 3 decreases the HOMO energy. The influence of the first group is strongly expressed; it lowers E_{HOMO} by 0.898 eV. The further additions of nitro groups just slightly affect the E_{HOMO}. The value of the addition of the first nitro group lowers the LUMO energy by about 2.7 eV and each next group by 0.68 eV. It is clear that while the addition of the π acceptor group ($-NO_2$) stabilizes the HOMO energy levels, the donor groups produce the opposite effect. Thus, the addition of the OH group to dinitrobenzene (**19**) increases E_{HOMO} by 0.735 eV, similarly CH_3 by 0.571 eV in the case of trinitrobenzene. An acceptor or donor group can influence the values of electron density in a molecule. The increase in the orbital energies by substitution of a donor group is due to an increase in the electron density leading to the contribution of a positive value to the electron–electron repulsion term in the total energy expression. The opposite is true for substitution with an acceptor.

However, the effect of an electron-withdrawing group on IP is essentially the result of the σ-inductive (or deshielding) effect. The addition of a second electron-withdrawing group must lower the HOMO energy by the same mechanism. Comparisons of the Mulliken charges on the electron-withdrawing group monosubstituted benzene relative to the parent molecule show very little change in charge distribution at the unsubstituted sites. This implies that, regardless of where the substitution occurs, the addition of a second electron-withdrawing group to molecules of this type will have essentially the same effect on the HOMO energy as the first one. This can be seen by examining the effect on HOMO in both nonconjugated and conjugated electron-withdrawing substituted species. Thus, 1-chloro-4-nitrobenzene (**11**, Cl is conjugated) and 4-nitrobenzylchloride (**14**, CH_2Cl is nonconjugated) are characterized by E_{HOMO} -7.864 and -7.978 eV, respectively and an almost identical vertical IP, respectively. Similarly, the effect of the electron-donating group is realized by the same mechanism. The corresponding values of E_{HOMO} and IPs for isomeric nitrophenols (**6–8**) as well as nitrotoluenes (**4, 5**) are very close to each other.

A discussion of the role of two substituents should, in principle, be extendable to the case of polysubstituted NACs. However, we can see significant nonadditivity in their

Table 3. Calculated frontier orbital energies, HOMO-LUMO gap (HLG) in eV and ionization potentials in eV

	B3LYP[a]			HF[b]	B3LYP			
System	HOMO	LUMO	HLG	HLG	$IP_{adiabatical}$	$IP_{vertical}$	Expt. IP, eV[c]	Technique[d]
1	−7.050	−0.435	6.615	12.955	9.11	9.26	(9.24384 ± 0.00006)	TE
2	−6.727	−0.395	6.332	12.637	8.64	8.80	(8.8276 ± 0.0006)	TE
3	−7.938	−2.909	5.029	11.350	9.87	10.03	9.85 ± 0.03	
4	−7.595	−2.721	4.871	11.219	9.35	9.57	9.51 ± 0.03	
5	−7.698	−2.789	4.909	11.282	9.48	9.65	9.46 ± 0.05	
6	−7.206	−3.170	4.035	10.373	9.02	9.23	9.1 (9.29)	PE
7	−7.184	−2.893	4.289	10.691	9.02	9.24	9.0 (9.33)	PE
8	−7.331	−2.735	4.596	11.135	9.08	9.34	9.1 (9.38)	PE
9	−7.687	−2.876	4.811	11.412	9.47	9.66		
10	−7.755	−3.148	4.607	11.061	9.56	9.74	(9.9 ± 0.1)	EI
11	−7.864	−3.108	4.757	11.236	9.55	9.82	(10.0 ± 0.1)	EI
12	−8.329	−3.420	4.909	11.491	9.88	10.25	(10.3 ± 0.2)	EI
13	−8.229	−3.382	4.846	10.901	10.01	10.22	(10.2 ± 0.2)	EI
14	−7.978	−2.988	4.991	11.385	9.68	9.83		
15	−7.312	−2.555	4.759	11.421	9.05	9.23		
16	−7.279	−2.656	4.626	10.920	8.95	9.17		
17	−7.608	−3.094	4.517	11.138	9.28	9.48		
18	−7.848	−3.282	4.563	11.007	9.54	9.71		
19	−8.776	−3.589	5.186	11.475	10.34	10.47	10.43 ± 0.02	PI
20	−8.378	−3.453	4.925	11.214	10.06	10.27		
21	−8.438	−3.396	5.042	11.363	9.96	10.33		
22	−8.036	−3.287	4.748	11.325	9.64	9.93	9.57	PE
23	−8.719	−3.619	5.102	11.668	10.29	10.73		
24	−9.415	−4.125	5.290	11.900	11.01	11.17	10.96 ± 0.02	PI
25	−7.747	−3.233	4.514	11.061	9.31	9.52		
26	−7.704	−3.393	4.308	10.623	9.38	9.60		
27	−8.844	−2.909	5.935	11.614	10.42	10.68	10.59 ± 0.04	PI
28	−7.810	−3.088	4.724	11.753	9.32	9.49		

[a]B3LYP / 6-311+G(d,p).
[b]RHF / 6-31G(d).
[c]Experimerntal from ref [44] adiabatical and/or vertical values in parenthesis.
[d]TE – Threshold electron detection, PE – Photoelectron spectroscopy, EI – Electron impact techniques ("Electron impact"), PI – Photoionization mass spectrometry.

combined effect which has been shown by Di Labio and coworkers [42] for polymethylbenzenes.

The molecular dipole moment is perhaps the simplest experimental measure of charge density distribution in a molecule. The calculated values of dipole moments are displayed in Table 4. For the molecules considered in this work, several experimental values of dipole moment were found in literature, including values for all isomeric nitrotoluenes and nitrochlorobenzenes [47], nitrobenzene [48], and toluene [32]. For **1** and **24**, the experimental dipole moments are considered to be 0. For these molecules the nitro groups are located in symmetrically opposed positions, and the components of μ compensate, yielding a total $\mu = 0$.

Overall, the quality of the DFT estimation is a little better than the Hartree-Fock estimation. Both methods usually overestimate dipole moment approximately 0.7 D and 0.9 D, correspondingly. For the set of 10 compounds, it gives an average error of 18% for B3LYP/6-311+G(d) and of 24% for HF/6-31G(d). The presence of a π-conjugated $-NO_2$ group dramatically decreases the accuracy of the dipole moment calculation. For instance, for toluene (**2**) $\mu = 0.37$, the DFT calculation reproduces the experimental value within 2% and HF with a 22% error. However, with replacement of $-NO_2$ by a $-CH_3$ group, the error increases to 22% and 25%, respectively. Both methods satisfactorily reproduce the tendency of the decrease of μ from *o*-nitro- to *p*-nitro-chlorobenzene. However, they failed to describe the same tendency along isomeric nitrotoluenes. The value of μ for *m*-nitrotoluene is dropped out. The predicted values of μ for *metha*-isomers in both rows are much worse, than for the corresponding *ortho*- and *para*-isomers. This behavior is probably related to the possible nonlinear optical properties for these molecules.

In conclusion, the values of the calculated dipole moments cannot be considered as reliable molecular descriptors due to the large uncertainty of their estimations.

Table 4. Dipole moments calculated at B3LYP 6-311+G(d) and RHF/6-31G(d) and experimental values

System	μ, D (B3LYP)	μ, D (RHF)	Expt, μ, D	Ref
1	0.00	0.00	0[a]	
2	0.38	0.29	0.37	32
3	4.95	5.27	4.22	48
4	4.62	4.93		
5	5.65	5.85		
6	3.98	4.29	3.13	47
7	6.21	6.41	3.90	47
8	5.58	5.58	4.83	47
9	5.20	5.76	4.64	47
10	4.23	4.48	3.40	47
11	3.39	3.31	2.83	47
12	2.90	3.11		
13	3.84	4.05		
14	4.40	4.54		
15	3.90	4.36		
16	4.91	5.17		
17	4.22	4.61		
18	3.10	3.02		
19	4.56	4.95		
20	2.99	3.36		
21	5.26	5.56		
22	3.67	3.92		
23	4.01	4.37		
24	0.00	0.00	0[a]	
25	3.12	3.15		
26	5.70	5.75		
27	1.72	1.63		
28	2.78	3.04		

[a]By the symmetry reason.

Table 5. Correlation matrix for physicochemical and 2D- 3D-descriptors used in present work

	Ms	**X5Av**	**PCR**	**PCD**	**C-026**	**BELe1**	$\mathbf{E_{LUMO}}$
Ms	1	.716	.484	.598	.358	.059	.813
X5Av		1	.497	.505	.037	.242	.459
PCR			1	.632	.015	.464	.367
PCD				1	.263	.288	.398
C-026					1	.120	.288
BELe1						1	.048
$\mathbf{E_{LUMO}}$							1

Ms – mean electrotopological state; **X5Av** – average valence connectivity index; **PCR** – account of ratio of multiple path counts to path counts; **PCD** – difference of multiple path counts to path counts; **BELe1** – positive and negative eigenvalues of the adjacency matrix; $\mathbf{E_{LUMO}}$ – Energy of Lowest Unoccupied Molecular Orbital.

QSAR

The results of the best structure-toxicity models using one to three descriptors are given by Equations 1–7. While constructing the models, great care was taken in order to avoid inclusion of highly collinear descriptors. The correlation matrix for the 2D–3D descriptors together with physicochemical descriptors used in this study is given in Table 5. Table 5 includes only those variables that have comprised the most populated models selected by the variable selection GA method. We excluded from consideration all descriptors which have a cross-correlation of more than 0.6 between each other. The exclusion has been made only for the **Ms**-**X5Av**, **Ms**-$\mathbf{E_{LUMO}}$ and **PCR**-**PCD** cross- correlations ($r = 0.716$, 0.813 and 0.632, respectively), because these descriptors probably encode partly similar information; however, they are presented in different equations.

In summary, we have described as one-descriptor models, as two- and three-descriptor models, where in multi-descriptor models are present the combinations of descriptors from other models. In the models 1–7 present three types of descriptors: 2D (encodes two-dimensional information), 3D (encodes three-dimensional information) and quantum-chemical ($\mathbf{E_{LUMO}}$ energy) descriptors. The descriptors **X5Av** and **Ms** to a greater extent describe the content of the heteroatoms (N and O) in a molecule, while **PCR**, **BELe1** and **PCD** describe information related to bonds and distances (bond orders, saturation and ratio of multiple bonds to single bonds). In addition, the quantum-chemical descriptor $\mathbf{E_{LUMO}}$ allows the reactivity of explored compounds to be directly described.

The first one-parametric model is represented by model (1) and includes the parameter **X5Av**:

$$\mathbf{Log[LD_{50}]^{-1}} = -77.5336(\pm 24.6603)\mathbf{X5Av} - 4.4177(\pm 0.6454) \quad (1)$$
$$(n = 28;\ r = 0.785;\ s = 0.394;\ F = 41.624;\ Q^2 = 0.552;\ \text{SPRESS} = 0.425)$$

where n is the number of compounds in training set; r is the correlation coefficient of regression; s is the standard error; F is the **F**-ratio between the variances of observed and calculated activities; Q^2 is the leave-one-out cross-validation coefficient; SPRESS is the predictive error sum of squares.

The topological descriptor **X5Av** is an average valence connectivity index which encodes the presence of double and triple bonds as well as heteroatoms in the molecule [24]. Hence, the negative coefficient of **X5Av** indicates that a decrease in the descriptor value results in an increase in toxicity. Based on the descriptor nature and qualitative changes, one can conclude that the presence of nitrogen and oxygen atoms (nitro group) makes a larger impact than the halogen atoms.

Model (2) shows the relationship of the **Ms** parameter which has a correlation coefficient $R = 0.827$:

$$\mathbf{Log[LD_{50}]^{-1}} = 1.0995(\pm 0.3009)\,\mathbf{Ms} - 1.1062(\pm 0.9821) \quad (2)$$

($n = 28$; $r = 0.827$; $s = 0.357$; $F = 56.202$;
$Q^2 = 0.640$; SPRESS = 0.381)

The descriptor **Ms** describes a mean electrotopological state. It is a one of the most interesting indices of this type, which is based on the electronic state (E-state) of each atom type and its topological nature in a molecular graph, introduced by Kier et al. [49]. E-state indexes have been shown to be useful for the prediction of various properties [50–53]. Especially useful are those indexes which describe the structural-biological activity (property) relationships where an electronic property of a molecule plays an important role. Electrotopological state indexes encode the number of valence electrons and the degree of branching at each atom. Lower E-state indexes are assigned to atoms that have fewer valence electrons and that are away from the periphery of the molecule. The descriptor **Ms** expresses a mean value of the E-states of all atoms in the molecule. Hence, based on model (2) one can conclude that an increase in the number of atoms with a high number of valence electrons will increase the toxicity of the compound. The presence of the nitrogroups increases the electrotopological state and hence the toxicity of nitrobenzenes; this statement confirms the conclusions from model (1). Clearly, such a descriptor does provide fundamental molecular information.

Model (3) consists of two topological descriptors:

$$\mathbf{Log[LD_{50}]^{-1}} = 1.4564(\pm 0.3728)\,\mathbf{Ms} - 0.5132(\pm 0.3731)\,\mathbf{PCR} - 0.1120(\pm 1.3375) \quad (3)$$
($n = 28$; $r = 0.872$; $s = 0.317$; $F = 39.663$;
$Q^2 = 0.697$; SPRESS = 0.357)

The descriptor **PCR** is the ratio of multiple path counts to path counts [24]. As it can be seen from the behavior of the **PCR** descriptor in equation (3) the decrease of the **PCR** value leads to an increase in toxicity. Most likely, descriptors **Ms** and **PCR** encode partially similar structural information in relation to valence; furthermore, these are complementary indexes. The cross-correlation between these descriptors provides values smaller than 0.5. There is good reason to believe that in this case the value of the **Ms** descriptor mainly depends on the presence of heteroatoms in the molecule, and this impact is essential.

The three-parametric model (4, Figure 1) includes the **PCD** and **C-026** descriptors in addition to descriptor **X5Av**:

$$\mathbf{Log[LD_{50}]^{-1}} = -103.8848(\pm 23.5754)\,\mathbf{X5Av} - 0.1573(\pm 0.0802)\,\mathbf{PCD} + 0.2906(\pm 0.0966)\,\mathbf{C\text{-}026} + 5.2116(\pm 0.8447) \quad (4)$$
($n = 028$; $r = 0.924$; $s = 0.253$; $F = 46.647$;
$Q^2 = 0.792$; SPRESS = 0.302)

PCD represents the difference of multiple path counts to path counts. This descriptor encodes almost the same information as the PCR descriptor; therefore, the correlation coefficient between them is 0.632. The **C-026** descriptor represents the number of R-CX-R like fragments as defined by Ghost-Grippen [52], where X is a heteroatom that attaches to carbon (particularly, halogen atoms). While the **X5Av** descriptor indicates that a low content of multiple bonds increase toxicity (this is also confirmed by the PCD descriptor); the C-026 descriptor indicates that an increasing number of halogen atoms to some extent increases toxicity. The **PCD** and **X5Av** descriptors encode information concerning the bond types. Jointly with the fragment type descriptor **C-026**, they sufficiently describe features of the molecule in relation to toxicity. Model (4) also shows that using a combined description of the molecular structure by topological descriptors (**PCD** and **X5Av**) describing the presence of heteroatoms and bond types as well as by using fragment descriptors, it is possible to predict the toxicity of nitrobenzenes with high probability.

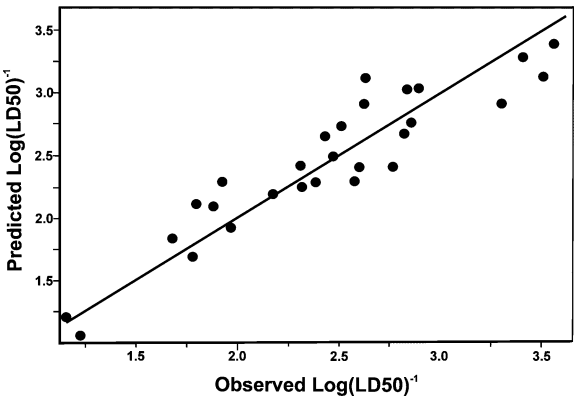


Figure 1. A plot of observed and GA-MLR-predicted $\mathbf{Log[LD_{50}]^{-1}}$ values for a set of 28 nitrobenzenes. Model (4), linear correlation.

As already mentioned in the Introduction, for a better understanding of the experimental results, the number of quantum-chemical parameters of the nitro-compounds was calculated. One of the calculated descriptors is the energy of the LUMO. The $\mathbf{E_{LUMO}}$ has been selected by a GA variable selection algorithm as a descriptor, which has better correlation in comparison with other quantum-chemical descriptors. The $\mathbf{E_{LUMO}}$ energies, calculated by DFT, show better agreement with the experimental toxicity data than the UHF level values.

In the following model (5, Figure 2) a mixed approach using quantum-chemical parameters and topological descriptors have been used to predict toxicity properties of NACs. In our case the energy of the lowest unoccupied molecular orbital $\mathbf{E_{LUMO}}$ has shown good correlation with the toxicity of nitrobenzenes. Thus, decreasing the energy of the LUMO orbital increases the toxicity.

$$\mathbf{Log[LD_{50}]^{-1}} = -77.5474(\pm 24.8304)\,\mathbf{X5Av} - 9.9137(\pm 4.0661)\,\mathbf{BELe1} - 5.5814(\pm 4.7390)\,\mathbf{E_{LUMO}} + 22.1802(\pm 7.9850) \quad (5)$$

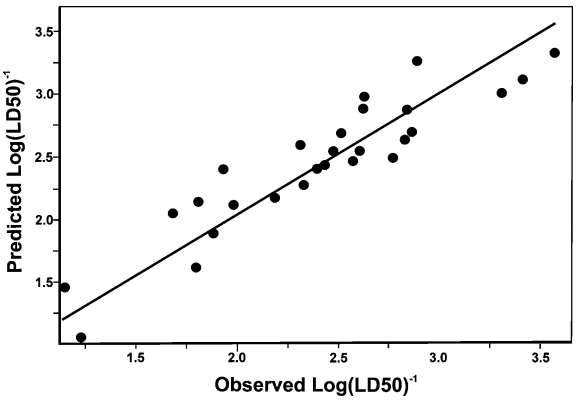


Figure 2. A plot of observed and GA-MLR-predicted $\mathbf{Log[LD_{50}]^{-1}}$ values for a set of 27 nitrobenzenes. Model (5), linear correlation.

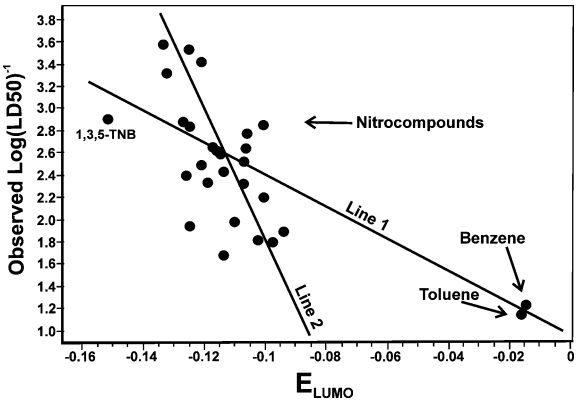


Figure 3. Plot for the relationship between the logarithm of toxicity and the calculated electrophilic reactivity $\mathbf{E_{LUMO}}$.

$$(n = 027;\ r = 0.919;\ s = 0.251;\ F = 41.455;\ Q^2 = 0.768;\ \text{SPRESS} = 0.306)$$

Compound (**26**), 4,6-dinitrocresol, has been selected as an outlier.

BELe1 is a molecular descriptor, the lowest eigenvalue of the Burden matrix, weighted by atomic Sanderson electronegativities, obtained from the positive and negative eigenvalues of the adjacency matrix, and weighting the diagonal elements with atom weights; it is a **BCUT** descriptor [54, 55]. More specifically, **BCUT** descriptors describe the surface distributions of positive charges, negative charges, H-bond donors, H-bond acceptors, regions of high polarizability ("greasiness"), and regions of low polarizability, which indirectly affect lipophilicity. One can see from model (5) that the decreases of the values of all presented descriptors increase the toxicity. The **BELe1** descriptor encodes information concerning distances, electronegativities, and atom types. Jointly with **X5Av** and $\mathbf{E_{LUMO}}$, the **BELe1** descriptors sufficiently describe features of the molecule in relation to toxicity. This is in good agreement with one of the toxicity mechanisms which is based on the reactivity properties of NACs (nitrogroup reduction). This model describes the basic properties of nitrobenzenes quite well and determines the toxicity, such as reactivity properties ($\mathbf{E_{LUMO}}$) and lipophilicity properties (**BELe1**, **X5Av**).

The next model is represented by a simpler equation which demonstrates the linear relationship between E_{LUMO} and toxicity:

$$\mathbf{Log[LD_{50}]^{-1}} = -15.3011(\pm 6.0168)\,\mathbf{E_{LUMO}} + 0.7863(\pm 0.6750) \quad (6)$$

$$(n = 28;\ r = 0.715;\ s = 0.444;\ F = 27.232;\ Q^2 = 0.453;\ \text{SPRESS} = 0.470)$$

This model has shown satisfactory correlation between the $\mathbf{E_{LUMO}}$ energy and toxicity ($R = 0.715$). However, it seems that this model incorrectly describes the function for this dependence. The inclusion of non-nitrocompounds (benzene and toluene) in the dataset may dramatically deform the expected dependence but also results in an increase in the correlation coefficient (Figure 3, line 1). As is seen, the position of benzene and toluene strongly differ from other compounds. According to it, one may assume that in nitrobenzenes toxicity exhibiting the main role is played by its reactivity. It is obvious that benzene and toluene are less reactive than nitrobenzenes. Model (6) shows that the more reactive is the compound, the more toxic it is. In an organism, such reactivity produces oxidative stress in cells, resulting in cell damage. Thus, according to other research [56], the process of hydroxylation by cytochrome P450 of nitrobenzenes directly depends on LUMO energy, and the reaction proceeds through the atom where the greatest density of the LUMO orbital of the substrate is concentrated. A low correlation coefficient of LUMO energy with toxicity [56] shows that there are at least several metabolic processes in an organism, and according to them the reactions could proceed by various reactive centers. In conclusion, some metabolites are formed, because of reactions both at the nitrogroup and at other positions of the benzene ring.

Taking into account the results for the previous model (6), model (7) considers E_{LUMO} correlation only for nitrobenzenes (benzene and toluene are excluded from the explored set; 1,3,5-TNB is a considerable outlier):

$$\mathbf{Log[LD_{50}]^{-1}} = -29.4911(\pm 16.9213)\,\mathbf{E_{LUMO}} - 0.8387(\pm 1.9411) \quad (7)$$

$$(n = 25,\ r = 0.600,\ s = 0.439,\ F = 12.960,\ Q^2 = 0.26,\ \text{SPRESS} = 0.476,\ p > 0.0015)$$

It is obvious that model (7) possesses a poor correlation coefficient ($R = 0.60$), but in this model is much more correct in describing of the $\mathbf{E_{LUMO}}$ energy–toxicity dependence (Figure 3), unlike in the case of model (6). As can be seen from Figure 3, line 1 (model 6) reflects a slightly wrong tendency

because of the presence of non-nitroaromatic compounds, benzene and toluene, while line 2 (model 7) shows a correct tendency (some divergence of points can be explained by deviation in experimental measurements). A poor correlation between $\mathbf{E_{LUMO}}$ and toxicity may be explained by the presence of other competitive processes during the toxic action of nitrobenzenes (e.g., multiple reactions and differences in absorption rates). Also one has to take into account that the experimental values have been collected from various sources, which sharply increases statistical error.

Many active compounds (such as drugs) showed a good correlation between lipophilicity and activity, it is implies that these compounds can be good absorbed through intestine (intestinal absorption). The fundamental importance of LogP (as measure of lipophilicity/hydrophobicity in the distribution of compounds in various biological systems (for example, human organism). Since we have not found any correlation of toxicity with the LogP descriptor (lipophilicity) of nitrobenzenes. Concluding, we can assume that at in vivo toxicity of nitrobenzenes the basic role is played by the reactivity of the compounds, whereas a cell permeability properties is much less important. Probably the rate of cell permeation for nitrobenzenes is quite low in comparison with the reaction rate during interaction with organism. Although, in our study, the presence of descriptors which encoding the electrtopological state of molecules (**Ms**, **BELe1**) indirectly indicate on possibility of the contribution to molecules ability to cell permeation.

An analysis of the descriptors in Equation (1–7) shows good agreement with the results obtained with the known mechanisms of nitrobenzene toxicity [57]. The mechanism of NAC action usually involves the reduction under hypoxic conditions of the nitrogroup to the corresponding amine via nitroradical, nitroso- and N-hydroxyl intermediates.

Thus, the given set of descriptors (2D, 3D and quantum-chemical descriptors) allowed us to determine not only structural (geometric and topologic) properties responsible for toxicity but also to determine the approximate mechanism of NAC action in the living organism. Based on obtained correlations it can be seen that toxicity is mainly related to intracellular oxidation and reduction. The basic relationship has been found for descriptor **Ms** which depends on the electrotopological state of the molecule and descriptor E_{LUMO} which defines the reactive property of nitrobenzenes (reduction of the nitrogroup). Thus, the toxic potency of these compounds could be explained by their molecular structure (**Ms, X5Av, PCR, PCD, BELe1** and **C-029**) and electronic property ($\mathbf{E_{LUMO}}$) involving in *in vivo* reactions with biomacromolecules.

According to the values of the LUMO energies one can assume that, with an increase in the number of nitrogroups, the toxicity also increases, and the orbital energy decreases. Hence, the $\mathbf{E}_{LUMO}$ values of nitrobenzenes are higher, while the corresponding values of dinitrobenzenes and trinitrobenzenes are lower. Dinitrobenzenes and trinitrobenzenes can be easily reduced; therefore, these compounds are more toxic compared with nitrobenzenes. Moderate correlation of the electrophilicity parameter ($\mathbf{E_{LUMO}}$) with experimental LD_{50} toxicity values confirms that the nitrobenzenes are involved in various reaction processes during metabolic modifications in living organism.

Conclusions

In the present work we made an attempt to describe the toxicity of NACs by simple one-descriptor models for toxicity description of nitrobenzenes and by multi-descriptor models for accurate toxicity description. The toxicity parameter LD_{50} for rats for the first time has been utilized for QSAR analysis of nitrobenzenes and is an important improvement. We believe that the obtained models reflect substantial factors concerning the action of toxicity of nitrobenzenes in mammals. An important distinctive feature that makes all of the obtained regressions unique is that it correlates LD_{50} of rats, rather than IC_{50} or Ki's which describe specific receptor sites or the inhibition of specific proteins. As such, most LD_{50} models have a much lower correlation coefficient.

The obtained results and their analysis discussion allow us to conclude that the developed QSARs with 2D and 3D based topological descriptors (distance-based parameters) can be used for estimating toxicity of nitrobenzenes and other nitroaromatic compounds (Models 1–4). Moreover, the quantum-chemical descriptor $\mathbf{E_{LUMO}}$ satisfactorily describes toxicity of nitrobenzenes as reactive chemicals which explains its nonselective toxicity. The major difference between QSAR models with 2D–3D based topological descriptors and quantum-chemical descriptors is a better statistical fit for the 2D–3D based models. An important conclusion is that the topological models can describe well the toxicity of nitroaromatic compounds as well as non-nitroaromatic compounds.

Molecular geometry and electronic properties (ionization potentials, dipole moment, orbital energies and charges) obtained by B3LYP/6-311+G(d,p) calculations were evaluated by comparison with available experimental data (where applicable). All of the above parameters and derived topological descriptors from the calculated geometry are found to be reliable, except for the dipole moments due to the large uncertainty of their estimation. At the same time, the calculated descriptors, in particular E_{LUMO}, have definite chemical meaning and direct relation with the mechanism of molecular action in living organism, despite the obtained poor correlation coefficient. It is clear that the mechanism of nitrogroup reduction plays a major role in the toxicity properties of nitrobenzenes and the rate of nitrogroup reduction, and consequently toxicity depends on electronic state of the compound.

Finally, the mechanism-based model (7) after validation can be used on expanded datasets as an efficient predictive tool for the evaluation of the toxicity of nitrobenzenes; – however, there very poor data on oral LD_{50} toxicity of nitrobenzenes for mammals. Regardless of the relatively small

set of compounds, the variation of substituents presented (in addition to the nitro groups) is quite wide, and this fact can prove the applicability of these models for further use for predicting *in vivo* toxicity.

Acknowledgment

This work was supported in part by National Science Foundation through EPSCoR Grant No. 94-4-756-01, National Institute of Health RCMI Grant No. G12RR13459. We thank the Mississippi Center for Supercomputer Research for a generous allotment of computer time.

References

1. Hartter, D.R., *The use and importance of nitroaromatic chemicals in the chemical industry*, In Rickert, D.E. (Ed.), Toxicity of nitroaromatic compounds. Chemical Industry Institute of Toxicology Series, Chemisphere, Washington, D.C., 1985, pp. 1–14.
2. Nitrobenzene. Initial report of the TSCA Interagency Testing Committee to the administrator. EPA 560-10-78/001. U.S. Environmental Protection Agency, Washington, D.C., 1978.
3. Kriek, E., *Aromatic amines and related compounds as carcinogenic hazards to man*, In Emmelot, P. and Kriek, E. (Eds.), Environmental carcinogenesis, Elsevier, Amsterdam, 1979, pp. 143–164.
4. Won, W.D., di Salvo, L.H. and Ng. J., *Toxicity and mutagenicity of 2,4,6-trinitrotoluene and its microbial metabolites*, Appl. Environ. Microbiol., 31 (1976) 576–580.
5. Slater, E.C., *Mechanism of uncoupling of oxidative phosphorylation by nitrophenols*, Comp Biochem Physiol., 4 (1962) 281–301.
6. Donlon, B.A., Razo-Flores, E., Field, J.A. and Lettinga, G., *Toxicity of N-substituted aromatics to acetoclastic methanogenic activity in granular sludge*, Appl. Environ. Microbiol., 61 (1995) 3889–3893.
7. Soffers, A.E.M.F., Boersma, M.G., Vaes, W.H.J., Vervoort, J., Tyrakowska, B., Hermens, J.L.M. and Rietjens, I.M.C.M., *Computer-modeling-based QSARs for analyzing experimental data on biotransformation and toxicity*, Toxicology in Vitro, 15 (2001) 539–551.
8. Katritzky, A.R., Oliferenko, P., Oliferenko, A., Lomaka, A. and Karelson, M., *Nitrobenzene toxicity: QSAR correlations and mechanistic interpretations*, J. Phys. Org. Chem., 16 (2003) 811–817.
9. Agrawal, W.K. and Khadikar, P.V., *QSAR prediction of toxicity of nitrobenzenes*, Bioorg. Med. Chem., 9 (2001) 3035–3040.
10. Cronin, M.T.D. and Schultz, T.W., *Development of Quantitative Structure-Activity Relationships for the Toxicity of Aromatic Compounds to Tetrahymena pyriformis: Comparative Assessment of the Methodologies*, Chem. Res. Toxicol., 14 (2001) 1284–1295.
11. Mekenyan, O., Roberts, D.W. and Karcher, W., *MO-Parameters as Predictors of Skin Sensitization Potential of Halo- and Pseudohalobenzenes Acting as SNAr Electrophiles*, Chem. Res. Toxicol., 10 (1997) 994–1000.
12. Cronin, M.T.D., Gregory, B.W. and Schultz, T.W., *Quantitative structure-activity analyses of nitrobenzene toxicity to Tetrahumena pyriformis*, Chem. Res. Toxicol., 11 (1998) 902–908.
13. Schmitt, H., Altenburger, R., Jastorff, B. and Schuurmann, G., *Quantitative stucture-activity analysis of the algae toxicity of nitroaromatic compounds*, Chem. Res. Toxicol., 13 (2000) 441–450.
14. Toxicological Profile For Nitrophenols: 2-Nitrophenol, 4-Nitrophenol, Agency for Toxic Substances and Disease Registry, U.S. Public Health Service, July 1992.
15. Toxicological Profile for Dinitrocresols. U.S. Department of Health and Human Services, US Public Health Service, Agency for Toxic Substances and Disease Registry, August 1995.
16. Toxicological Profile For Dinitrophenols. U.S. Department Of Health And Human Services, US Public Health Service, Agency for Toxic Substances and Disease Registry, August 1995.
17. Toxicological Profile For 1,3-Dinitrobenzene and 1,3,5-Trinitrobenzene. U.S. Department Of Health And Human Services, US Public Health Service, Agency for Toxic Substances and Disease Registry, August 1995.
18. Toxicological Profile For 2,4- and 2,6-Dinitrotoluene. U.S. Department Of Health And Human Services, US Public Health Service, Agency for Toxic Substances and Disease Registry, December 1998.
19. Toxicological Profile For Nitrobenzene. Agency for Toxic Substances and Disease Registry, U.S. Public Health Service, December 1990.
20. Becke, A.D., *Density-Functional Thermochemistry .3. The Role of Exact Exchange*, J. Chem. Phys., 98 (1993) 5648.
21. Lee, C., Yang, W. and Parr, R.G., *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density*, Phys. Rev. B, 37 (1988) 785.
22. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Zakrzewski, V.G., Montgomery, Jr., J.A.; Stratmann, R.E., Burant, J.C., Dapprich, S., Millam, J.M., Daniels, A.D., Kudin, K.N., Strain, M.C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G.A., Ayala, P.Y.; Cui, Q., Morokuma, K., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Cioslowski, J., Ortiz, J.V., Baboul, A.G., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R.L., Fox, D.J., Keith, T., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Gonzalez, C., Challacombe, M., Gill, P.M.W., Johnson, B., Chen, W., Wong, M.W., Andres, J.L., Gonzalez, C., Head-Gordon, M., Replogle, E.S., and Pople, J.A., Gaussian 98, Revision A.11, Gaussian, Pittsburgh PA, 1998.
23. Todeschini, R. and Consonni, V. DRAGON software for the Calculation of Molecular Descriptors, web version 3.0 for Windows, 2003.
24. Todeschini, R. and Consonni, V. Handbook of Molecular Descriptors, Wiley-VCH, Weinheim and New York, 2000.
25. Davis, L. Handbook of Genetic Algorithms, Van Nostrand Reinhold, N.Y. (USA), 1991.
26. Devillers, J. Genetic Algorithms in Molecular Modeling, Academic Press, Ltd., London, 1996.
27. de Oliveira, D.B. and Gaudio, A.C., *BuildQSAR: A new computer program for QSAR studies*, Quant. Struct.-Act. Relat., 19 (2000) 599–601.
28. Shishkov, I., Vilkov, L.V., Kovacs, A. and Hargittai. I., *Molecular geometry of 2-nitrotoluene from gas phase electron diffraction and quantum chemical study*, J. Mol. Structure, 445 (1998) 259–268.
29. Sadova, N.I., Khaikin, L.S. and Vilkov, L.V., *Certain questions of stereochemistry of nitrogen-compounds in the gaseous-phase*, Russ. Chem. Rev., 61 (1992) 2129–2171.
30. Chiş, V., *Molecular and vibrational structure of 2,4-dinitrophenol: FT-IR, FT-Raman and quantum chemical calculations*, Chem. Phys., 300 (2004) 1–3.
31. Lampert, H., Mikenda, W. and Karpfen, A., *Intramolecular hydrogen bonding in 2-hydroxybenzoyl compounds: Infrared spectra and quantum chemical calculations*, J. Phys. Chem., 100 (1996) 7418–7424.
32. Lide, D.R. (Ed.) CRC Handbook of Chemistry and Physics 74th Edition, CRC Press, Boca Raton, 2000.
33. Barve, J.V. and Pant, L.M., *Structure of para-nitrotoluene*, Acta Cryst., B27 (1971) 1158–1162.
34. Iwasaki, F. and Kawano, Y., *Crystal and molecular-structure of ortho-nitrophenol*, Acta Cryst., B34 (1978) 1286–1290.
35. Pandarese, F., Ungaretti, L. and Coda, A., *Crystal-structure of a monoclinic phase of meta-nitrophenol*, Acta Cryst., B31 (1975) 2671–2675.
36. Coppens, P. and Schmidt, G.M.J., The crystal structure of the a-modification of p-nitrophenol near 90 K, Acta Cryst., 18 (1965) 62–67.
37. Mak, T.C.W. and Trotter, J., *The crystal structure of p-chloronitrobenzene*, Acta Cryst., 15 (1962) 1078–1080.
38. Trotter, J. and Williston, C.S., *Bond lengths and thermal vibrations in m-dinitrobenzene*, Acta Cryst., 21 (1966) 285–288.

39. Kagawa, T., Kawai, R. and Haisa, M., *The crystal and molecular structure of 2,4-dinitrophenol*, Acta Cryst., B32 (1976) 3171–3175.
40. Wilkins, A. and Small, R.W.H., *Structure of 1-fluoro-2,4-dinitrobenzene*, Acta Cryst., C47 (1991) 220–221.
41. Kovac, A. and Hargittai, I., *Theoretical investigation of the additivity of structural substituent effects in benzene derivatives*, Struct. Chem., 11 (2000) 193–201.
42. Di Labio, G.A., Pratt, D.A. and Wright, J.S., *Theoretical calculation of gas-phase ionization potentials for mono- and polysubstituted benzenes*, Chem. Phys. Lett, 311 (1999) 215–220.
43. Krygowski, T.M., Ejsmont, K., Stepien, B.T., Cyranski, M.K., Poater, J. and Sola, M. *Relation between the substituent effect and aromaticity*, J. Org. Chem., 69 (2004) 6634–6640.
44. Linstrom, P.J. and Mallard, W.G., (Eds.) NIST Chemistry WebBook, NIST Standard Reference Database Number 69, (http://webbook.nist.gov), March 2003.
45. Bentley, T.W. and Johnstone, R.A.W., *Aspects of mass spectra of organic compounds .8. calculation of ionization potentials of disubstituted benzenes and their importance in hammett correlations in mass spectrometry*, J. Chem. Soc. B, 2 (1971) 263–270.
46. Hehre, W.J., Radom, L., Schleyer, P. and Pople, J.A. Ab initio Molecular Orbital Theory, Wiley, New York, 1986.
47. Prabhumirashi, L.S. and Kunte, S.S., *Solvent effects on electronic absorption spectra of nitrochlorobenzenes, nitrophenols and nitroanilines – I. Studies in nonpolar solvents*, Spectrochim Acta A, 42A (1986) 435–439.
48. Desfrançois, C., Périquet, V., Lyapustina, S.A., Lippa, T.P., Robinson, D.W., Bowen, K.H., Nonaka, H. and Compton, R.N., *Electron binding to valence and multipole states of molecules: Nitrobenzene, para- and meta-dinitrobenzenes*, J. Chem. Phys., 111 (1999) 4569–4576.
49. Hall, L.H., Mohney, B. and Kier, L.B., *The electrotopological state - structure information at the atomic level for molecular graphs*, J. Chem. Inf. Comp Sci., 31 (1991) 76–82.
50. Huuskonen, J., *Estimation of water solubility from atom-type electrotopological state indices*, Env. Toxicol. Chem., 20 (2001) 491–497.
51. Livingstone, D.J., Ford, M.G., Huuskonen, J.J. and Salt, D.W., *Simultaneous prediction of aqueous solubility and octanol/water partition coefficient based on descriptors derived from molecular structure*, J. Comp. Aided Mol. Design, 15 (2001) 741–752.
52. Ghose, A.K. and Crippen, G.M., *Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure-Activity Relationships I. Partition Coefficients ans a Measure of Hydrophobicity*, J. Comput. Chem., 7 (1986) 565–577.
53. Hall, L.H., Mohney, B. and Kier, L.B., *An Electrotopological-State: An Atom Index for QSAR*, Quant. Struct. Act. Relat., 10 (1991) 43–51.
54. Kubinyi, H., Folkers, G. and Martin, Y.C. (Eds.) 3D QSAR in Drug Design, Vol. 2: Ligand-Protein Interactions and Molecular Similarity, Kluwer/ESCOM, Dordrecht (The Netherlands), 1998, pp. 339–353.
55. Devillers, J. and Balaban, A.T. (Eds.) Topological indices and related descriptors in QSAR and Drug Design, Gordon & Breach, Amsterdam, 2000.
56. Rietjens, I.C.M.M., Cnubben, N.H.P., Haandel, M., Tyrakowska, B., Soffers, A.E.M.F. and Vervoort, J., *Different metabolic pathways of 2,5-difluoronitrobenzene and 2,5-difluoroaminobenzene compared to molecular orbital substrate characteristics*, Chem. Biol. Interact., 94 (1995) 49–72.
57. Bearden, A.P. and Schultz, T.W., *Comparison of Tetrahymena and Pimephales toxicity based on mechanism of action*, SAR and QSAR in Environ. Res., 9 (1998) 127–53.